BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Nucleic acid chemistry II

Edited by Hans-Achim Wagenknecht

Imprint

Beilstein Journal of Organic Chemistry
www.bjoc.org
ISSN 1860-5397
Email: journals-support@beilstein-institut.de

The *Beilstein Journal of Organic Chemistry* is published by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften.

Beilstein-Institut zur Förderung der Chemischen Wissenschaften
Trakehner Straße 7–9
60487 Frankfurt am Main
Germany
www.beilstein-institut.de

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Synthesis of oligonucleotides on a soluble support

Harri Lönnberg

Review

Open Access

Address:
Department of Chemistry, University of Turku, FIN-20014 Turku, Finland

Email:
Harri Lönnberg - harlon@utu.fi

Keywords:
DNA; oligonucleotides; RNA; soluble support; synthesis

Beilstein J. Org. Chem. **2017,** *13,* 1368–1387.
doi:10.3762/bjoc.13.134

Received: 27 March 2017
Accepted: 19 June 2017
Published: 12 July 2017

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

Oligonucleotides are usually prepared in lab scale on a solid support with the aid of a fully automated synthesizer. Scaling up of the equipment has allowed industrial synthesis up to kilogram scale. In spite of this, solution-phase synthesis has received continuous interest, on one hand as a technique that could enable synthesis of even larger amounts and, on the other hand, as a gram scale laboratory synthesis without any special equipment. The synthesis on a soluble support has been regarded as an approach that could combine the advantageous features of both the solution and solid-phase syntheses. The critical step of this approach is the separation of the support-anchored oligonucleotide chain from the monomeric building block and other small molecular reagents and by-products after each coupling, oxidation and deprotection step. The techniques applied so far include precipitation, extraction, chromatography and nanofiltration. As regards coupling, all conventional chemistries, viz. phosphoramidite, *H*-phosphonate and phosphotriester strategies, have been attempted. While P(III)-based phosphoramidite and *H*-phosphonate chemistries are almost exclusively used on a solid support, the "outdated" P(V)-based phosphotriester chemistry still offers one major advantage for the synthesis on a soluble support; the omission of the oxidation step simplifies the coupling cycle. Several of protocols developed for the soluble-supported synthesis allow the preparation of both DNA and RNA oligomers of limited length in gram scale without any special equipment, being evidently of interest for research groups that need oligonucleotides in large amounts for research purposes. However, none of them has really tested at such a scale that the feasibility of their industrial use could be critically judged.

Introduction

The synthesis of oligonucleotides (ONs) consists of linking nucleosides to each other in a specified order by esterification of phosphoric acid with the 3′-OH of one and the 5′-OH of the other nucleoside. Usually, the 3′-OH is first esterified with an appropriate derivative of phosphoric acid and the resulting building block is then reacted with the 5′-OH (Figure 1). Either a linear or a convergent strategy may be utilized, but the stepwise linear approach proceeding from the 3′- to the 5′-terminus of ON is nowadays almost exclusively exploited [1,2]. The coupling reaction may take place either at oxidation level III or V

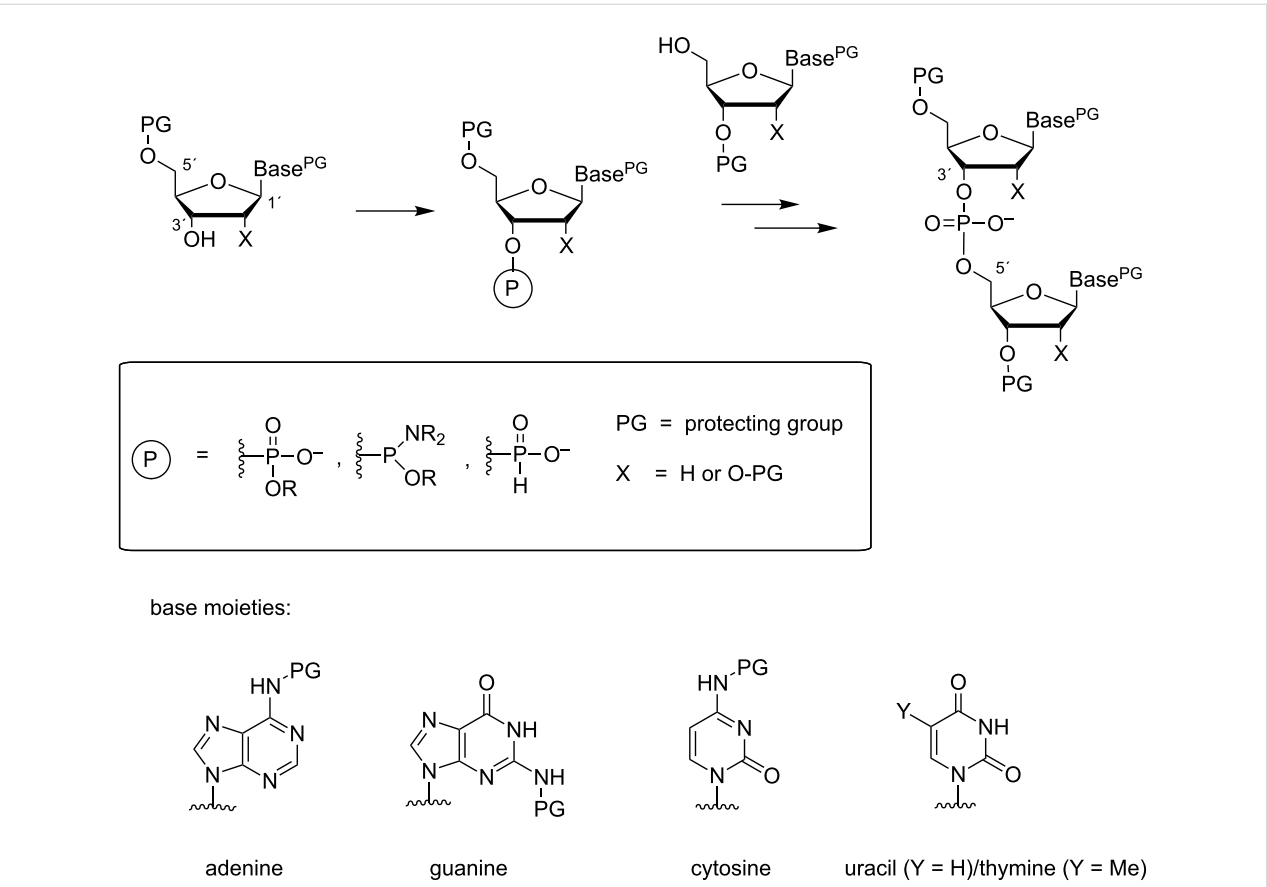
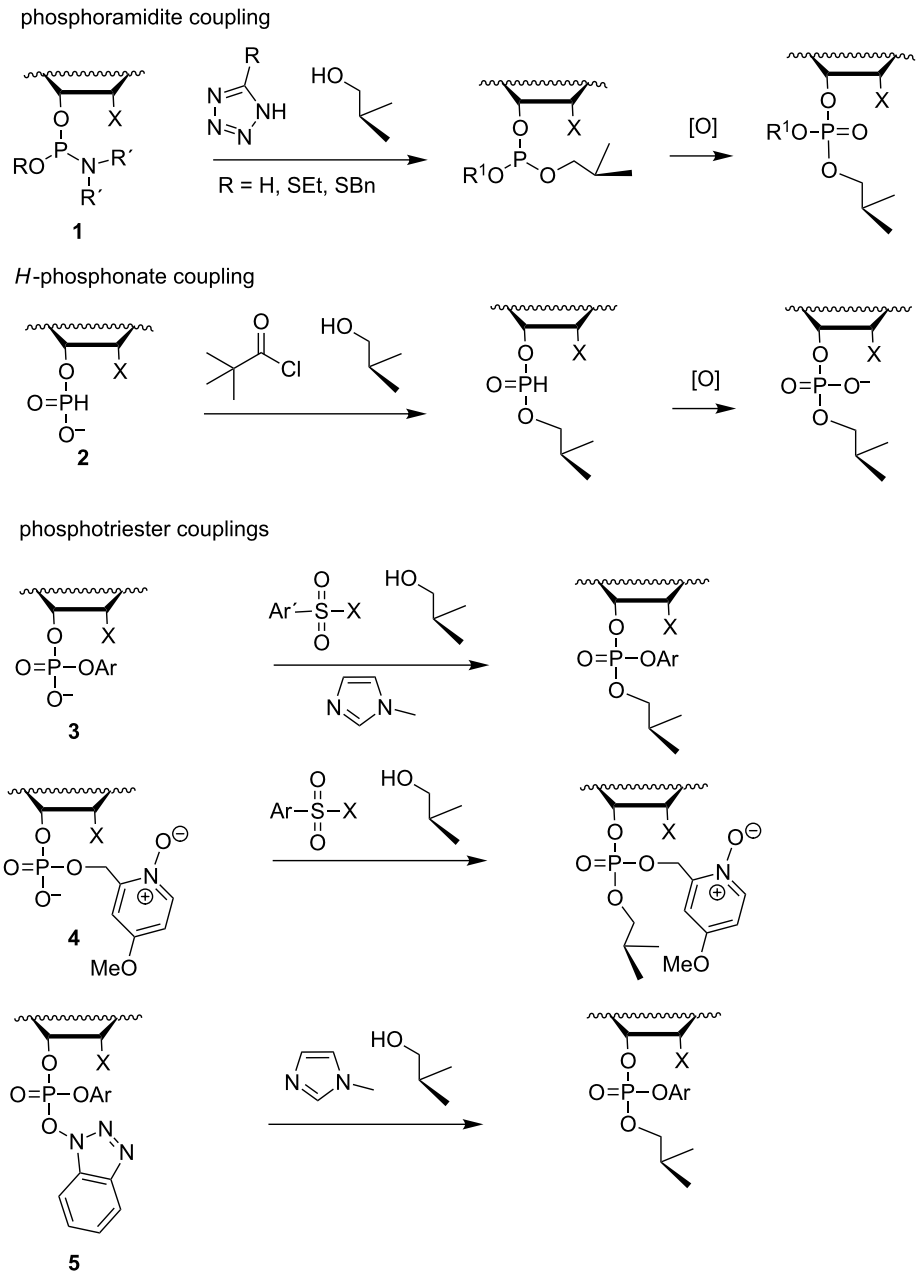


Figure 1: General principle of oligonucleotide synthesis.

of phosphorus. Owing to higher reactivity of P(III) centers, appropriately protected nucleoside 3′-(2-cyanoethyl-*N*,*N*-dialkylphosphoramidite)s (**1** in Scheme 1) or 3′-(*H*-phosphonate)s are usually preferred as building blocks [3] (**2** in Scheme 1). The attacking 5′-OH apart, all other nucleophilic functionalities must be kept protected during the coupling. The primary amino groups of the nucleobases are usually protected with acyl groups and the 5′-OH of the monomeric building block with a 4,4’-dimethoxytrityl group (DMTr), or sometimes with its monomethoxytrityl analog (MMTr) [4,5].

To achieve coupling, phosphoramidites are activated with azoles [6], such as tetrazole [7], its derivatives 2-ethyl- and 2-benzylthiotetrazole [8] or 4,5-dicyanoimidazole [9]. The activator has a dual role donating a proton to the departing dialkylamino group and attacking as an anionic species on phosphorus [10]. Nucleoside *H*-phosphonates are, in turn, converted in situ to reactive mixed anhydrides with acyl chlorides or chlorophosphates [11-13]. On applying the phosphoramidite chemistry, the phosphite triesters obtained are oxidized to phosphate triesters in each coupling cycle, whereas the *H*-phosphonate diesters may be stable enough to become oxidized only at the end of chain assembly. When the coupling is carried out at P(V) level, 3′-arylphosphate diesters (**3** in Scheme 1) are normally used as building blocks and activated with arylsulfonyl chloride or azolide in the presence of an auxiliary nucleophilic catalyst [14], or a catalytically active phosphate protecting group, such as the 4-methoxy-1-oxido-2-picolyl group [15], is used instead of a non-participating arylphosphate group (**4** in Scheme 1). Alternatively, prefabricated or in situ activated 1-hydroxybenzotriazole 3′-arylphosphotriesters may be used for coupling in the presence of a nucleophilic catalyst [16,17] (**5** in Scheme 1).

Compared to oligodeoxyribonucleotides (ODNs), the synthesis of oligoribonucleotides (ORNs) is complicated by the presence of an additional nucleophilic functionality, viz. the 2′-OH that has to be kept protected as long as basic conditions are required during synthesis and deprotection of the oligonucleotide. Since the phosphate protecting groups are normally base-labile and the repeatedly removable 5′-*O* protecting group is acid-labile, the 2′-*O*-protection should preferably be removable under orthogonal conditions. For this purpose, numerous protecting groups have been proposed [18,19], the fluoride ion labile *tert*-butyldimethylsilyl (TBDMS) [20]) and triisopropyl-



Scheme 1: Alternative coupling methods used in the synthesis of oligonucleotides.

silyloxymetyl (TOM) [21] groups being most widely used. Otherwise, the synthetic strategies are similar to those of ODNs.

The real breakthrough of the chemical synthesis of oligonucleotides was the finding of Beaucage and Caruthers in the early 1980s, according to which appropriately protected nucleosides could rapidly be coupled as 3′-(*O*-alkyl-*N*,*N*-dialkylphosphoramidite)s to 5′-OH of a support bound nucleoside by using tetrazole as an activator [7]. Since then, this solid-supported phosphoramidite chemistry has almost exclusively used for the preparation of oligonucleotides from lab scale [3,22] to industrial synthesis up to kilogram scale [23]. In spite of the obvious success of this methodology, synthesis in solution phase has received continuous interest as an alternative for large-scale synthesis, and the recent advances in the development of therapeutic oligonucleotides targeting either pre-mRNA [24,25], mature mRNA [26-28] or noncoding microRNA [29,30] have even increased this interest. It has been repeatedly argued that (i) the synthesis in solution could be carried out with a smaller excess of building blocks, (ii) the scale up procedure would be

more straightforward and (iii) expensive solid support material is not needed. In addition, the possibility to characterize the growing chain by mass or NMR spectroscopy after each coupling is an attractive feature, although not possible with all soluble supports. While major advances in the large scale solid-phase technology have been taken, the difference in the consumption of building blocks in solution and on a solid-support is not necessarily as substantial as previously assumed; the phosphoramidite-chemistry-based synthesis has been optimized to the level that building blocks are required only in a moderate excess, 1.5–2.0 equiv [23]. The obvious challenge is the separation of the support-anchored ON chain from small molecular reagents after each coupling cycle, a step that on a solid-support can be carried out by simple washing. Precipitation, chromatography, extraction and nanofiltration have been considered to be feasible approaches.

Even if the synthesis on a soluble support fails to compete with industrial solid-phase synthesis, it may still play an important role in up to gram scale laboratory synthesis, since no special equipment is usually needed. Spectroscopic studies on structure, dynamics and recognition of ONs by other biopolymers, small molecules or metal complexes, for example, may consume ONs in amounts that cannot be conveniently reached by lab-scale solid-phase synthesizers. In addition to synthesis on a soluble support, impressive examples of classical convergent synthesis [31-34] and exploitation of solid-supported reagents in solution [35,36] have been reported. The present review, however, surveys only the progress of ON synthesis on a soluble support.

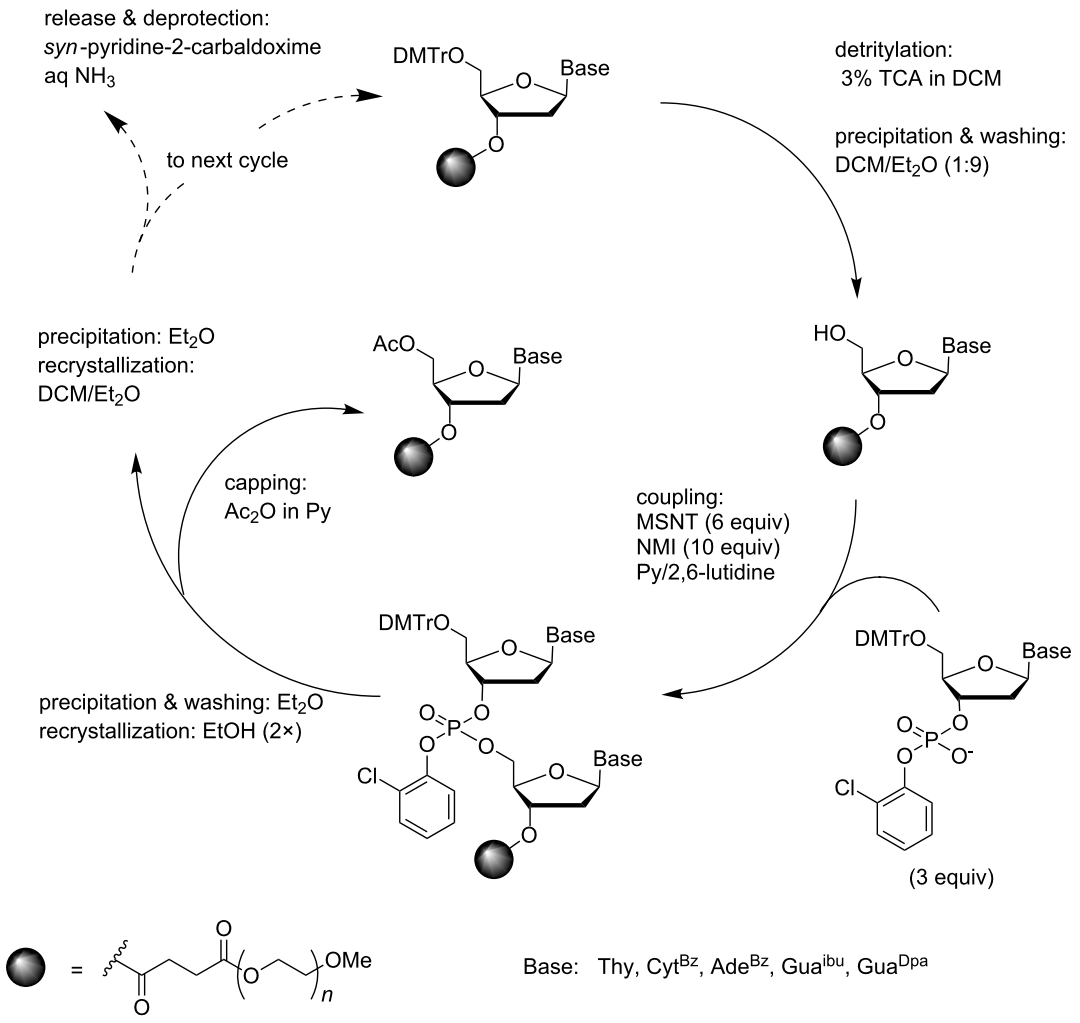
Review

Synthesis of oligodeoxyribonucleotides by phosphotriester chemistry

The pioneering syntheses of ONs on a soluble support were carried out by the phosphotriester strategy. Although this coupling chemistry is seldom used on a solid support where small molecule reagents and wastes can be removed by simple washing, the avoidance of the oxidation step due to use of P(V) synthons markedly simplifies the coupling cycle. This is a marked advantage in case of solution synthesis where the excess of reagents and wastes must be removed by a more laborious technique. The first synthesis of a reasonably long ODN, viz. an octamer d(5′-TAGCGCTA-3′), was carried out by Bonora et al. [37] on polyethylene glycol (PEG 5000) monomethyl ester. The overall strategy was rather similar to that of the solid-supported chemistry (Scheme 2). Accordingly, the 3′-terminal nucleoside, 5′-*O*-DMTr-N^6-Bz-dA, was attached to the support via a 3′-succinyl linker, the 5′-*O*-DMTr group was removed with 3% TCA in DCM and the derivatized support was isolated by precipitation with Et_2O and recrystallization from a 1:9 (v/v) mixture of DCM and Et_2O. 5′-*O*-DMTr-nucleosides (3.0 equiv of dT, dC^{Bz}, dG^{ibu}, dG^{Dpa}, dA^{Bz}) were then coupled as 3′-(2-chlorophenylphosphate)s in a mixture of pyridine and 2,6-lutidine using 1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole (MSNT; 6 equiv) as an activator and *N*-methylimidazole (NMI; 10 equiv) as a nucleophilic catalyst. Each coupling was followed by precipitation/recrystallization from EtOH, capping with Ac_2O in pyridine and precipitation from DCM/Et_2O. In spite of several precipitations and recrystallizations, one coupling cycle could be completed in 5 hours, the stepwise coupling yield ranging from 90% to 95% and the crude PEG-bound octamer was obtained in 79% yield. The coupling of dG^{ibu} proceeded, however, in more than 100% yield, which was interpreted as an indication of a side product formation. Evidently, the MSNT activation had resulted in displacement of O6 by the 3-nitro-1,2,4-triazol-1-yl group [38]. The oligomer was released from the support and deprotected by successive treatments with *syn*-pyridine-2-carbaldoxime and tetramethylguanidine in aq dioxane [39] and aq ammonia, and purified by ion-exchange chromatography on DEAE cellulose. From 980 mg of crude PEG-octamer, 85 mg of pure lyophilized TEA salt of d(5′-TAGCGCTA-3′) was obtained. In other words, the yield of the isolation step was less than 30%.

To avoid the modification of dG^{ibu} during the MSNT treatment, activation by 1-hydroxybenzotriazole, as originally introduced by Marugg et al. [40], was then attempted on the same PEG-support [41]. Accordingly, 3′-(2-chlorophenyl benzotriazol-1-yl phosphate)s of conventionally protected 2′-deoxynucleosides (3 equiv) were used as building blocks, and the coupling was carried out in a mixture of pyridine and dioxane in the presence of NMI (5 equiv). Otherwise, the protocol was similar to the previous one. The average stepwise coupling yield upon the assembly of octamer d(5′-TAGCGCTA-3′) was 93.5%, and 55% of the PEG-anchored oligomer could be isolated in pure deprotected form. No base modification reactions were now detected.

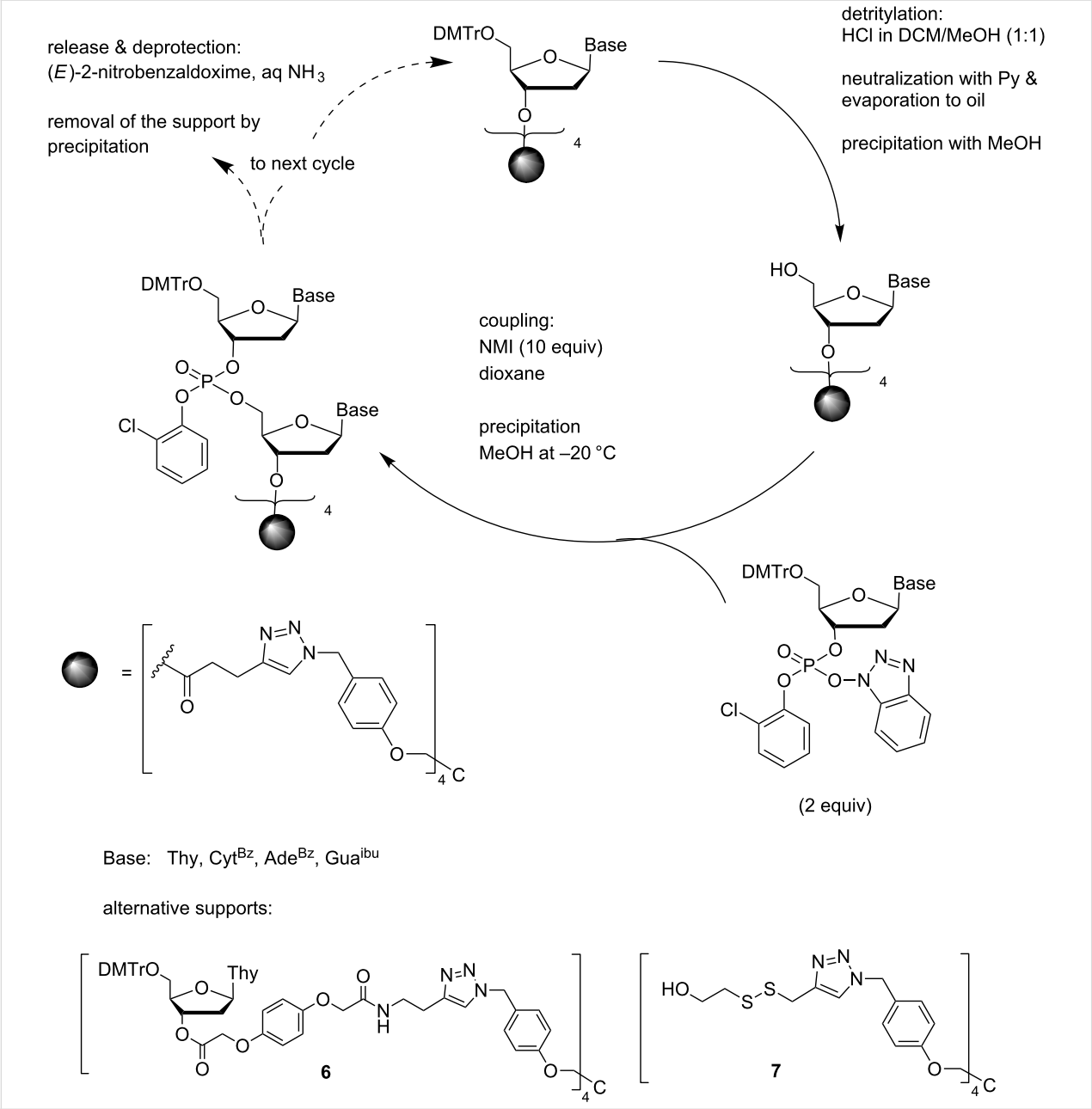
The phosphotriester approach based on hydroxybenzotriazole activation has more recently applied to the synthesis of short ODNs on a branched tetrakis-*O*-[4-(azidomethyl)phenyl]pentaerythritol-derived support (Scheme 3) [42]. Owing to the symmetrical structure of the support, NMR and mass spectroscopic characterization is possible at any stage of the chain assembly. The 3′-terminal nucleoside was immobilized to this support as a 3′-*O*-(4-pentynoyl) derivative by Cu(I)-catalyzed 1,3-dipolar cycloaddition [43]. This support is soluble in MeCN and dioxane but precipitates quantitatively in MeOH. Each coupling cycle contained two precipitations, one after removal of the 5′-*O*-DMTr group and the second after the coupling step. Detritylation was catalyzed with HCl in a 1:1 (v/v) mixture of



Scheme 2: Synthesis of ODNs on a precipitative PEG-support by phosphotriester chemistry using MSNT/NMI activation [37].

MeOH and DCM and coupling was carried out in dioxane in the presence of NMI. Precipitations were achieved by 10-fold dilution with MeOH. All small-molecule compounds remained in solution. Removal of the 2-chlorophenyl protections with the tetramethylguanidium salt of (*E*)-2-nitrobenzaldoxime in aqueous dioxane, followed by ammonolysis, removal of the support by precipitation and conversion to the sodium salt, completed the synthesis. A pentamer, d(5′-CGCAT-3′), homogeneous by HPLC, was obtained in 55% yield on using 2 equiv of building block in each coupling step. The advantages of such a tetrapodal support appear to be good atomic economy, i.e., small amount of support material compared to the amount of ORN obtained and the moderate consumption of solvent (MeOH) required for really quantitative precipitation of the support-bound oligonucleotides. However, only short oligomers have been so far prepared on this support. Support loaded with longer fully protected oligomers may precipitate less quantitatively or interchain aggregation may reduce the coupling efficiency.

A closely related support **6**, incorporating additionally a Q-linker moiety [44], has been used for preparation of fully protected ODN trimers having only the 3′-terminal hydroxy function unprotected and, hence, available for one step conversion to a phosphoramidite building block [45]. Such phosphoramidites are widely used for the assembly of ODNs useful in protein engineering by oligonucleotide directed mutagenesis [46-49]. Cleavage of the linker by 5 mmol L^{-1} K_2CO_3 in a 3:43:10 mixture of DCM, dioxane and MeOH (30 min), followed by neutralization with pyridinium chloride, left the 5′-*O*-DMTr group, 2-chlorophenyl phosphate protections and base moiety protections untouched. Silica gel chromatographic purification and conventional phosphitylation with 1-chloro-1-(2-cyanoethoxy)-*N*,*N*-diisopropylphosphoramidite gave the

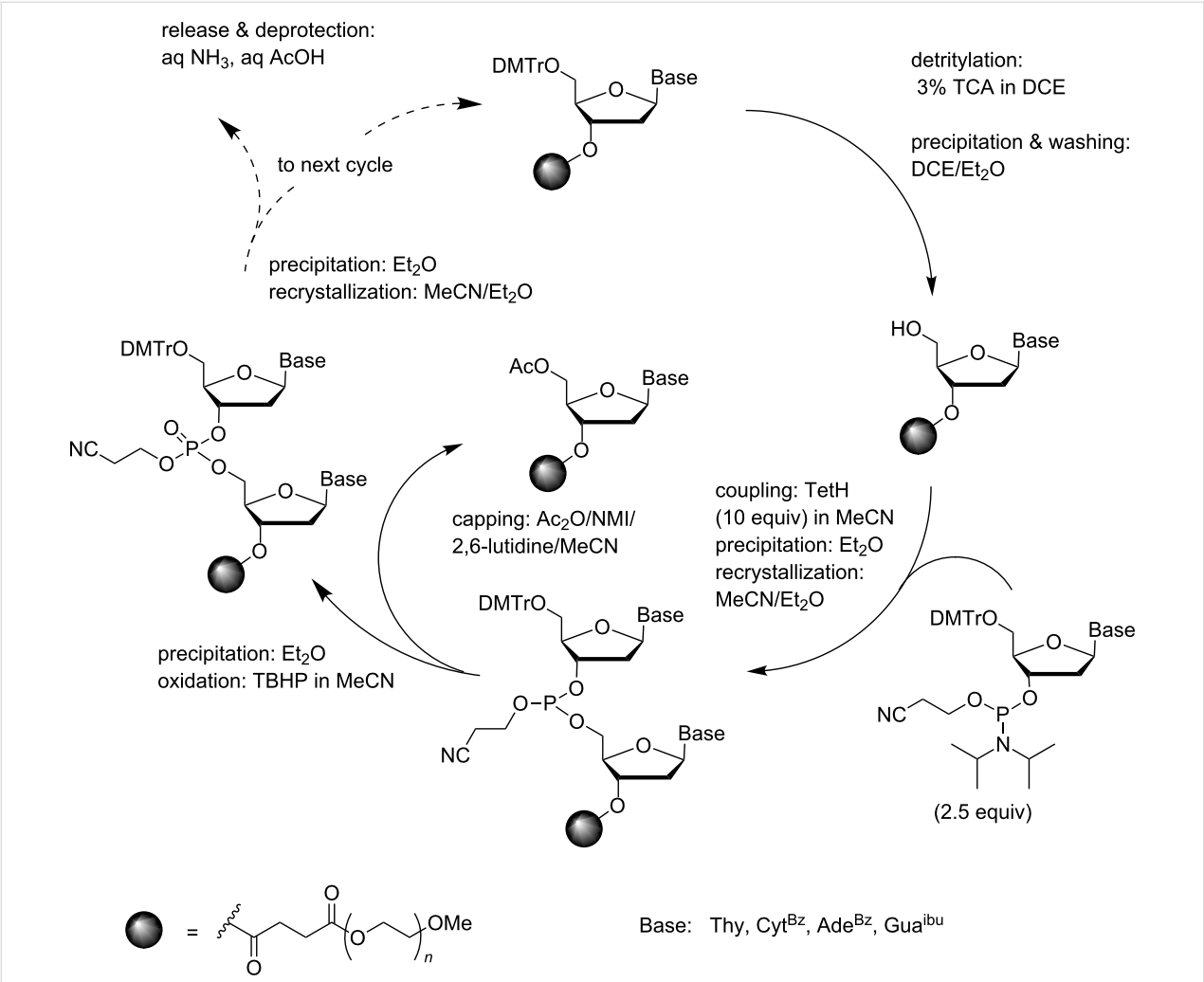


Scheme 3: Synthesis of ODNs on a precipitative tetrapodal support by phosphotriester chemistry using 1-hydroxybenzotriazole activation [42].

desired building blocks, the applicability of which in a solid-phase synthesis was demonstrated [45]. 3´-(2-Chlorophenyl)phosphates of protected trimeric ODNs, useful for phosphotriester coupling, have been prepared on a related reductively cleavable disulfide-linked support **7** [50].

Synthesis of oligodeoxyribonucleotides by phosphoramidite chemistry

As mentioned above, phosphoramidite chemistry is nowadays the method of choice for the solid-supported synthesis of oligonucleotides both in small and large scale. The first attempt to apply the phosphoramidite chemistry to synthesis on a soluble support dates back to 1993. Both the support (PEG) and overall strategy of chain assembly were in this pioneering study of Bonora et al. [51] similar to those used earlier in their synthesis of ODNs by the phosphotriester method. In other words, the support-bound material was separated from the low molecular weight substances by precipitation from Et_2O and recrystallization from a mixture of MeCN and Et_2O. In this case, four precipitation/recrystallization steps were needed in each coupling cycle: after detritylation, coupling, capping and oxidation (Scheme 4). The building blocks were base-moiety protected



Scheme 4: Synthesis of ODNs on a precipitative PEG-support by conventional phosphoramidite chemistry [51].

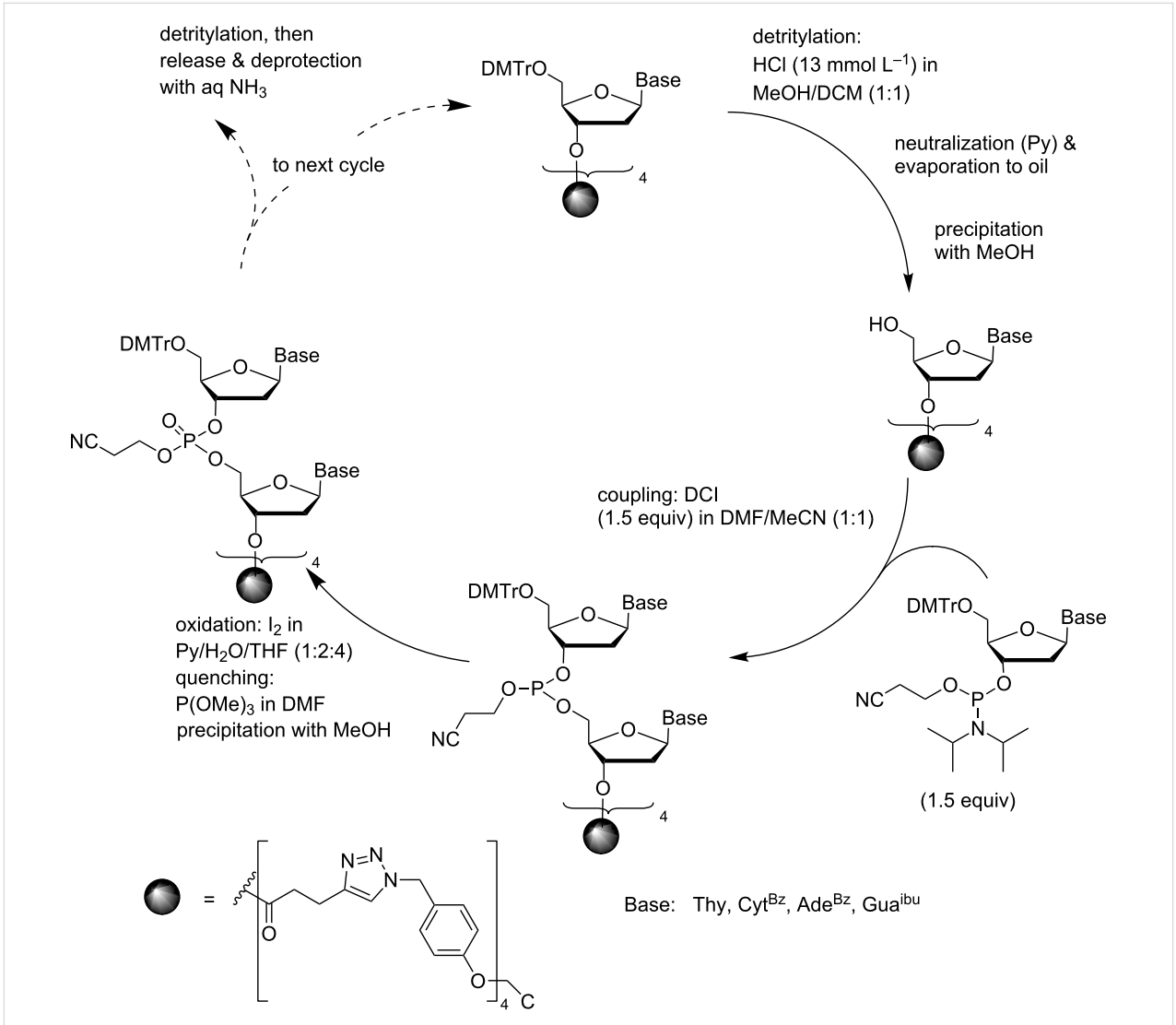
5′-*O*-DMTr-nucleoside 3′-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidites), i.e., the ones used in standard solid-supported synthesis. Phosphite triesters were oxidized to phosphate triesters after each coupling with *tert*-butyl hydroperoxide in MeCN [52]. On using 2.5 equiv of the phosphoramidite block and 10 equiv of tetrazole as an activator in MeCN, 98–99% coupling yields were obtained. Support-bound octamer, DMTr-d(5′-TAGCGCTA-3′)-PEG could be obtained in 93% yield and a 20-mer in 85% yield. These yields are surprisingly high, requiring 99% yield per coupling cycle. Release/deprotection by conventional ammonolysis followed by acidolytic detritylation and removal of the PEG-support by precipitation was reported to give the pure octamer in 50% higher yield than the phosphotriester approach.

The essentially same approach was later applied to the synthesis of a PEG-conjugated 12-mer antisense ODN [53] and a 13-mer purine-rich triple-helix forming sequence [54]. Immobilization of the 3′-terminal nucleosides via a succinyl linker was, however, replaced by direct phosphoramidite coupling to the terminal OH of PEG, which gave a stable phosphodiester linkage upon ammonolytical deprotection. In other words, the ODNs were used as PEG-conjugates in biological studies. In addition, a bifunctionalized PEG, bearing the acid labile DMTrO group at one end and a base labile Fmoc-NH functionality at the other end, has been used as a soluble support to obtain oligonucleotide–PEG–peptide conjugates [55,56]. The Fmoc protecting group was first removed and the peptide was assembled on the exposed amino function. Since the peptide moiety did not contain acid labile side chain protections, the oligonucleotide sequence could then be assembled by the protocol discussed above.

Another precipitative support that has been used for the synthesis of ODNs is the tetrapodal tetrakis-*O*-[4-(azidomethyl)phenyl)]pentaerythritol-derived support discussed above [43]. Two

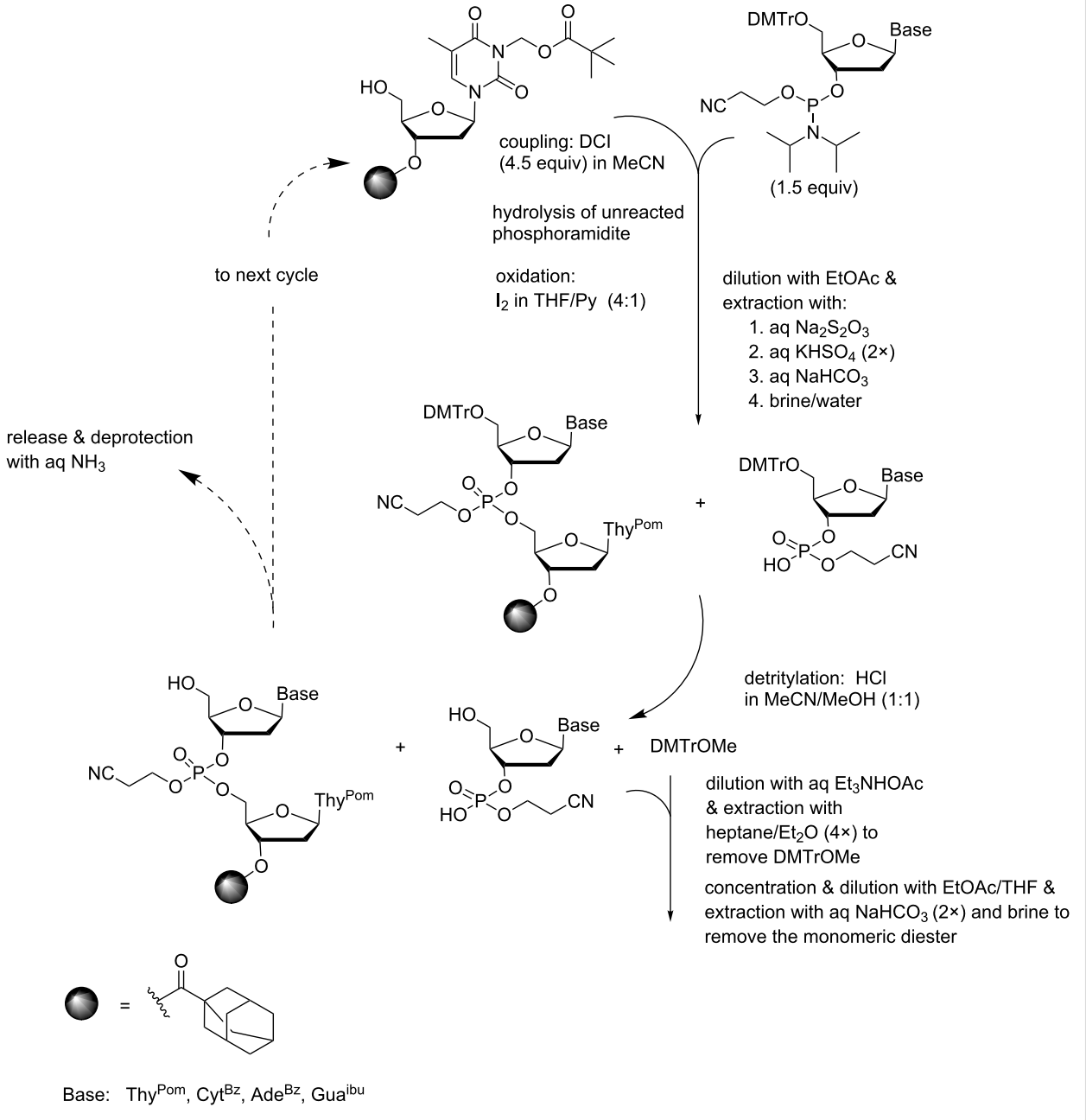
precipitations from MeOH were carried out in each coupling cycle: one after the 5′-O-detritylation and the second after the coupling/oxidation step (Scheme 5). The detritylation was carried out with HCl in a 1:1 (v/v) mixture of MeOH and DCM under carefully controlled conditions. The acid was neutralized with slight excess of pyridine. To prevent re-tritylation of the exposed 5′-OH by trityl carbocation, prolonged heating of the oily residue was avoided. Precipitation from MeOH quantitatively removed the traces of the DMTr carbocation as a methyl ether. Couplings were carried out in a 1:1 (v/v) mixture of DMF and MeCN using standard phosphoramidite building blocks (1.5 equiv) and 4,5-dicyanoimidazole (DCI, 1.5 equiv) as an activator. The resulting phosphite triesters were converted to phosphate esters by conventional aq iodine oxidation. Precipitation by dilution with MeOH removed all traces of reagents and monomeric nucleoside derivatives. As a proof of concept, a pentamer, d(5′-AGCCT-3′), was assembled. Release and deprotection of the oligomer by conventional ammonolysis were accompanied by precipitation of the support. The pentamer, homogeneous by HPLC, was obtained in a 43% yield.

Besides precipitation, extraction offers a possible approach for the separation of the soluble-supported oligonucleotides from small molecular materials. The underlying idea is to keep the growing oligonucleotide chain sufficiently hydrophobic to enable removal of the excess of building blocks, activators and wastes by water extraction, but still allow removal of highly hydrophobic substances, above all DMTrOMe, by extraction with very nonpolar solvents. The feasibility of this concept has been demonstrated by assembling a hexamer, d(5′-ATGCTT-3′), on 3′-(*O*-adamant-1-yl)acetyl-3-pivaloyloxymethylthymidine [57]. Twelve individual extractions had to be carried out in



Scheme 5: Synthesis of ODNs on a precipitative tetrapodal support by conventional phosphoramidite chemistry [43].

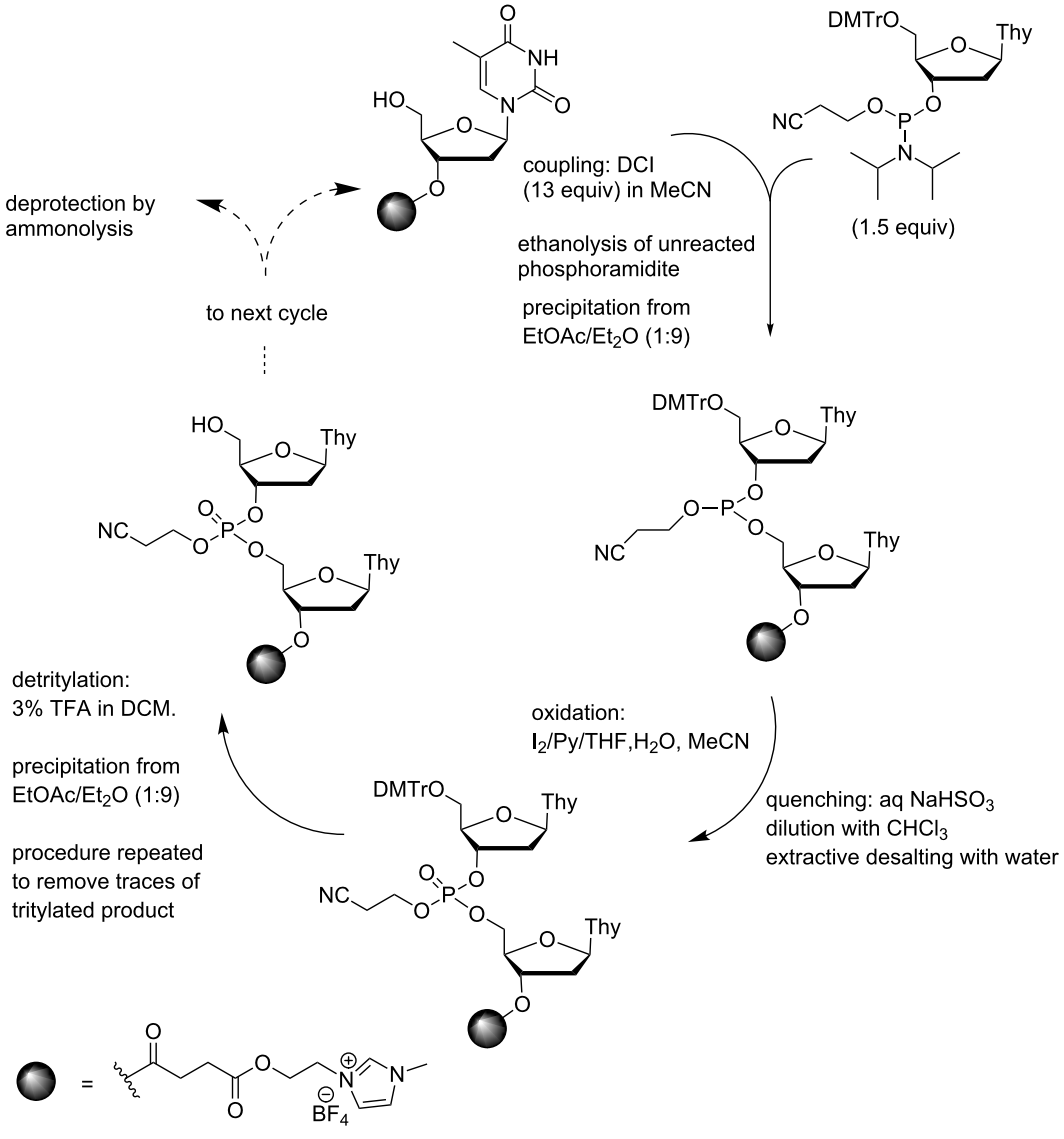
each synthetic cycle, as indicated in Scheme 6. First, DCI activated coupling in MeCN, hydrolysis of the unreacted phosphoramidite and subsequent I_2 oxidation in aq THF/pyridine was followed by dilution with EtOAc and washing with aq $Na_2S_2O_3$, aq $KHSO_4$ (twice), aq $NaHCO_3$ and brine. After HCl catalyzed detritylation in a 6:1 mixture of MeOH and MeCN, another set of extraction was performed. The mixture was neutralized with Et_3NHOAc and diluted with aq MeCN to give a 2:2:1 mixture of MeCN, MeOH and H_2O. The DMTrOMe byproduct was first removed by extracting four times with a 2:1 mixture of heptane and Et_2O. The polar phase was concentrated, diluted with a 5:2 mixture of EtOAc and THF, and washed twice with aq $NaHCO_3$ and then with diluted brine. Standard base moiety protections (dA^{Bz}, dC^{Bz}, dG^{ibu}) were employed, with the exception of thymine, which was used as a 3-pivaloyloxymethyl derivative to ensure sufficient hydrophobicity. On using 1.5 equiv of the phosphoramidite for coupling, the fully protected hexamer was obtained in 67% yield. Ammonolysis and ion-exchange chromatographic purification then gave hexamer d(5′-ATGCTT-3′) in isolated 39% yield.



Scheme 6: Synthesis of ODNs by an extractive strategy on an adamant-1-ylacetyl support [57].

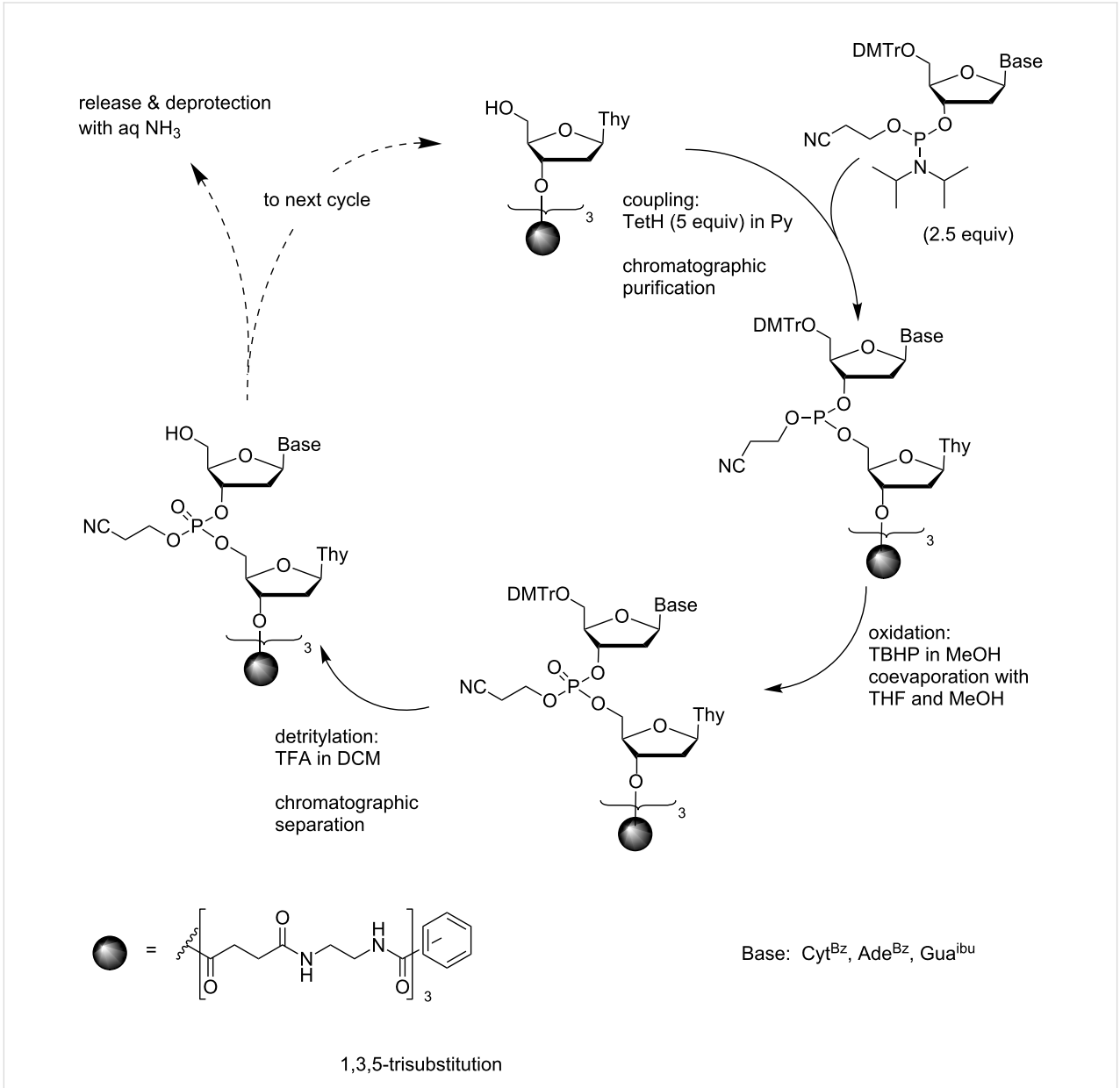
Esterification of a 5′-*O*-DMTr-3′-*O*-succinylthymidine with 3-(2-hydroxyethyl)-1-methyl-1*H*-imidazol-3-ium tetrafluoroborate has given another soluble support that allows utilization of extractive techniques, in this case in combination with precipitation [58] (Scheme 7). The support precipitates from a 1:9 mixture of EtOAc and Et_2O, but is soluble in chloroform, which allows removal of salts by extraction with water. The couplings were carried out with 1.5 equiv of standard 2-cyanoethyl-*N*,*N*-diisopropylphosphoramidites in THF or MeCN, using DCI as an activator. Unreacted phosphoramidites were quenched by EtOH and the support was precipitated before the oxidation step, repeatedly when needed. The precipitate was dissolved in MeCN and conventional aq I_2 oxidation was performed. After bisulfite quenching, the mixture was diluted with chloroform and washed with water to remove salts. The organic phase was evaporated to foam and subjected to detritylation with TFA in DCM or MeCN. The detritylated material was then precipitated with the EtOAc/Et_2O mixture. The product was, however, still partly tritylated, and the detritylation had therefore to be repeated. The longest oligomer synthesized was a thymidine tetramer. The yield of the support-bound tetramer was 87%, but no isolated yield was reported.

Although chromatographic separation appears to be a tedious procedure compared to precipitation or extraction, it has been successfully applied to the synthesis of ODNs on a soluble support. The studies of Wörl and Köster on N^1,N^3,N^5-tris(2-aminoethyl)benzene-1,3,5-tricarboxamide derivatized with



Scheme 7: Synthesis of ODNs by a combination of extractive and precipitative strategy [58].

3′-*O*-succinylthymidine offered an early example [59] (Scheme 8). Owing to poor solubility of the support into MeCN, elongation of the branches by tetrazole promoted coupling of nucleoside phosphoramidites was carried out in pyridine under argon. On using 2.5 equiv of the phosphoramidite and 5 equiv of tetrazole, the average coupling yield was 96%. The mixture was concentrated and subjected to gel permeation chromatography in MeOH to remove the low molecular weight compounds. The pooled fractions containing the support-bound oligonucleotides were concentrated and oxidized with *tert*-butyl hydroperoxide. The excess of oxidizing agent was removed by coevaporation with THF and MeOH, and the residue was dissolved into an 80:19:1 mixture of DCM, $MeNO_2$ and MeOH. Finally, the 5′-terminal DMTr groups were removed by adding 2% TFA. After neutralization with Et_3N, the chromatographic separation was repeated. Upon assembly of a fully protected 10-mer, d(5′-*O*-DMTr-$G^{ibu}A^{Bz}C^{Bz}G^{ibu}G^{ibu}C^{Bz}C^{Bz}A^{Bz}G^{ibu}T)_3$-support, the average yield of an entire coupling cycle was 87% and the overall yield 33%. Conventional ammonolysis was used for the release from the support. Since no capping reaction had been carried out in any coupling cycle, the n − 1 fragment was formed in a considerable amount. Assembly from dimeric phos-



Scheme 8: Synthesis of ODNs by phosphoramidite chemistry on a N^1,N^3,N^5-tris(2-aminoethyl)benzene-1,3,5-tricarboxamide support by making use of chromatographic separation [59].

phoramidites was additionally attempted, but the chromatographic separation was not efficient enough to remove the excess of the dimeric building block.

Much later chromatographic separation was exploited for the assembly of short ODNs from base-moiety-protected 5′-(1-methoxy-1-methylethyl)-2′-deoxyribonucleoside 3′-phosphoramidites on a fully methylated β-cyclodextrin support [60]. The 1-methoxy-1-methylethyl group may be removed by acid-catalyzed methanolysis approximately as readily as the DMTr group, but it gives only volatile products. Accordingly, after removal of the 5′-protection, only evaporation was needed. The subsequent flash chromatographic purification was, in turn, rather straightforward owing to the hydrophobic support. After ammonolytic release and deprotection, the methylated cyclodextrin support could be removed by simple extraction with DCM. A pentameric oligonucleotide, 5′-TACTT-3′, was obtained in 52% yield on using 1.5 equiv of phosphoramidites and 1.5 equiv of DCI as an activator.

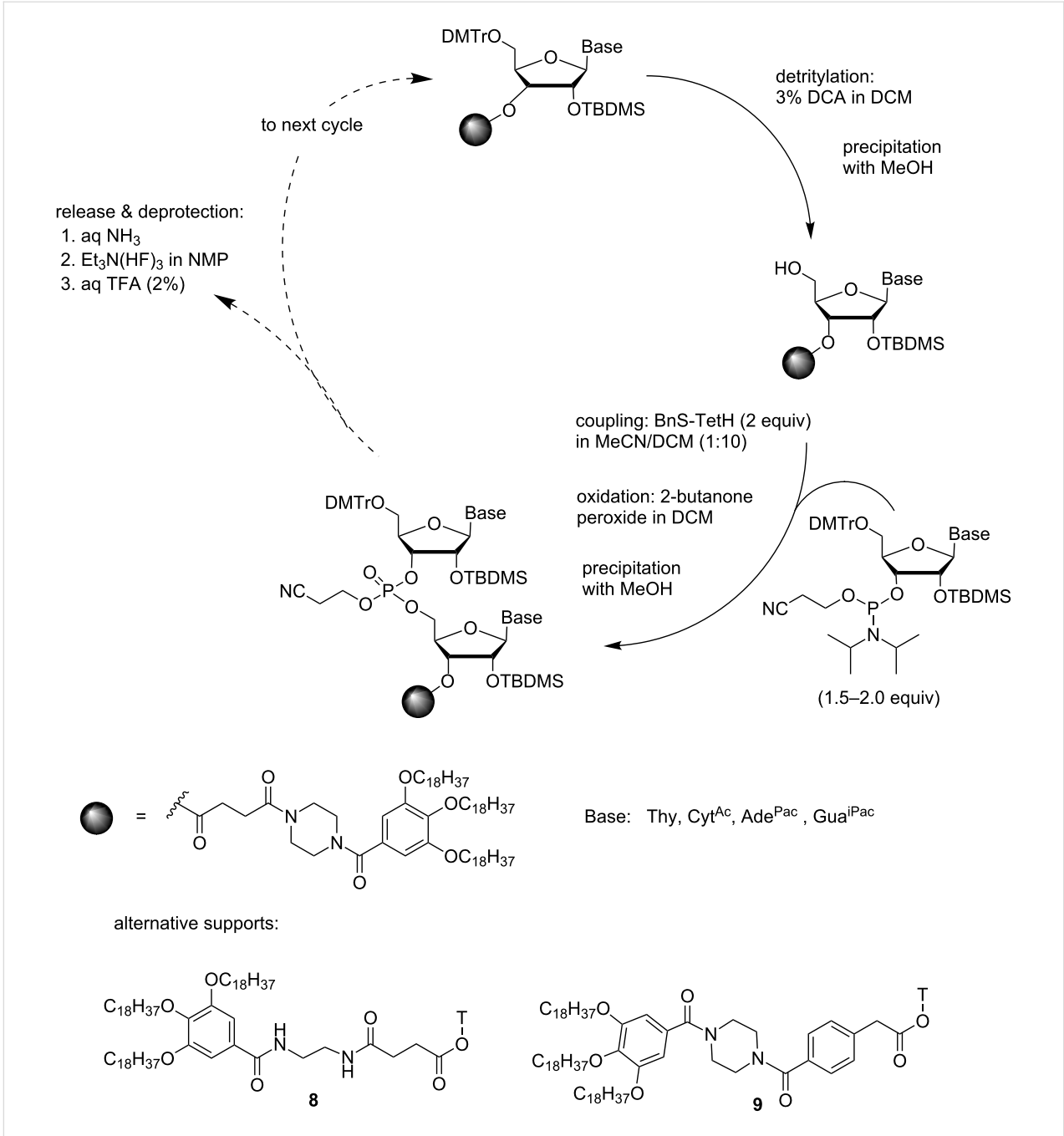
Synthesis of oligoribonucleotides by the phosphoramidite chemistry

Three different protocols, all based on separation of the support-bound oligonucleotide from low-molecular weight compounds by precipitation, have been utilized for the synthesis of oligoribonucleotides by phosphoramidite chemistry. A highly hydrophobic support that is well soluble in THF, $CHCl_3$ and DCM, but insoluble in MeOH, MeCN and EtCN, has been used to assemble a 21-mer RNA sequence in gram scale [61] (Scheme 9). First, the DMTr group was removed with DCA in DCM and the detritylated support was precipitated from MeOH. A base-moiety-protected (A^{Pac}, G^{iPac}, C^{Ac}) 5′-*O*-DMTr-2′-*O*-TBDMS-nucleoside 3′-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidite) (1.5–2.0 equiv) was then coupled in a 1:10 mixture of MeCN and DCM using 5-(benzylthio)-1*H*-tetrazole as an activator. After completion of the coupling, oxidation to the phosphate ester was carried out in the same pot by addition of 2-butanone peroxide in DCM. Dilution with MeOH precipitated the support. With 15–21-mer oligomers, some support-bound material, however, remained in solution and was recovered by adsorption to C18-coated silica gel. The cycle was completed by detritylation with DCA (3%) in DCM. Cleavage and deprotection was conventional: ammonolysis in aqueous EtOH, followed by desilylation with $Et_3N(HF)_3$ in *N*-methylpyridinone (NMP) and removal of the 5′-*O*-DMTr with aq TFA (2%). The isolated yield, 26%, is surprisingly high, taking into account that the synthesis involves more than 60 steps. In fact, the fully protected sequence was reported to be obtained in 46% yield, which corresponds to 98% yield per coupling cycle. Evidently the lack of amide hydrogens on the support is essential for the desired solubility properties, since replacement of the piperazine fragment within the linker structure with ethylene diamine **8** gave considerably less satisfactory results.

When the succinyl linker was replaced with the 4-carboxymethylbenzoic acid linker **9**, the fully protected oligomer could be released by catalytic hydrogenation. This allowed the preparation of appropriately protected dimeric and trimeric building blocks having only the 3′-terminal hydroxy function unprotected and, hence, subject to phosphitylation [62].

The terapodal tetrakis-*O*-[4-(azidomethyl)phenyl]pentaerythritol-derived support has also been used for the synthesis of short ORNs [63] (Scheme 10). Unusual 2′-*O*-(2-cyanoethyl)-5′-*O*-(1-methoxy-1-methylethyl)ribonucleoside 3′-phosphoramidites were used, since common commercially available building blocks turned out to be too hydrophobic to allow precipitation of the support-bound oligonucleotides from MeOH. The 1-methoxy-1-methylethyl group could be removed quantitatively as a dimethyl acetal of acetone upon acid-catalyzed transesterification in MeOH. The 3'-terminal nucleoside was attached to the support as a 3′-*O*-(4-pentynoyl) derivative, essentially as with 2′-deoxyribonucleosides. The acid-catalyzed removal of the 5′-*O*-1-methoxy-1-methylethyl group by 0.015 mol L^{-1} HCl in MeOH was essentially as fast as that of the DMTr group and no additional scavengers were needed to push the reaction to completion. Precipitation of the support from cold MeOH was quantitative. The phosphoramidite blocks were used in 50% excess and the coupling was promoted with DCI in a mixture of MeCN and DMF (1:1, v/v) under N_2. The phosphite triester obtained was oxidized to phosphate triester by conventional aqueous iodine treatment. The support was separated from all small molecular reagents by concentration to oil and subsequent precipitation from cold MeOH. Finally, the support-bound ORNs were subjected to consecutive treatments with triethylamine, ammonia and with TBAF. The fully deprotected ORNs were precipitated with NaOAc from EtOH. The hexamer, 5'-ACGUUU-3', was obtained in 54% yield, which means that the average coupling yield was 86%.

When 5′-*O*-DMTr-2′-*O*-TBDMS protected building blocks were used [64], instead of two precipitations from MeOH, each coupling cycle involved one precipitation from water and one flash chromatography (Scheme 11). Detritylation was carried out with HCl in a 2:5 (v/v) mixture of MeOH and DCM. The acid was neutralized with pyridine, the mixture concentrated to oil and subjected to flash column chromatography on silica gel. For subsequent coupling, the desired commercial block was used in 50% excess and DCI as an activator. After standard I_2 oxidation, the support-bound material was precipitated from water. The precipitation was quantitative, but some reagents and

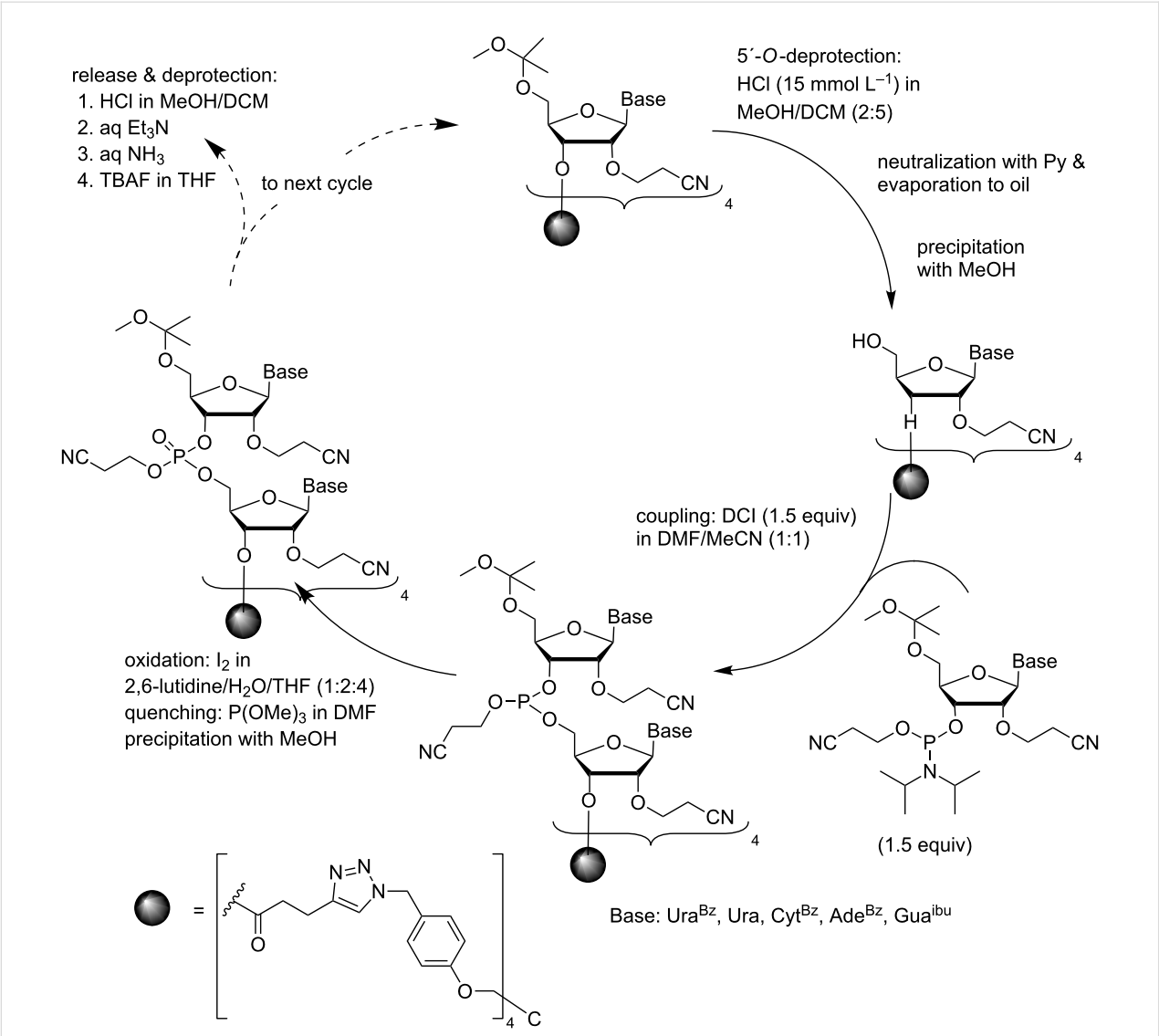


Scheme 9: Synthesis of ORNs by phosphoramidite chemistry on a hydrophobic support [61].

byproducts, above all DCI, coprecipitated with the support. The flash chromatography after next detritylation, however, removed these impurities. It is worth noting that the hydrophobic support greatly facilitated the chromatographic separation. After completion of the chain assembly, treatment with Et_3N, followed by ammonolysis and finally $Et_3N(HF)_3$ treatment, released the ORN, which was precipitated from cold MeOH with NaOAc. By this method, pentamer 5′-AGCUU-3′ was prepared in 46% yield.

Synthesis of oligodeoxyribonucleotides by the alkyl *H*-phosphonate chemistry

Surprisingly few attempts have been made to apply the *H*-phosphonate chemistry to the soluble-supported synthesis of oligonucleotides and most of these attempts have concerned the preparation of phosphorothioate ODNs, as discussed below. The only successful synthesis of unmodified ODNs was based on oxidative coupling of alkyl *H*-phosphonates on a PEG support [65]. The 3′-terminal nucleoside was immobilized to a

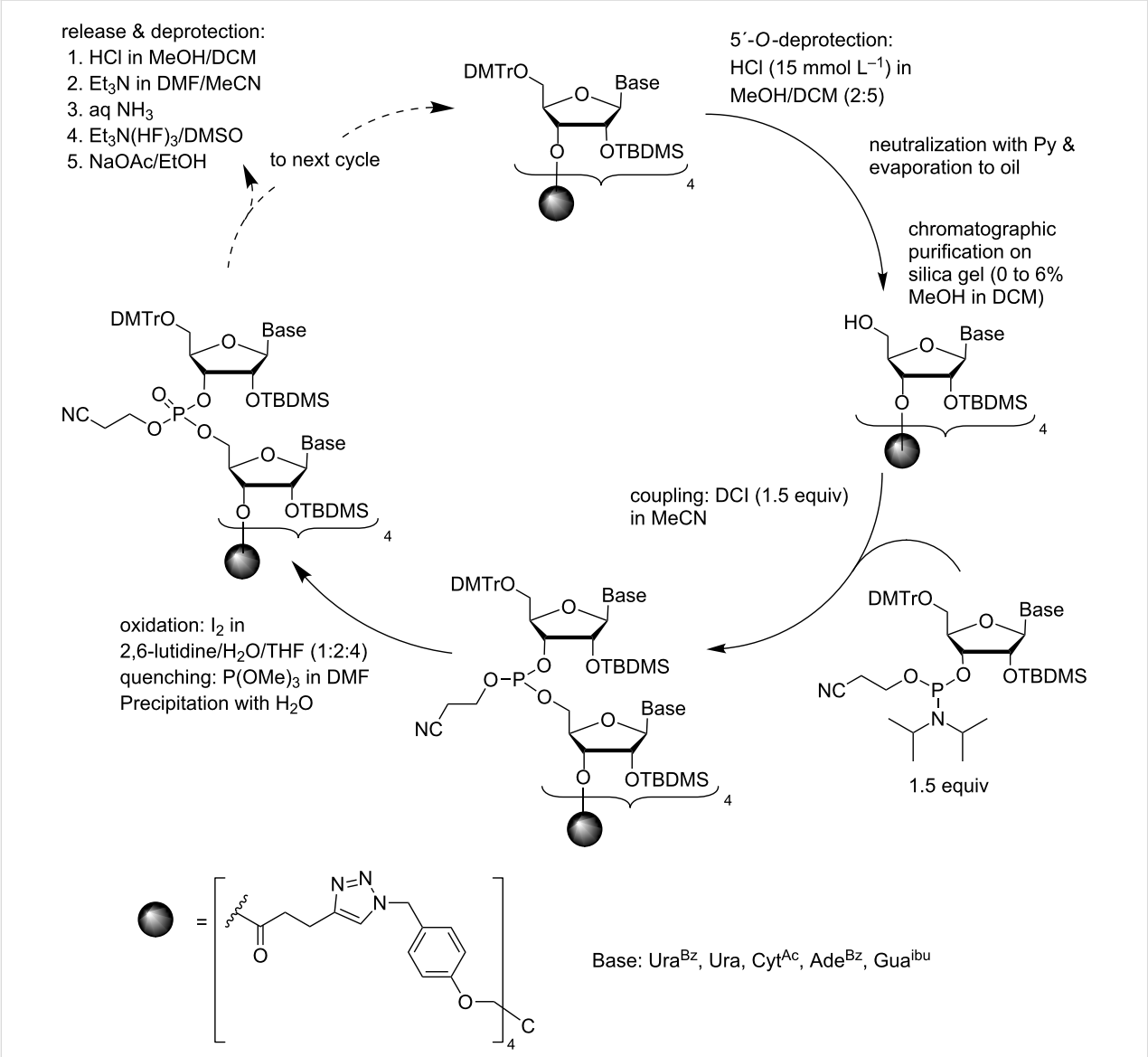


Scheme 10: Synthesis of ORNs by the phosphoramidite chemistry on a precipitative tetrapodal support using 2′-*O*-(2-cyanoethyl)-5′-*O*-(1-methoxy-1-methylethyl) protected building blocks [63].

PEG support via a succinyl linker, detritylated with DCA in DCM and precipitated and washed with Et_2O (Scheme 12). 3′-(2-Cyanoethyl *H*-phosphonate)s of 5′-*O*-DMTr-2′-deoxyribonucleosides were then used as synthons for the chain elongation. The oxidative couplings were carried out in a 4:1 (v/v) mixture of MeCN and Et_3N using *N*-bromosuccinimide (NBS) as an activator. The coupling efficiency was high (98%) on using 2.5 equiv of the *H*-phosphonate synthon and 5 equiv of the activator. After each coupling step, the support was precipitated from Et_2O and recrystallized from MeCN/Et_2O. The unreacted hydroxy groups were capped by acetylation and the support was again precipitated with Et_2O. Finally, ammonolysis was carried out and the oligonucleotide was separated from the PEG support by precipitation from MeOH. The feasibility of the method was tested by the synthesis of d(5′-ACGGGCCCGT-3′) in 75% yield.

Synthesis of oligonucleotide phosphorothioates

Phosphorothioate oligonucleotides have largely been synthesized by the same approaches as their oxygen counterparts. In fact, the only major difference is that oxidative sulfurization has been applied instead of oxidation. For example, when the phosphoramidite chemistry on a precipitative PEG support was applied, tetraethylthiuram disulfide (TETD; 0.5 mol L^{-1} in MeCN; 10-fold excess) was used as the sulfurization reagent [66] instead of *tert*-butyl hydroperoxide used for the oxidation in the synthesis of unmodified ODNs [51]. On using 2.5 equiv

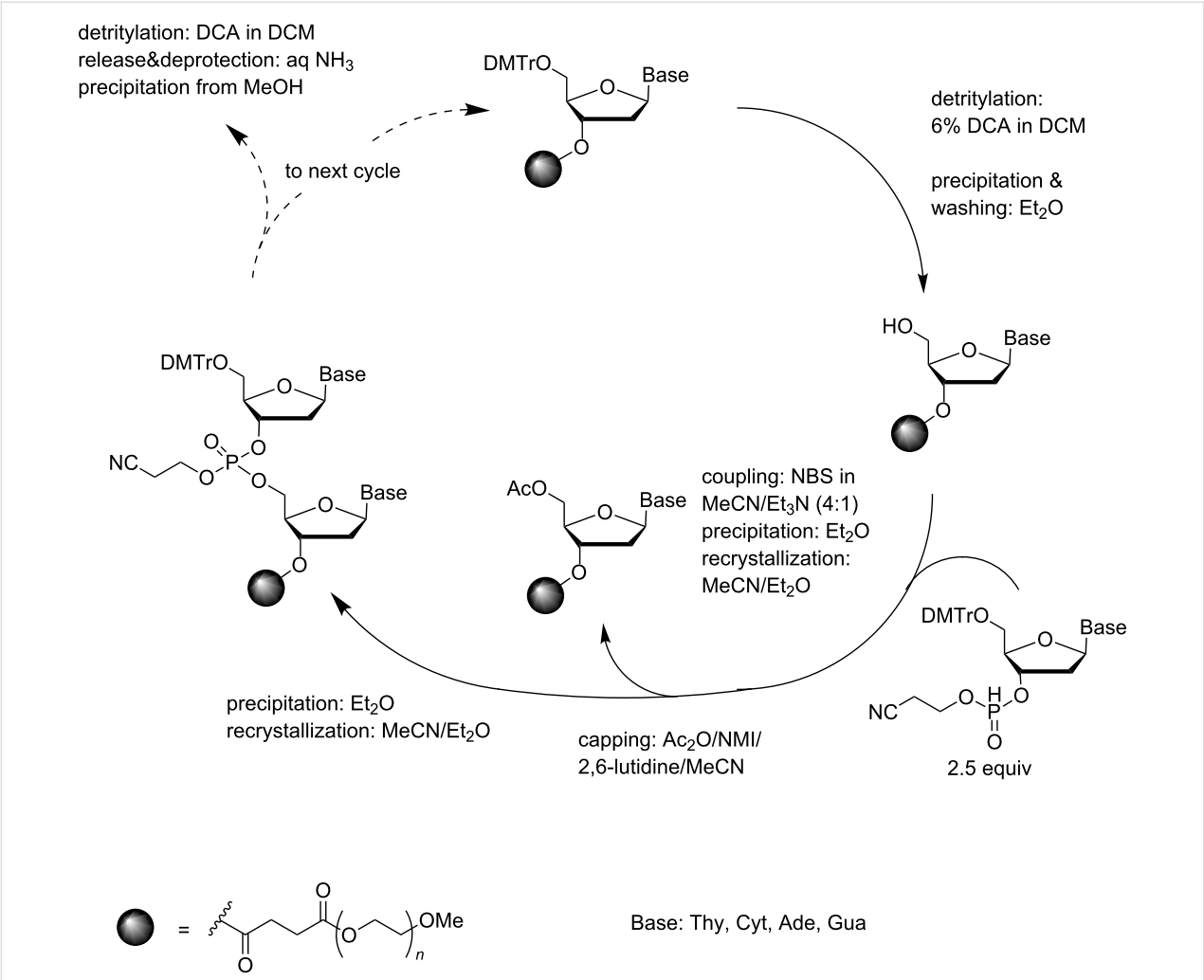


Scheme 11: Synthesis of ORNs by phosphoramidite chemistry on a precipitative tetrapodal support from commercially available building blocks [64].

of phosphoramidites for coupling, a support-bound 20-mer was obtained in 83% yield, and the pure oligomer could be isolated from the crude in 55% yield.

As a modification of this approach, phosphorothioate ODNs have been prepared by using 3′-phosphoramidites of dinucleoside-3′,5′-phosphorothioates as building blocks. The coupling efficiency was 99% on using 3.0 equiv of the dimeric building block [67]. The resulting phosphite triester was after each coupling oxidatively sulfurized with a 10-fold excess of diethyldithiocarbonate disulfide (DDD) [68]. The capping step after each sulfurization was carried out at 0 °C to avoid cleavage of the 2-cyanoethyl groups from the phosphorothioate triester linkages. Methyl *tert*-butyl ether was used for precipitations after detritylation and coupling/sulfurization steps. Detritylation with DCA in DCE, however, turned out to be somewhat problematic, since the procedure had often to be repeated. Conventional ammonolysis was used for the release from support and removal of base and phosphate protections. By this approach, a 15-mer phosphorothioate ODN (sequence not given, one G coupled as a monomer) was synthesized in 58% overall yield.

The development of new materials that allow nanofiltration in organic solvents has offered an entirely new paradigm for the soluble-supported synthesis of oligonucleotides. The underlying idea is that on passing the reaction mixture by high pressure through a membrane, small molecules pass through the

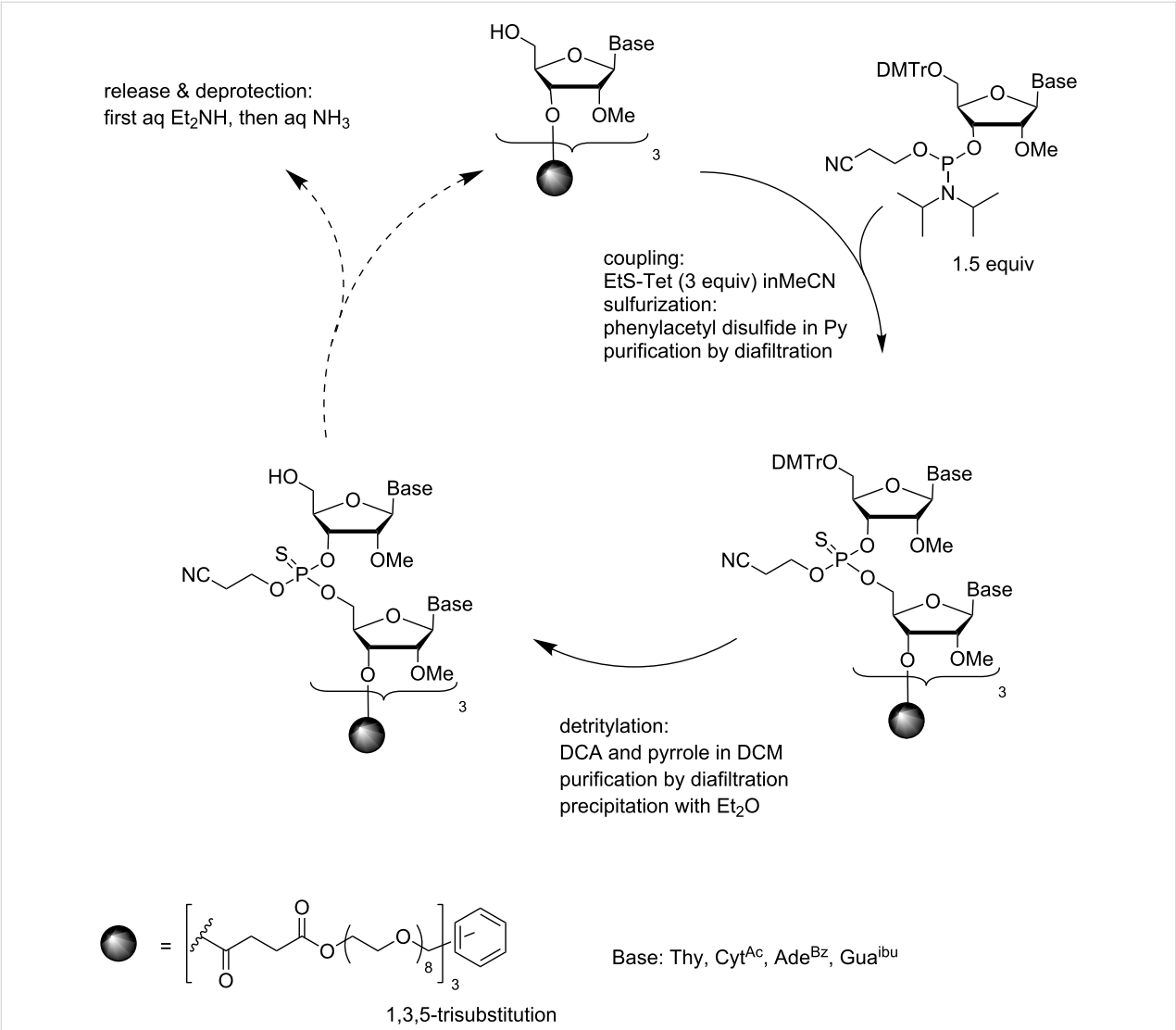


Scheme 12: Synthesis of ODNs on a precipitative PEG-support by *H*-phosphonate chemistry [65].

membrane, while the support is too bulky to escape through the nanopores of the membrane material. As a proof of concept, a 9-mer 2′-*O*-methyl oligoribonucleotide phosphorothioate has been synthesized [69,70]. The soluble support was 1,3,5-tris(hydroxymethyl)benzene derivatized with an eight units long PEG chain (Scheme 13), called homostar by the authors. The 3′-terminal nucleoside, in this case 5′-*O*-DMTr-2′-*O*-methyluridine, was attached via a succinyl linker to the terminal hydroxy functions of the support. Commercially available 5′-*O*-DMTr-2′-*O*-methylribonucleoside 3′-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidites (U, C^{Ac}, G^{ibu}, A^{Bz}) were employed for chain elongation. Ethylthiotetrazole-activated coupling (3 equiv per OH) in MeCN was followed by sulfurization with phenylacetyl disulfide in pyridine. All small molecule compounds were removed by the so-called diafiltration through a polybenzimidazole-based membrane PBI-17DBX [71,72]. In other words, the volume of the reaction mixture was kept unchanged during the filtration by continuous addition of pure solvent. After changing the solvent to DCM, detritylation with dichloroacetic acid was performed using pyrrole as a scavenger for the DMTr cation [73]. It turned out, however, that the DMTr-pyrrole formed could not be entirely removed by filtration, but a precipitation of the support with Et_2O was required for quantitative removal of this impurity. During the first four coupling cycles, the coupling yields gradually increased from 75 to 90%, and remained after that high (90–95%). Isolation of pure deprotected 9-mer, however, required HPLC purification and could be obtained in only 16% yield calculated from the crude support-bound material.

Conclusion

Several approaches based on precipitation, extraction, chromatographic separation or nanofiltration of a soluble support have been developed by making use of phosphoramidite, phosphotriester or *H*-phosphonate coupling. Usually these methods are aimed to be utilized for an industrial-scale synthesis of



Scheme 13: Synthesis of 2′-*O*-methyl ORN phosphorothioates by phosphoramidite chemistry by making use of nanofiltration in organic solvents [69,70].

oligonucleotides. However, none of them has really tested at such a scale that the feasibility of industrial use could be critically judged. The method based on nanofiltration in organic solvents clearly differs from the other approaches and represents a genuine effort towards an industrial process. The results still are very preliminary and the success will undoubtedly depend on further development of the membrane material and efficiency of the recyclization of the large solvent amounts.

All the other approaches discussed above allow lab-scale syntheses of oligonucleotides used for research purposes in gram scale. The advantage of the proposed soluble support strategies is that no special equipment is needed, and hence, they evidently are of interest for research groups that only now and then require large amounts of oligonucleotides for research in their main field. Comparison of the applicability of these methods is difficult on the basis of the data available. One interesting point is that some of the methods use capping, as usually on a solid support, whereas others omit it. None of the groups has carried out comparative studies that would shed light on the necessity of capping. Capping increases the number of manipulation but evidently simplifies the final purification. Which one is more important? Similarly, the phosphoramidite coupling is more efficient than phosphortriester or oxidative *H*-phosphonate coupling, but requires a separate oxidation step. Which one is more important, high coupling efficiency or simpler coupling cycle? Finally, it is worth noting that all the strategies proposed so far are based on acid-labile 5′-*O*-protection, although it inevitably leads to depurination as a side reaction, in particular on using acyl protections for the amino functions.

May be a proper solution to this old problem would open doors for success of oligonucleotide synthesis on a soluble support.

Abbreviations

Table 1: List of abbreviations.

A	adenosine
Ade	adenine
BnS-TetH	5-benzylthiotetrazole
Cyt	cytosine
dA	2´-deoxyadenosine
dC	2´-deoxycytidine
DCE	1,2-dichloroethane
DCI	4,5-dicyanoimidazole
DCM	dichloromethane
DDD	diethyldithiocarbonate disulfide
DEAE	2-(diethylamino)ethyl
dG	*2´-deoxyguanosine*
DMF	*N*,*N*-dimethylformamide
DMTr	4,4´-dimethoxytrityl
Dpa	diphenylacetyl
dT	thymidine
EtS-TetH	5-ethylthiotetrazole
G	guanosine
Gua	guanine
ibu	isobutyl
iPac	4-isopropylphenoxyacetyl
MMTr	4-methoxytrityl
MSNT	1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole
NBS	*N*-bromosuccinimide
NMI	*N*-methylimidazole
NMP	*N*-methyl-2-pyrrolidone
ODN	oligodeoxyribonucleotide
ON	oligonucleotide
ORN	oligoribonucleotide
Pac	phenoxyacetyl
PEG	polyethylene glycol
PG	protecting group
Pom	pivaloyloxymethyl
Py	pyridine
TBAF	tetrabutylammonium fluoride
TBDMS	*tert*-butyldimethylsilyl
TBHP	*tert*-butyl hydroperoxide
TCA	trichloroacetic acid
TEA	triethylammonium
TetH	tetrazole
TETD	tetraethylthiuram disulfide
THF	tetrahydrofuran
Thy	thymine
TOM	triisopropylsilyloxymethyl
U	uridine
Ura	uracil

References

1. Reese, C. B. *Org. Biomol. Chem.* **2005,** *3,* 3851–3868. doi:10.1039/b510458k
2. Abramova, T. *Molecules* **2013,** *18,* 1063–1075. doi:10.3390/molecules18011063
3. Roy, S.; Caruthers, M. *Molecules* **2013,** *18,* 14268–14284. doi:10.3390/molecules181114268
4. Agarwal, K. L.; Yamazaki, A.; Cashion, P. J.; Khorana, H. G. *Angew. Chem., Int. Ed. Engl.* **1972,** *11,* 451–459. doi:10.1002/anie.197204511
5. Reese, C. B. *Tetrahedron* **2002,** *58,* 8893–8920. doi:10.1016/S0040-4020(02)01084-0
6. Wei, X. *Tetrahedron* **2013,** *69,* 3615–3637. doi:10.1016/j.tet.2013.03.001
7. Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981,** *22,* 1859–1862. doi:10.1016/S0040-4039(01)90461-7
8. Stutz, A.; Höbartner, C.; Pitsch, S. *Helv. Chim. Acta* **2000,** *83,* 2477–2503. doi:10.1002/1522-2675(20000906)83:9<2477::AID-HLCA2477>3.0.CO;2-9
9. Vargeese, C.; Carter, J.; Yegge, J.; Krivjansky, S.; Settle, A.; Kropp, E.; Peterson, K.; Pieken, W. *Nucleic Acids Res.* **1998,** *26,* 1046–1050. doi:10.1093/nar/26.4.1046
10. Nurminen, E.; Lönnberg, H. *J. Phys. Org. Chem.* **2004,** *17,* 1–17. doi:10.1002/poc.681
11. Froehler, B. C.; Ng, P. G.; Matteucci, M. D. *Nucleic Acids Res.* **1986,** *14,* 5399–5407. doi:10.1093/nar/14.13.5399
12. Garegg, P. J.; Lindh, I.; Regberg, T.; Stawinski, J.; Strömberg, R. *Tetrahedron Lett.* **1986,** *27,* 4051–4054. doi:10.1016/S0040-4039(00)84908-4
13. Stawinski, J.; Strömberg, R. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2005,** *288,* 81–100. doi:10.1385/1-59259-823-4:081
14. Reese, C. B.; Zhang, P.-Z. *J. Chem. Soc., Perkin Trans. 1* **1993,** 2291–2301. doi:10.1039/p19930002291
15. Efimov, V. A.; Buryakova, A. A.; Dubey, I. Y.; Polushin, N. N.; Chakhmakhcheva, O. G.; Ovchinnikov, Y. A. *Nucleic Acids Res.* **1986,** *14,* 6525–6540. doi:10.1093/nar/14.16.6525
16. van der Marel, G.; van Boeckel, C. A. A.; Wille, G.; van Boom, J. H. *Tetrahedron Lett.* **1981,** *22,* 3887–3890. doi:10.1016/S0040-4039(01)91336-X
17. de Vroom, E.; Fidder, A.; Marugg, J. E.; van der Marel, G. A.; van Boom, J. H. *Nucleic Acids Res.* **1986,** *14,* 5885–5900. doi:10.1093/nar/14.14.5885
18. Somoza, A. *Chem. Soc. Rev.* **2008,** *37,* 2668–2675. doi:10.1039/b809851d
19. Dellinger, D. J.; Timár, Z.; Myerson, J.; Sierzchala, A. B.; Turner, J.; Ferreira, F.; Kupihár, Z.; Dellinger, G.; Hill, K. W.; Powell, J. A.; Sampson, J. R.; Caruthers, M. H. *J. Am. Chem. Soc.* **2011,** *133,* 11540–11556. doi:10.1021/ja201561z
20. Usman, N.; Ogilvie, K. K.; Jiang, M. Y.; Cedergren, R. J. *J. Am. Chem. Soc.* **1987,** *109,* 7845–7854. doi:10.1021/ja00259a037
21. Pitsch, S.; Weiss, P. A.; Jenny, L.; Stutz, A.; Wu, X. *Helv. Chim. Acta* **2001,** *84,* 3773–3795. doi:10.1002/1522-2675(20011219)84:12<3773::AID-HLCA3773>3.0.CO;2-E
22. Beaucage, S. L.; Iyer, R. P. *Tetrahedron* **1992,** *48,* 2223–2311. doi:10.1016/S0040-4020(01)88752-4

23. Sanghvi, Y. S. A Status Update of Modified Oligonucleotides for Chemotherapeutics Applications. *Current Protocols in Nucleic Acid Chemistry;* Wiley, 2011; Vol. 46, 4.1.1.–4.1.22. doi:10.1002/0471142700.nc0401s46
24. Havens, M. A.; Hastings, M. L. *Nucleic Acids Res.* **2016,** *44,* 6549–6563. doi:10.1093/nar/gkw533
25. Järver, P.; O'Donovan, L.; Gait, M. J. *Nucleic Acid Ther.* **2014,** *24,* 37–47. doi:10.1089/nat.2013.0454
26. Potaczek, D. P.; Garn, H.; Unger, S. D.; Renz, H. *J. Allergy Clin. Immunol.* **2016,** *137,* 1334–1346. doi:10.1016/j.jaci.2015.12.1344
27. Wittrup, A.; Lieberman, J. *Nat. Rev. Genet.* **2015,** *16,* 543–552. doi:10.1038/nrg3978
28. Sahin, U.; Karikó, K.; Türeci, Ö. *Nat. Rev. Drug Discovery* **2014,** *13,* 759–780. doi:10.1038/nrd4278
29. Cheng, C. J.; Bahal, R.; Babar, I. A.; Pincus, Z.; Barrera, F.; Liu, C.; Svoronos, A.; Braddock, D. T.; Glazer, P. M.; Engelman, D. M.; Saltzman, W. M.; Slack, F. J. *Nature* **2015,** *518,* 107–110. doi:10.1038/nature13905
30. Li, Z.; Rana, T. M. *Nat. Rev. Drug Discovery* **2014,** *13,* 622–638. doi:10.1038/nrd4359
31. Reese, C. B.; Yan, H. *J. Chem. Soc., Perkin Trans. 1* **2002,** 2619–2633. doi:10.1039/b208802a
32. Rejman, D.; Masojídková, M.; Rosenberg, I. *Nucleosides, Nucleotides Nucleic Acids* **2004,** *23,* 1683–1705. doi:10.1081/NCN-200033912
33. Efimov, V. A.; Aralov, A. V.; Fedunin, S. V.; Klykov, V. N.; Chakhmakhcheva, O. G. *Russ. J. Bioorg. Chem.* **2009,** *35,* 250–253. doi:10.1134/S1068162009020149
34. Aralov, A. V.; Klykov, V. N.; Chakhmakhcheva, O. G.; Efimov, V. A. *Russ. J. Bioorg. Chem.* **2011,** *37,* 586–592. doi:10.1134/S1068162011050025
35. Dueymes, C.; Schönberger, A.; Adamo, I.; Navarro, A.-E.; Meyer, A.; Lange, M.; Imbach, J.-L.; Link, F.; Morvan, F.; Vasseur, J.-J. *Org. Lett.* **2005,** *7,* 3485–3488. doi:10.1021/ol0511777
36. Adamo, I.; Dueymes, C.; Schönberger, A.; Navarro, A.-E.; Meyer, A.; Lange, M.; Imbach, J.-L.; Link, F.; Morvan, F.; Vasseur, J.-J. *Eur. J. Org. Chem.* **2006,** 436–448. doi:10.1002/ejoc.200500547
37. Bonora, G. M.; Scremin, C. L.; Colonna, F. P.; Garbesi, A. *Nucleic Acids Res.* **1990,** *18,* 3155–3159. doi:10.1093/nar/18.11.3155
38. Reese, C. B.; Song, Q. *J. Chem. Soc., Perkin Trans. 1* **1999,** 1477–1486. doi:10.1039/a901301f
39. Reese, C. B.; Zard, L. *Nucleic Acids Res.* **1981,** *9,* 4611–4626. doi:10.1093/nar/9.18.4611
40. Marugg, J. E.; McLaughlin, L. W.; Piel, N.; Tromp, M.; van der Marel, G. A.; van Boom, J. H. *Tetrahedron Lett.* **1983,** *24,* 3989–3992. doi:10.1016/S0040-4039(00)88244-1
41. Colonna, F. P.; Scremin, C. L.; Bonora, G. M. *Tetrahedron Lett.* **1991,** *32,* 3251–3254. doi:10.1016/S0040-4039(00)79736-X
42. Kungurtsev, V.; Virta, P.; Lönnberg, H. *Eur. J. Org. Chem.* **2013,** 7886–7890. doi:10.1002/ejoc.201301352
43. Kungurtsev, V.; Laakkonen, J.; Gimenez Molina, A.; Virta, P. *Eur. J. Org. Chem.* **2013,** 6687–6693. doi:10.1002/ejoc.201300864
44. Pon, R. T.; Yu, S. *Nucleic Acids Res.* **1997,** *25,* 3629–3635. doi:10.1093/nar/25.18.3629
45. Kungurtsev, V.; Lönnberg, H.; Virta, P. *RSC Adv.* **2016,** *6,* 105428–105432. doi:10.1039/C6RA22316H
46. Sondek, J.; Shortle, D. *Proc. Natl. Acad. Sci. U. S. A.* **1992,** *89,* 3581–3585. doi:10.1073/pnas.89.8.3581
47. Neylon, C. *Nucleic Acids Res.* **2004,** *32,* 1448–1459. doi:10.1093/nar/gkh315
48. Fellouse, F. A.; Esaki, K.; Birtalalan, S.; Raptis, D.; Cancasdi, V. J.; Koide, A.; Jhurani, P.; Vasser, M.; Wiesmann, C.; Kossiakoff, A. A.; Koide, S.; Sidhu, S. S. *J. Mol. Biol.* **2007,** *373,* 924–940. doi:10.1016/j.jmb.2007.08.005
49. Arunachalam, T. S.; Wichert, C.; Appel, B.; Müller, S. *Org. Biomol. Chem.* **2012,** *10,* 4641–4650. doi:10.1039/c2ob25328c
50. Jabgunde, A. M.; Molina, A. G.; Virta, P.; Lönnberg, H. *Beilstein J. Org. Chem.* **2015,** *11,* 1553–1560. doi:10.3762/bjoc.11.171
51. Bonora, G. M.; Biancotto, G.; Maffini, M.; Scremin, C. L. *Nucleic Acids Res.* **1993,** *21,* 1213–1217. doi:10.1093/nar/21.5.1213
52. Hayakawa, Y.; Uchiyama, M.; Noyori, R. *Tetrahedron Lett.* **1986,** *27,* 4191–4194. doi:10.1016/S0040-4039(00)84946-1
53. Bonora, G. M.; Ivanova, E.; Zarytova, V.; Burcovich, B.; Veronese, F. M. *Bioconjugate Chem.* **1997,** *8,* 793–797. doi:10.1021/bc970082p
54. Ballico, M.; Drioli, S.; Morvan, F.; Xodo, L.; Bonora, G. M. *Bioconjugate Chem.* **2001,** *12,* 719–725. doi:10.1021/bc010034b
55. Driolo, S.; Adamo, I.; Ballico, M.; Morvan, F.; Bonora, G. M. *Eur. J. Org. Chem.* **2002,** 3473–3480. doi:10.1002/1099-0690(200210)2002:20<3473::AID-EJOC3473>3.0.CO;2-2
56. Adamo, I.; Ballico, M.; Campaner, P.; Drioli, S.; Bonora, G. M. *Eur. J. Org. Chem.* **2004,** 2603–2609. doi:10.1002/ejoc.200400076
57. de Koning, M. C.; Ghisaidoobe, A. B. T.; Duynstee, H. I.; Ten Kortenaar, P. B. W.; Filippov, D. V.; van der Marel, G. A. *Org. Process Res. Dev.* **2006,** *10,* 1238–1245. doi:10.1021/op060133q
58. Donga, R. A.; Khaliq-Uz-Zaman, S. M.; Chan, T.-H.; Damha, M. J. *J. Org. Chem.* **2006,** *71,* 7907–7910. doi:10.1021/jo061279q
59. Wörl, R.; Köster, H. *Tetrahedron* **1999,** *55,* 2957–2972. doi:10.1016/S0040-4020(99)00059-9
60. Molina, A. G.; Kungurtsev, V.; Virta, P.; Lönnberg, H. *Molecules* **2012,** *17,* 12102–12120. doi:10.3390/molecules171012102
61. Kim, S.; Matsumoto, M.; Chiba, K. *Chem. – Eur. J.* **2013,** *19,* 8615–8620. doi:10.1002/chem.201300655
62. Matsuno, Y.; Shoji, T.; Kim, S.; Chiba, K. *Org. Lett.* **2016,** *18,* 800–803. doi:10.1021/acs.orglett.6b00077
63. Molina, A. G.; Jabgunde, A. M.; Virta, P.; Lönnberg, H. *Beilstein J. Org. Chem.* **2014,** *10,* 2279–2285. doi:10.3762/bjoc.10.237
64. Molina, A. G.; Jabgunde, A. M.; Virta, P.; Lönnberg, H. *Curr. Org. Synth.* **2015,** *12,* 202–207. doi:10.2174/1570179411666141120215703
65. Padiya, K. J.; Salunkhe, M. M. *Bioorg. Med. Chem.* **2000,** *8,* 337–342. doi:10.1016/S0968-0896(99)00287-4
66. Scremin, C. L.; Bonora, C. M. *Tetrahedron Lett.* **1993,** *34,* 4663–4666. doi:10.1016/S0040-4039(00)60651-2
67. Bonora, G. M.; Rossin, R.; Zaramella, S.; Cole, D. L.; Eleuteri, A.; Ravikumar, V. T. *Org. Process Res. Dev.* **2000,** *4,* 225–231. doi:10.1021/op990096l
68. Eleuteri, A.; Cheruvallath, Z. S.; Capaldi, D. C.; Cole, D. L.; Ravikumar, V. T. *Nucleosides Nucleotides* **1999,** *18,* 1803–1807. doi:10.1080/07328319908044844
69. Gaffney, P. R. J.; Kim, J. F.; Valtcheva, I. B.; Williams, G. D.; Anson, M. S.; Buswell, A. M.; Livingston, A. G. *Chem. – Eur. J.* **2015,** *21,* 9535–9543. doi:10.1002/chem.201501001
70. Kim, J. F.; Gaffney, P. R. J.; Valtcheva, I. B.; Williams, G.; Buswell, A. M.; Anson, M. S.; Livingston, A. G. *Org. Process Res. Dev.* **2016,** *20,* 1439–1452. doi:10.1021/acs.oprd.6b00139

71. Valtcheva, I. B.; Marchetti, P.; Livingston, A. G. *J. Membr. Sci.* **2015,** *493,* 568–579. doi:10.1016/j.memsci.2015.06.056
72. Valtcheva, I. B.; Kumbharkar, S. C.; Kim, J. F.; Bhole, Y.; Livingston, A. G. *J. Membr. Sci.* **2014,** *457,* 62–72. doi:10.1016/j.memsci.2013.12.069
73. Reese, C. B.; Serafinowska, H. T.; Zappia, G. *Tetrahedron Lett.* **1986,** *27,* 2291–2294. doi:10.1016/S0040-4039(00)84511-6

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Metal-mediated base pairs in parallel-stranded DNA

Jens Müller

Review

Open Access

Address:
Westfälische Wilhelms-Universität Münster, Institut für Anorganische und Analytische Chemie, Corrensstraße 30, 48149 Münster, Germany

Email:
Jens Müller - mueller.j@uni-muenster.de

Keywords:
DNA; metal-mediated base pairs; nucleic acids

Beilstein J. Org. Chem. **2017,** *13,* 2671–2681.
doi:10.3762/bjoc.13.265

Received: 26 September 2017
Accepted: 29 November 2017
Published: 13 December 2017

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

In nucleic acid chemistry, metal-mediated base pairs represent a versatile method for the site-specific introduction of metal-based functionality. In metal-mediated base pairs, the hydrogen bonds between complementary nucleobases are replaced by coordinate bonds to one or two transition metal ions located in the helical core. In recent years, the concept of metal-mediated base pairing has found a significant extension by applying it to parallel-stranded DNA duplexes. The antiparallel-stranded orientation of the complementary strands as found in natural B-DNA double helices enforces a cisoid orientation of the glycosidic bonds. To enable the formation of metal-mediated base pairs preferring a transoid orientation of the glycosidic bonds, parallel-stranded duplexes have been investigated. In many cases, such as the well-established cytosine–Ag(I)–cytosine base pair, metal complex formation is more stabilizing in parallel-stranded DNA than in antiparallel-stranded DNA. This review presents an overview of all metal-mediated base pairs reported as yet in parallel-stranded DNA, compares them with their counterparts in regular DNA (where available), and explains the experimental conditions used to stabilize the respective parallel-stranded duplexes.

Introduction

Nucleic acids are increasingly being applied in areas beyond their original biological context, e.g., as a scaffold for the defined spatial arrangement of functional entities [1-3]. This often goes along with the formal substitution of a canonical nucleoside (or any other nucleic acid component) by an artificial one that either bears the desired functionality or contains an anchor for a postsynthetic introduction of the functional moiety [4]. The site-specific incorporation of transition metal ions is nowadays typically achieved by introducing so-called metal-mediated base pairs into the duplex. In a metal-mediated base pair, the complementary nucleobases are pairing via coordinate bonds rather than hydrogen bonds (Figure 1). Metal-mediated base pairs can be obtained from natural nucleobases such as cytosine or thymine [5]. In addition, many artificial nucleobases have been developed for an application in metal-mediated base pairing [6,7]. Structural analyses have shown that their formation is possible without major conformational changes of the nucleic acid [8], even though metal-modified nucleic acids may very well adopt non-helical topologies [9]. It is even possible to create DNA duplexes composed of metal-

mediated base pairs only [10]. Possible applications of nucleic acids with metal-mediated base pairs exist in numerous fields [11]. More recently investigated areas include charge transfer in metal-modified DNA [12-14], the recognition of specific nucleic acid sequences [15-17], the creation of dynamic and switchable DNA nanostructures [18,19], and an exploitation of their processing by polymerases [20-24].

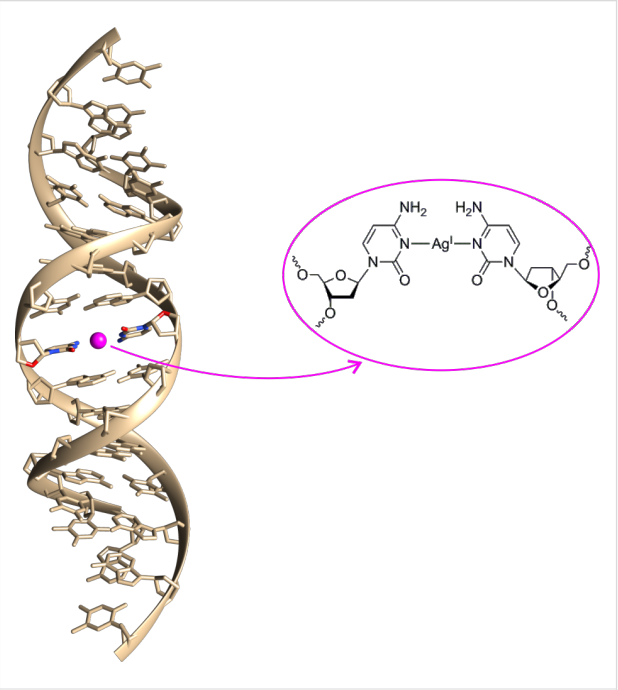


Figure 1: Structure of a B-DNA duplex comprising a central C–Ag(I)–C base pair. This figure was created based on the coordinates reported in PDB entry 2RVP [25].

Most metal-mediated base pairs reported so far have been introduced into canonical antiparallel-stranded nucleic acid duplexes [26], even though the idea of ligand-based nucleosides has also been applied to triplexes and quadruplexes [27-30]. More recently, metal-mediated base pairs have also been investigated in the context of parallel-stranded duplexes. This review presents an overview of metal-mediated base pairs introduced into parallel-stranded duplexes so far and compares them with the corresponding base pairs in regular antiparallel-stranded DNA. In the next section, it first introduces into the concept of parallel-stranded DNA and explains different experimental approaches to enforce a parallel alignment of the complementary oligonucleotide strands.

Review

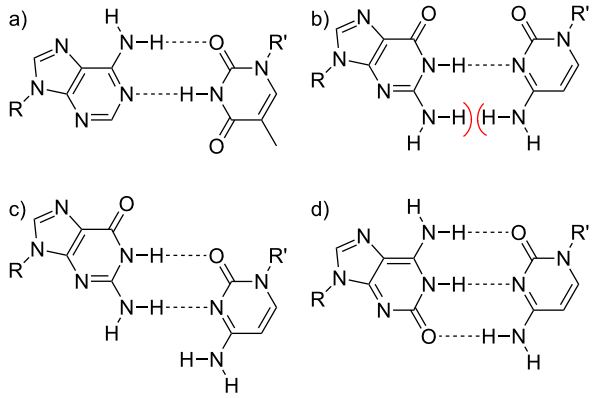
Parallel-stranded DNA

In canonical DNA duplexes, the complementary oligonucleotide strands are oriented in an antiparallel fashion. From a geometrical point of view, this correlates with a cisoid orientation of the glycosidic bonds in Watson–Crick base pairs. A parallel-stranded orientation of oligonucleotide strands may also occur in nature, albeit in more complex topologies such as triple helices or quadruplexes [31,32]. Nonetheless, the formation of parallel-stranded DNA duplexes can be induced in various ways, leading to a variety of non-canonical DNA duplex topologies that depend on the experimental approach taken to enforce a parallel alignment of the strands. Typically, parallel-stranded duplexes are less stable than their respective antiparallel-stranded counterparts [33,34]. This makes them of interest for the incorporation of metal-mediated base pairs, because the formation of such a base pair within an intrinsically unstable duplex is often accompanied by an exceptional thermal stabilisation, which in turn is advantageous for possible sensor applications. A feature of many base pairs in parallel-stranded duplexes is the transoid orientation of their glycosidic bonds, even though their formation is in principle also compatible with cisoid glycosidic bonds. Hence, metal-mediated base pairs that require a transoid orientation of the glycosidic bonds may be ideally generated in a parallel-stranded double helix. This section summarizes base pairing patterns established for parallel-stranded DNA duplexes in general and highlights experimental approaches feasible for the generation of such double helices.

Reversed Watson–Crick base pairing

From a geometrical point of view, the simplest way to convert an antiparallel-stranded duplex with Watson–Crick base pairs into a parallel-stranded one is the formal dissociation of one of its component strands into nucleotides, the rotation of each nucleotide by 180° along the long axis of the base pair, and reconnection of the backbone of that strand. This essentially reverts the Watson–Crick base pairs to give reversed Watson–Crick base pairs [35-37]. As can be seen in Scheme 1a, the resulting A:T base pair contains two hydrogen bonds and hence can be expected to be of similar stability as is canonical counterpart. In contrast, application of the above-mentioned formalism to a Watson–Crick G:C pair leads to a base pair comprising one hydrogen bond only and in addition a destabilizing steric clash between two opposing amino groups (Scheme 1b). As a result, most reports on parallel-stranded DNA involving reversed Watson–Crick base pairs focus on A:T rich-sequences. The presence of interspersed G:C base pairs within a duplex strongly destabilizes its structure [38]. Interestingly, a slight displacement of one of the bases in a G:C pair along the short axis could enable the formation of a more stable base pair with two hydrogen bonds (Scheme 1c) [39], albeit at the cost of a backbone distortion due to the displaced positions of the glycosidic bonds. Hence, when contiguous stretches of G:C base pairs are present in a parallel-stranded duplex, thereby reducing the effect of a local backbone distortion, they are much less destabi-

lizing. An elegant way to circumvent the low stability of a G:C base pair in parallel-stranded DNA is the use of isoguanine (^{i}G) or 5-methylisocytosine (^{i}C) to form ^{i}G:C or G:^{i}C base pairs with three hydrogen bonds each (Scheme 1d) [40,41].

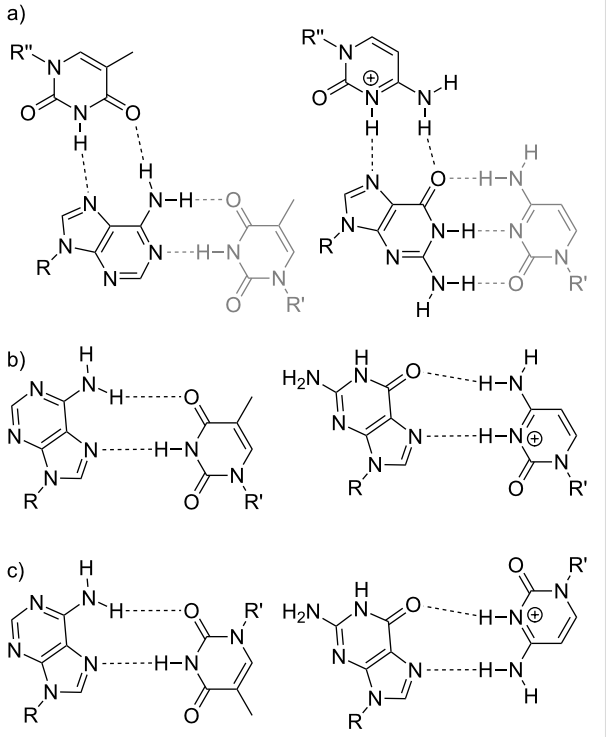


Scheme 1: Hydrogen-bonding patterns of various reversed Watson–Crick base pairs. a) A:T; b) G:C (steric clash between opposing amino groups indicated by red hemicycles); c) G:C in an alternative geometry; d) ^{i}G:C. R, R' = DNA backbone.

Hoogsteen and reversed Hoogsteen base pairing

Hoogsteen and reversed Hoogsteen base pairing can commonly be found in triple helices. The triplex most relevant in the context of this review is the pyrimidine:purine:pyrimidine triplex, where each base triple formally comprises a regular Watson–Crick base pair and an additional pyrimidine residue hydrogen-bonded to the central purine moiety via its Hoogsteen edge. Conceptually, this Hoogsteen-bonded part of the triplex represents a duplex of its own. Scheme 2a indicates how the A:T and G:CH^{+} Hoogsteen base pairs are formally derived from the respective base triples. As can be seen, the cytosine residue needs to be protonated to engage in this hydrogen-bonding pattern. Based on the pK_a value of a cytosine residue within an oligonucleotide single strand of about 4.3 [42], it can be anticipated that this base pair is ideally stabilized under slightly acidic conditions. However, triple helices including a protonated cytosine are stable under physiological conditions, too. In this context, the apparent pK_a value of a cytosine moiety within a CH^{+}:G:C triple in a triplex was reported to amount to 6.7 [43]. Hence, while preferring slightly acidic conditions, Hoogsteen base pairs may also be stable at near-neutral pH values.

When considering a duplex comprising Hoogsteen-type base pairs, the correlation between the relative orientation of the glycosidic bonds (cisoid vs transoid) and the relative orientation of the oligonucleotide strands (parallel vs antiparallel) is rather complex. As can be derived from several calculated or experimental duplex and triplex structures, a parallel strand orientation is adopted for cisoid Hoogsteen base pairing (Scheme 2b) when both nucleotides involved in the base pair adopt an identical glycosidic bond conformation, i.e., when both are oriented *anti* or both are oriented *syn* [44-46]. If they adopt opposing glycosidic bond conformations, an antiparallel strand orientation results [46-51]. The opposite is found for the transoid reversed Hoogsteen base pairing (Scheme 2c). Here, an identical glycosidic bond formation correlates with an antiparallel strand orientation [46,50,51], whereas a parallel arrangement of the strands results from opposing glycosidic bond conformations [46]. It needs to be noted that these correlations are derived from base pairs and triples comprising canonical purine and pyrimidine nucleobases only. In particular, it is assumed that both Hoogsteen and reversed Hoogsteen pairing involve the Hoogsteen edge of one purine residue and the Watson–Crick edge of the complementary pyrimidine or purine moiety. Artificial base pairs involving two purine entities facing each other via their respective Hoogsteen edge (vide infra, Scheme 8b) need to be treated differently, as this additional structural change leads to a change from parallel-stranded to antiparallel-stranded (and vice versa) in the above-made correlations. These general considerations on how the type of hydrogen-bonding

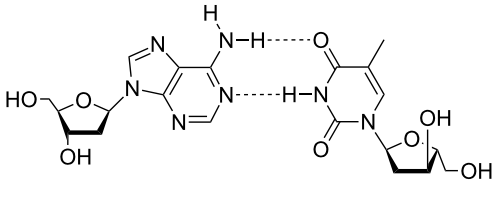


Scheme 2: Various hydrogen-bonding patterns. a) T:A:T and CH^{+}:G:C base triples indicating how T:A and CH^{+}:G Hoogsteen-type base pairs are formally derived from a pyrimidine:purine:pyrimidine base triple. The pyrimidine residue involved in the Watson–Crick pairing within each triple is indicated in grey. b) Hoogsteen-type A:T and G:CH^{+} base pairs. c) Reversed Hoogsteen-type A:T and G:CH^{+} base pairs. R, R', R'' = DNA backbone.

pattern, the orientation of the glycosidic bonds (*syn* vs *anti*) and their relative position (cisoid vs transoid) correlates with the relative strand direction of the oligonucleotide chains are also known as Westhof's rule [52].

Chimeric base pairs of α- and β-deoxyribonucleosides

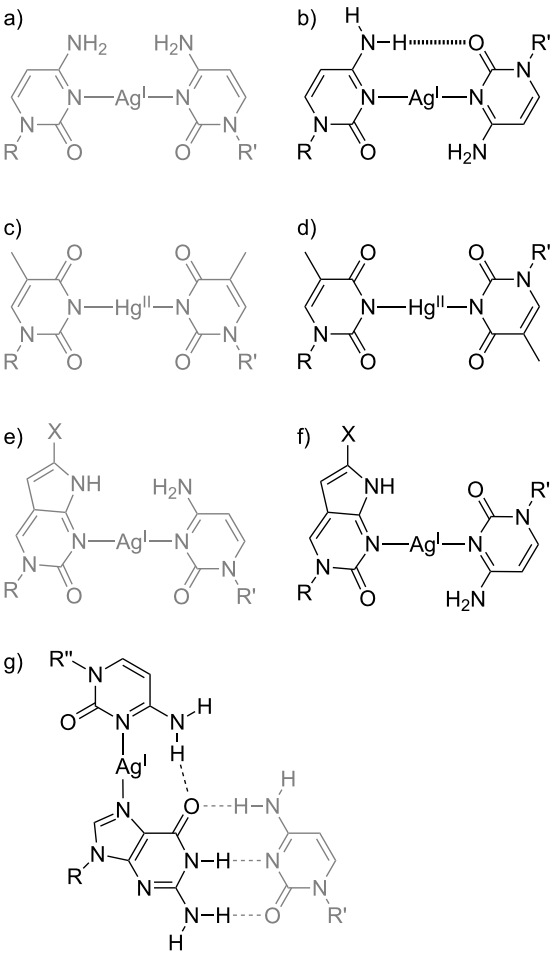
Another possibility to create parallel-stranded duplexes is the use of α-anomeric nucleic acids. These oligonucleotides are formally derived via an inversion of the configuration at the C1' position of the deoxyribonucleoside. When pairing an oligonucleotide comprising α-deoxyribonucleosides with an oligonucleotide consisting of canonical β-deoxyribonucleosides, a parallel-stranded duplex is formed [53]. Due to the reversal of the strand polarity and the concomitant inversion of the configuration, this duplex contains cisoid Watson–Crick base pairs (Scheme 3) [54,55]. Transoid reversed Watson–Crick base pairs can be obtained by introducing individual α-deoxyribonucleosides into a regular antiparallel-stranded duplex of β-deoxyribonucleosides [56].



Scheme 3: Representation of a β-dA:α-dT base pair.

Mononuclear metal-mediated base pairs

One of the first metal-mediated base pairs investigated both in parallel- and antiparallel-stranded DNA duplexes is the C–Ag(I)–C pair. The geometry of this base pair within regular B-DNA is depicted in Scheme 4a. It has been unambiguously proven by experimental structure determinations [10,25]. A comparison of this cisoid base pair with its transoid counterpart (Scheme 4b) suggests that the latter geometry may be additionally stabilized by a synergistic hydrogen bond. Indeed, computations indicate that the transoid base pair is favoured by 7.6 kcal mol^{-1} in the gas phase [57]. It was found to be slightly asymmetric with an N–Ag(I)–N angle of 161.7°. This asymmetry contrasts that of the symmetric hemiprotonated CH^+:C base pair known from i-motif structures and is the result of the larger size of the Ag(I) ion compared with a proton. Hence, only one rather than two hydrogen bonds is formed. This was corroborated by a different theoretical study for a solvated (aquated) transoid C–Ag(I)–C base pair [58]. Experimentally, formation of a transoid C–Ag(I)–C base pair was achieved in two independent manners. In the first report, base pairing of the surrounding canonical base pairs in a reversed Watson–Crick pattern was achieved by covalently linking the complementary strands, fixing them in a parallel-stranded fashion [59]. The second example for a transoid C–Ag(I)–C base pair involves the use of an α-deoxycytidine residue introduced into a regular B-DNA duplex [60]. Table 1 lists the melting temperatures of the respective duplexes. Interestingly, the transoid C–Ag(I)–C base pair was found to exert a larger stabilizing effect than the corresponding cisoid C–Ag(I)–C pair in the latter example only. It is tempting to speculate that the covalent linkage used in the first study additionally influences the duplex stability, leading to a decreased stabilizing effect of the metal-mediated base pair. This hypothesis is corroborated by the fact that the transoid T–Hg(II)–T base pair (Scheme 4d) is likewise less stabilizing than the corresponding cisoid pair (Scheme 4c) when covalently linked duplexes are considered (Table 1) [59].



Scheme 4: Representation of mononuclear metal-mediated base pairs involving canonical pyrimidine nucleobases. Base pairs within an antiparallel-stranded context are displayed in grey, whereas base pairs in a parallel-stranded sequence alignment are shown in black.
a, b) C–Ag(I)–C [59,60]; c, d) T–Hg(II)–T [59]; e) ^{X}PC–Ag(I)–C (X = CH_3, 2-pyridyl, 3-pyridyl) [61,62]; f) ^{X}PC–Ag(I)–C (X = 2-pyridyl, 3-pyridyl) [62]; g) C–Ag(I)–G:C base triple [63]. R, R', R'' = DNA backbone.

Table 1: Increase in melting temperature ΔT_m of duplexes bearing C–Ag(I)–C or T–Hg(II)–T base pairs upon formation of that metal-mediated base pair.

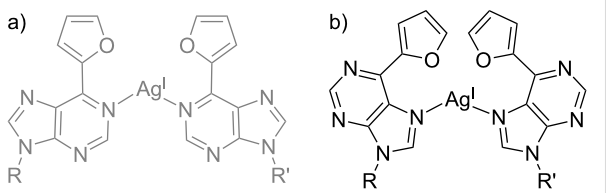
Base pair	ΔT_m [°C]		Nucleic acid	Type of stabilization	Ref.
	Cisoid	Transoid			
C–Ag(I)–C	+13 °C	+8 °C	DNA	covalently linked duplex	[59]
C–Ag(I)–C	+7.5 °C	+15.0 °C	DNA	α-deoxyribonucleoside	[60]
C–Ag(I)–C	+5.5 °C	+12.0 °C	DNA/RNA hybrid	α-deoxyribonucleoside	[60]
T–Hg(II)–T	+9 °C	+6 °C	DNA	covalently linked duplex	[59]

Without the constraint of being incorporated in a nucleic acid duplex comprising Watson–Crick or reversed Watson–Crick base pairs [64], additional geometries may be adopted by the C–Ag(I)–C base pair. A recent computational study on a duplex bearing Ag(I)-mediated base pairs formed from a d(CC) dinucleotide indicates significantly tilted nucleobases, leading to a conformation in-between cisoid and transoid [65]. Such an arrangement was later found in the crystal structure of an Ag(I) complex of the model nucleobase 1-hexylcytosine as well as in a non-canonical DNA structure [9,66].

Pyrrolocytosine (PC) represents a fluorescent analogue of cytosine that still retains the base pairing properties of its parent nucleobase [67]. Accordingly, its application in metal-mediated base pairing was probed, too. An initial report on the ^{Me}PC–Ag(I)–C base pair (Scheme 4e) within regular B-DNA did not include any data on the stabilizing effect of the Ag(I) ion coordination but unequivocally confirmed metal-mediated base pair formation via the quenching of the intrinsic fluorescence of ^{Me}PC [61]. The ^{2Pyr}PC–Ag(I)–C and ^{3Pyr}PC–Ag(I)–C base pairs (Scheme 4e,f) were investigated both in antiparallel-stranded and in parallel-stranded DNA. For both base pairs, the increase in melting temperature T_m upon formation of the metal-mediated base pair in parallel-stranded DNA slightly exceeded that observed for the antiparallel-stranded duplex ($\Delta T_m \geq 6$ °C vs $\Delta T_m = 5.5$ °C) [62]. Even though this difference is not significant, it may be assumed that it is the result of one synergistic hydrogen bond, just like in the case of C–Ag(I)–C. In this study, the parallel-stranded alignment of the duplex was achieved by enforcing reversed Watson–Crick base pairs via the use of G:^{i}C and ^{i}G:C base pairs.

One metal-modified nucleic acid has been reported with a C–Ag(I)–G pair in which the Ag(I) ion binds to the guanine residue via its Hoogsteen edge [63]. As mentioned above, Hoogsteen-type duplexes may be considered an excerpt from a pyrimidine:purine:pyrimidine triplex. In fact, the reported C–Ag(I)–G pair was essentially a component of a C–Ag(I)–G:C base triple within a triple helix, in which the proton of a CH^+:G:C triple was formally replaced by an Ag(I) ion [63].

In addition to the canonical pyrimidine nucleobases such as cytosine and thymine or pyrrolocytosine as a derivative thereof, 6-furylpurine (FP) was reported as an artificial purine derivative for metal-mediated base pairing. When introduced into a regular antiparallel-stranded sequence context, the thermal stabilization upon incorporation of Ag(I) was rather low (ΔT_m = 2 °C) and could not be unequivocally distinguished from unspecific binding to the canonical nucleobases [68]. However, when the FP–Ag(I)–FP base pair was incorporated into a parallel-stranded DNA duplex of the same sequence, a significant stabilization of almost 15 °C was observed [69]. In this study, a parallel-stranded orientation of the duplex was achieved by enforcing Hoogsteen base pairing via the selection of a low pH of 5.5. The strong preference for a parallel strand alignment was explained by comparing the proposed base pairing patterns for antiparallel-stranded DNA (Scheme 5a) and parallel-stranded DNA (Scheme 5b) [69]. While the relative location of the glycosidic bonds shows an enormous discrepancy between the Ag(I)-mediated Watson–Crick pair and its surrounding canonical base pairs (Δ > 2.7 Å), a perfect match was found for the Hoogsteen geometry (Δ = 0.01 Å). Hence, despite the fact that the Ag(I)-mediated Watson–Crick pair is more stable than the Ag(I)-mediated Hoogsteen pair by 15.3 kcal mol^{-1}, the Ag(I)-mediated Hoogsteen base pair displays a very favourable geom-



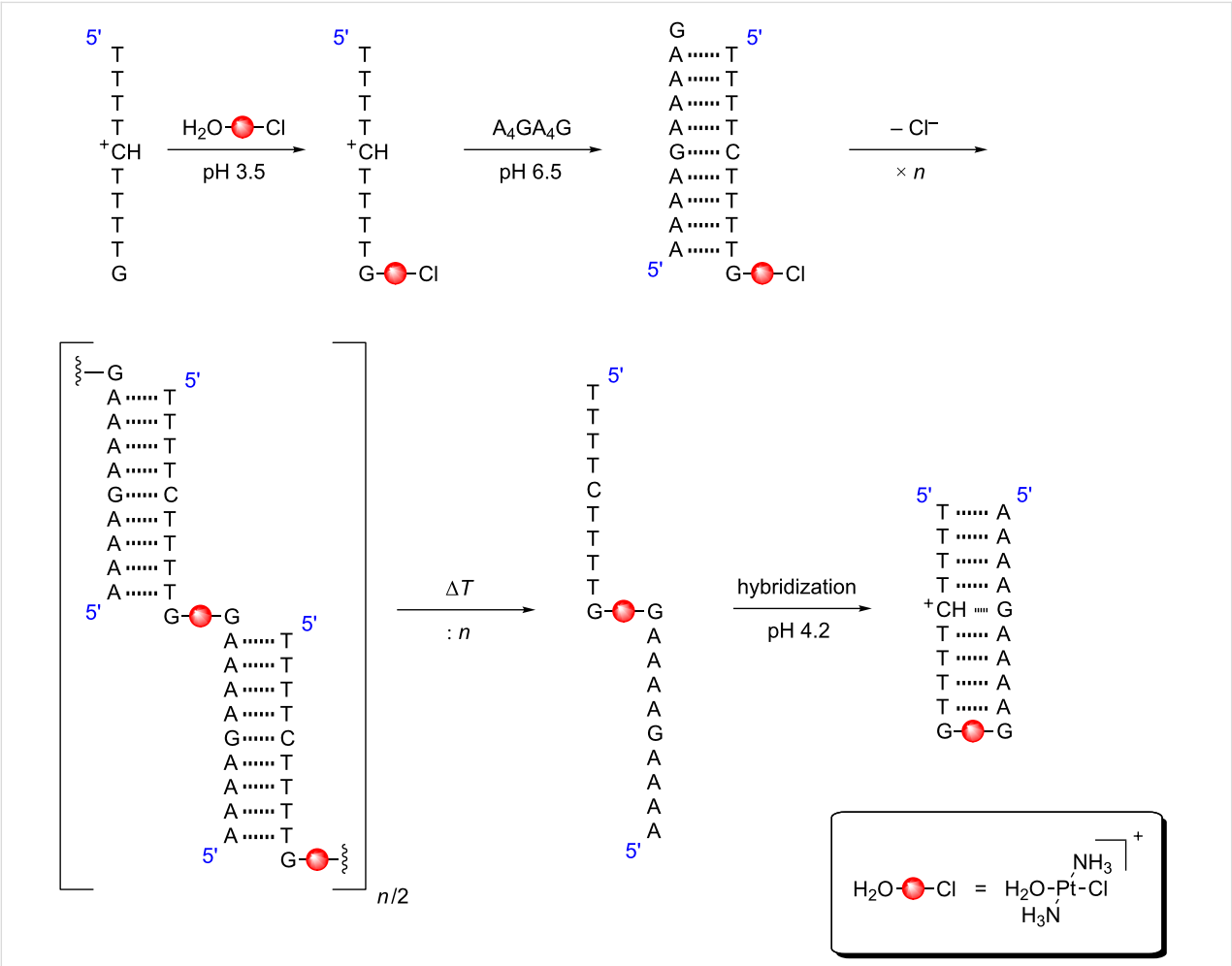
Scheme 5: Proposed base pairing patterns of FP–Ag(I)–FP, involving a) the Watson–Crick edge (antiparallel-stranded) and b) the Hoogsteen edge (parallel-stranded) of the purine derivative [68,69]. R, R' = DNA backbone.

etry and perfectly fits the steric requirements of the parallel-stranded duplex geometry.

Finally, a Pt(II)-mediated base pair has been reported in which a G–Pt(II)–G crosslink enforces a parallel strand orientation [70]. The preparation of this base pair was very distinct from the procedure commonly applied for the generation of Ag(I)- or Hg(II)-mediated base pairs [11], because Pt(II) reacts under kinetic control and has a high affinity for all canonical nucleobases [71]. As shown in Scheme 6, a single-stranded pyrimidine sequence with a terminal guanine residue was initially platinated with the monoaqua species of transplatin, i.e., with *trans*-$[PtCl(NH_3)_2(OH_2)]^+$. As a result of the reaction conditions (pH 3.5), the platination selectively took place at the sole guanine residue. In the second step, a complementary oligonucleotide was added to form an antiparallel-stranded duplex with a dangling guanine moiety at each 3' terminus, one of them being monofunctionally platinated. As all other nucleobases were involved in base pairing, the next platination proceeded after slow dissociation of the remaining chlorido ligand in an intermolecular fashion. Once this slow reaction was completed, brief heating to interrupt the hydrogen-bonded base pairs and subsequent slow annealing under slightly acidic conditions favoured the formation of a parallel-stranded duplex with Hoogsteen base pairing and one G–Pt(II)–G base pair.

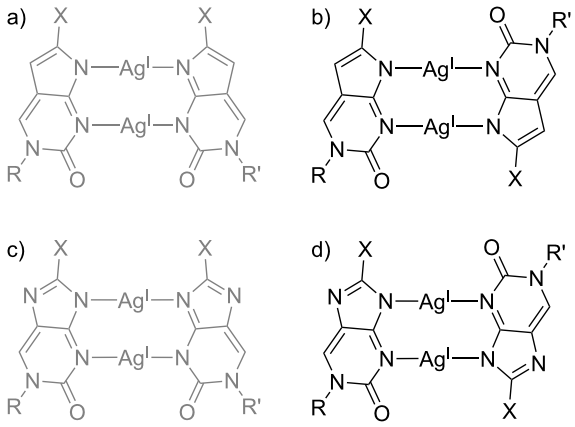
Dinuclear metal-mediated base pairs

In addition to engaging in base pairing with a complementary cytosine residue (vide supra), pyrrolocytosine derivatives were also investigated with respect to their propensity to form metal-mediated pyrrolocytosine:pyrrolocytosine base pairs [62]. These investigations not only included the ^{2Pyr}PC and ^{3Pyr}PC residues reported above, but also a ^{Ph}PC nucleoside bearing a phenyl substituent. Accordingly, a possible additional stabilization due to the presence of the (potentially coordinating) endocyclic nitrogen atom of the pyridine substituent was investigated. A series of 12-mer duplexes and 25-mer duplexes were studied. In all cases, dinuclear Ag(I)-mediated base pairs formed



Scheme 6: Reaction pathway towards a parallel-stranded DNA duplex bearing a G–Pt(II)–G base pair. Adapted from reference [70].

(Scheme 7). Table 2 lists the thermal stabilization upon formation of the respective metal-mediated base pairs. As can be seen, base pairs including at least one ^{2Pyr}PC moiety display the largest stabilization, which points towards an involvement of the endocyclic pyridyl nitrogen atom in metal coordination particularly for the parallel-stranded duplexes. For the shorter 12-mer duplexes, the stabilizing effect of metal-mediated base pair formation found in antiparallel-stranded DNA exceeds that observed in parallel-stranded DNA. For the longer 25-mer duplexes, the opposite is true, indicating the relevance of the sequence context on the observed stabilization. For all ^{X}PC-derived base pairs, reversed Watson–Crick base pairing was enforced to ensure the formation of parallel-stranded duplexes.



Scheme 7: Possible base pairing patterns of dinuclear Ag(I)-mediated base pairs. Base pairs within an antiparallel-stranded context are displayed in grey, whereas base pairs in a parallel-stranded sequence alignment are shown in black. a, b) ^{X}PC–Ag(I)$_2$–^{X}PC (X = phenyl, 2-pyridyl, 3-pyridyl) [62]; c, d) ^{X}IC–Ag(I)$_2$–^{X}IC (X = H, phenyl, 2-furyl) [72]. R, R' = DNA backbone.

Table 2: Increase in melting temperature ΔT_m of two sets of DNA duplexes bearing a ^{X}PC–Ag(I)$_2$–^{Y}PC base pair (X, Y = phenyl, 2-pyridyl, 3-pyridyl) [62].

X	Y	ΔT_m [°C] for 12-mer duplex[a]		ΔT_m [°C] for 25-mer duplex[b]	
		aps[c]	ps[c]	aps[c]	ps[c]
2Pyr	2Pyr	+21.5	+21.0	+8.5	+13.5
3Pyr	3Pyr	+26.0	+10.0	+5.0	+7.0
Ph	2Pyr	+26.0	+14.5	+7.5	+12.0
Ph	3Pyr	+27.0	+5.5	n.d.	n.d.
3Pyr	2Pyr	+26.5	+19.0	+9.5	+13.0

[a]Parallel-stranded DNA obtained by using ^{i}G:C and G:^{i}C base pairs; [b]Parallel-stranded DNA with A:T base pairs, determined by the sequence only. [c]aps: antiparallel-stranded (i.e., Watson–Crick base pairs, cisoid glycosidic bonds); ps: parallel-stranded (i.e., reversed Watson–Crick base pairs, transoid glycosidic bonds).

Formal replacement of one C–H group in pyrrolocytosine by a nitrogen atom leads to imidazolocytosine. A series of substituted imidazolocytosine (^{X}IC) nucleobases were investigated with respect to their metal-binding properties, too [72]. In analogy to ^{X}PC, ^{X}IC forms dinuclear metal-mediated homo base pairs with Ag(I). Table 3 lists representative changes in the melting temperature upon formation of an ^{X}IC–Ag(I)$_2$–^{X}IC base pair within a DNA duplex. As can be seen, these base pairs are extremely stabilizing both in antiparallel-stranded and in parallel-stranded duplexes. In fact, the ^{2Fur}IC–Ag(I)$_2$–^{2Fur}IC pair represents the most stabilizing Ag(I)-mediated base pair reported to date. The trend in stabilization (^{H}IC ≈ ^{Ph}IC < ^{2Fur}IC) allows two different explanations. As the furyl substituent is the only one with an potential donor atom [73], the extraordinary stability of the ^{2Fur}IC–Ag(I)$_2$–^{2Fur}IC base pair may be the direct result of the formation of additional coordinate bonds. Alternatively (or in addition), the deprotonation of ^{X}IC, which is a prerequisite for Ag(I) binding, is facilitated in the order ^{H}IC < ^{Ph}IC < ^{2Fur}IC (pK_a = 8.8, 7.9, and 7.3 for the respective nucleosides), which is identical to the trend in stabilization [72]. For all ^{X}IC-derived base pairs, parallel-stranded duplexes were obtained by enforcing reversed Watson–Crick base pairing.

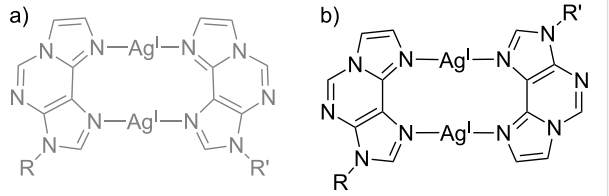
Table 3: Increase in melting temperature ΔT_m of a representative DNA duplex bearing one ^{X}IC–Ag(I)$_2$–^{X}IC base pair (X = H, phenyl, 2-furyl) [72].

Base pair	ΔT_m [°C]	
	aps[a]	ps[a]
^{H}IC–Ag(I)$_2$–^{H}IC	+39.0	+27.0
^{Ph}IC–Ag(I)$_2$–^{Ph}IC	+38.5	+27.0
^{2Fur}IC–Ag(I)$_2$–^{2Fur}IC	+48.0	+38.0

[a]aps: antiparallel-stranded (i.e., Watson–Crick base pairs, cisoid glycosidic bonds); ps: parallel-stranded (i.e., reversed Watson–Crick base pairs, transoid glycosidic bonds).

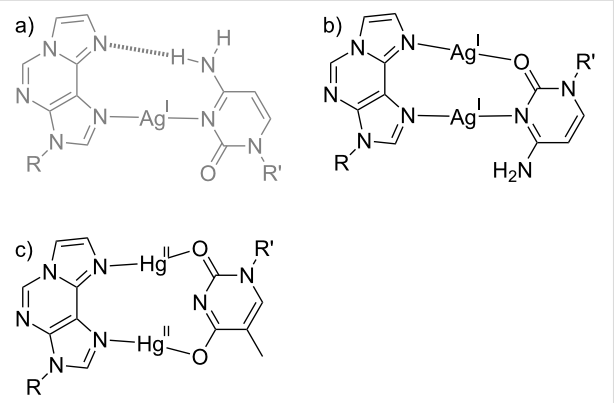
1,N^6-Ethenoadenine (εA) is an exocyclic etheno adduct of adenine which was shown to bind transition metal ions better than its parent nucleobase [74]. Accordingly, its propensity to engage in metal-mediated base pairing was investigated in detail. As it turned out, εA is capable of simultaneously binding two metal ions with an almost parallel alignment of the N→M bonds [75]. In principle, both a cisoid and a transoid arrangement of the glycosidic bonds are feasible (Scheme 8). The former is adopted when the εA–Ag(I)$_2$–εA base pair is incorporated in-between canonical Watson–Crick base pairs in a B-DNA duplex [76]. The stabilization observed upon formation of this cisoid εA–Ag(I)$_2$–εA pair amounts to 12 °C. Interestingly, when the dinuclear $[Ag_2(εA)_2]^{2+}$ complex is formed outside a DNA context, e.g., in a crystal structure using the

model nucleobase 9-ethyl-1,N^6-ethenoadenine or adsorbed onto HOPG applying 9-icosyl-1,N^6-ethenoadenine (HOPG: highly-ordered pyrolytic graphite), the transoid conformation is preferred [75]. Accordingly, the transoid εA–Ag(I)$_2$–εA base pair incorporated in-between reversed Hoogsteen base pairs in a parallel-stranded duplex brings about a stabilization of ≈16 °C, exceeding that of the cisoid pair. It should be noted that Hoogsteen (rather than reversed Hoogsteen) base pairing could not be ruled out completely for this parallel-stranded duplex, so that in principle a cisoid εA–Ag(I)$_2$–εA base pair may also form when the complementary oligonucleotides are aligned in a parallel fashion. However, considering the intrinsic preference of εA–Ag(I)$_2$–εA to adopt a transoid geometry and in line with the larger stabilization in a parallel-stranded duplex context, reversed Hoogsteen (Scheme 8b) represents the most likely base pairing pattern in this example [75].



Scheme 8: Base pairing patterns of the dinuclear Ag(I)-mediated homo base pair of 1,N^6-ethenoadenine (εA) with a) cisoid arrangement of the glycosidic bonds (i.e., in an antiparallel-stranded duplex) [76] and b) transoid arrangement of the glycosidic bonds (i.e., in a parallel-stranded duplex) [75]. R, R' = DNA backbone.

Its property to align the N→M vectors in an almost parallel manner has led to the application of εA in a series of other metal-mediated base pairs. When locating a cytosine residue opposite εA, a fascinating influence of the relative strand orientation of the DNA duplex on the number of metal ions per base pair was observed [77]. When the εA:C pair is present inside an antiparallel-stranded duplex, it incorporates one Ag(I). The formation of the resulting εA–Ag(I)–C base pair (Scheme 9a) is accompanied by an increase in T_m of 15 °C. According to a geometry optimization of the proposed base pair structure, it also contains a synergistic hydrogen bond, as is evident from the non-linear N–Ag–N angle of 167°. Using the same sequence context, albeit with a parallel alignment of the strands with reversed Hoogsteen base pairing, a dinuclear εA–Ag(I)$_2$–C pair is formed (Scheme 9b). The Ag···Ag distance within this base pair was calculated as 2.92 Å [77], suggesting the presence of a stabilizing argentophilic interaction [78]. The parallel-stranded duplex bearing an εA:C pair is stabilized by ≈20 °C upon incorporation of the two Ag(I) ions. This stabilization is not twice as large as that observed for the mononuclear base pair, indicating that the introduction of the second Ag(I) ion is only slightly more stabilizing than the synergistic hydrogen bond.



Scheme 9: Additional metal-mediated base pairs involving 1,N^6-ethenoadenine (εA). Base pairs within an antiparallel-stranded context are displayed in grey, whereas base pairs in a parallel-stranded sequence alignment are shown in black. a) εA–Ag(I)–C with a synergistic hydrogen bond [77]; b) εA–Ag(I)$_2$–C [77]; c) εA–Hg(II)$_2$–T [79,80]. R, R' = DNA backbone.

When a thymine residue is paired with εA in a parallel-stranded double helix with reversed Watson–Crick base pairs, then the resulting εA:T pair incorporates two Hg(II) ions, yielding εA–Hg(II)$_2$–T as the first example of a dinuclear metal-mediated base pair bearing divalent metal ions [79]. In the proposed base pair structure, both Hg(II) ions are coordinated by an endocyclic nitrogen atom of εA and an exocyclic oxygen atom of the thymine residue [80]. This structure (Scheme 9c) differs slightly from the originally proposed one containing one additional bond from the endocyclic nitrogen atom of thymine to one of the Hg(II) ions, because a calculation of the Hg···N force constant [81] had resulted in an exceptionally low value of 0.7 N cm^{-1} [80]. A re-inspection of the originally proposed structure indicated that it represents a local energy minimum rather than the global one. In the structure shown in Scheme 9c, all Hg···N and Hg···O force constants amount to ≈2 N cm^{-1} and hence indicate strong bonds [80]. The fourfold positive charge of the two Hg(II) ions in the εA–Hg(II)$_2$–T base pair is stabilized by three factors. First of all, the thymine residue is deprotonated upon coordination to Hg(II), as is also the case for the well-established T–Hg(II)–T base pair. Hence, the negative charge of the thyminate helps shielding the positive charge introduced by the metal ions. Second, the propensity of εA to bind two metal ions with the N→M bonds aligned in parallel brings together the Hg(II) ions at close distance. Third, the εA–Hg(II)$_2$–T base pair appears to be formed in a parallel-stranded duplex only. Attempts to introduce it into an antiparallel sequence context were unsuccessful so far. The reason for this is unknown yet, but may be related to the transoid orientation of the glycosidic bonds in parallel-stranded DNA. The combination of these three effects thus allowed the formation of the first dinuclear metal-mediated base pair with two divalent metal ions.

Conclusion

This review summarizes recent efforts to extend the principle of metal-mediated base pairing to parallel-stranded nucleic acid duplexes. It indicates the many experimental possibilities to enforce a parallel strand alignment. Depending on the requirements of the metal-mediated base pair to be formed, different strategies (e.g., using different pH values) can be followed. The exhaustive list of examples presented in this review allows drawing some general conclusions. The most important one probably is that in many cases it is not possible to predict a priori whether a metal-mediated base pair is more stabilizing in an antiparallel-stranded or a parallel-stranded duplex. The intrinsic stability of the metal-free duplex, its sequence and the method used to enforce a parallel orientation of the complementary strands play important roles, too. This becomes evident for example for the ^{X}PC–Ag(I)$_2$–^{Y}PC base pairs, which were found to be more stabilizing in an antiparallel-stranded duplex for 12-mer oligonucleotides, whereas longer 25-mers showed a larger stabilization in a parallel-stranded orientation. Nonetheless, the use of parallel-stranded duplexes significantly extents the scope of metal-mediated base pairing, because it has been shown that artificial nucleobases can be designed in a way that they form metal-mediated base pairs that are more stabilizing in a parallel-stranded context than in an antiparallel-stranded one (e.g., FP–Ag(I)–FP, εA–Ag(I)$_2$–C, ^{2Fur}IC–Ag(I)$_2$–^{2Fur}IC). It is therefore beyond doubt that parallel-stranded DNA will find an important place in research on metal-mediated base pairs, in particular when the metal complex prefers a C_2-symmetric geometry and hence a transoid orientation of the glycosidic bonds.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for its support of our research on metal-mediated base pairing, Dr. Soham Mandal for his help in creating Figure 1, and Prof. Dr. Jörg Grunenberg for his comments regarding the structure of the εA–Hg(II)$_2$–T base pair.

ORCID® iDs

Jens Müller - http://orcid.org/0000-0003-4713-0606

References

1. Bandy, T. J.; Brewer, A.; Burns, J. R.; Marth, G.; Nguyen, T.; Stulz, E. *Chem. Soc. Rev.* **2011,** *40,* 138–148. doi:10.1039/B820255A
2. Ensslen, P.; Wagenknecht, H.-A. *Acc. Chem. Res.* **2015,** *48,* 2724–2733. doi:10.1021/acs.accounts.5b00314
3. Copp, S. M.; Schultz, D. E.; Swasey, S.; Gwinn, E. G. *ACS Nano* **2015,** *9,* 2303–2310. doi:10.1021/nn506322q
4. Gramlich, P. M. E.; Wirges, C. T.; Manetto, A.; Carell, T. *Angew. Chem., Int. Ed.* **2008,** *47,* 8350–8358. doi:10.1002/anie.200802077
5. Tanaka, Y.; Kondo, J.; Sychrovský, V.; Šebera, J.; Dairaku, T.; Saneyoshi, H.; Urata, H.; Torigoe, H.; Ono, A. *Chem. Commun.* **2015,** *51,* 17343–17360. doi:10.1039/C5CC02693H
6. Takezawa, Y.; Müller, J.; Shionoya, M. *Chem. Lett.* **2017,** *46,* 622–633. doi:10.1246/cl.160985
7. Clever, G. H.; Kaul, C.; Carell, T. *Angew. Chem., Int. Ed.* **2007,** *46,* 6226–6236. doi:10.1002/anie.200701185
8. Mandal, S.; Müller, J. *Curr. Opin. Chem. Biol.* **2017,** *37,* 71–79. doi:10.1016/j.cbpa.2017.01.019
9. Liu, H.; Shen, F.; Haruehanroengra, P.; Yao, Q.; Cheng, Y.; Chen, Y.; Yang, C.; Zhang, J.; Wu, B.; Luo, Q.; Cui, R.; Li, J.; Ma, J.; Sheng, J.; Gan, J. *Angew. Chem., Int. Ed.* **2017,** *56,* 9430–9434. doi:10.1002/anie.201704891
10. Kondo, J.; Tada, Y.; Dairaku, T.; Hattori, Y.; Saneyoshi, H.; Ono, A.; Tanaka, Y. *Nat. Chem.* **2017,** *9,* 956–960. doi:10.1038/nchem.2808
11. Jash, B.; Müller, J. *Chem. – Eur. J.* **2017,** *23,* 17166–17178. doi:10.1002/chem.201703518
12. Hensel, S.; Eckey, K.; Scharf, P.; Megger, N.; Karst, U.; Müller, J. *Chem. – Eur. J.* **2017,** *23,* 10244–10248. doi:10.1002/chem.201702241
13. Ehrenschwender, T.; Schmucker, W.; Wellner, C.; Augenstein, T.; Carl, P.; Harmer, J.; Breher, F.; Wagenknecht, H.-A. *Chem. – Eur. J.* **2013,** *19,* 12547–12552. doi:10.1002/chem.201300593
14. Liu, S.; Clever, G. H.; Takezawa, Y.; Kaneko, M.; Tanaka, K.; Guo, X.; Shionoya, M. *Angew. Chem., Int. Ed.* **2011,** *50,* 8886–8890. doi:10.1002/anie.201102980
15. Jash, B.; Müller, J. *Eur. J. Inorg. Chem.* **2017,** 3857–3861. doi:10.1002/ejic.201700665
16. Jash, B.; Scharf, P.; Sandmann, N.; Fonseca Guerra, C.; Megger, D. A.; Müller, J. *Chem. Sci.* **2017,** *8,* 1337–1343. doi:10.1039/C6SC03482A
17. Taherpour, S.; Golubev, O.; Lönnberg, T. *Inorg. Chim. Acta* **2016,** *452,* 43–49. doi:10.1016/j.ica.2016.01.025
18. Swasey, S. M.; Gwinn, E. G. *New J. Phys.* **2016,** *18,* 045008. doi:10.1088/1367-2630/18/4/045008
19. Takezawa, Y.; Yoneda, S.; Duprey, J.-L. H. A.; Nakama, T.; Shionoya, M. *Chem. Sci.* **2016,** *7,* 3006–3010. doi:10.1039/C6SC00383D
20. Kobayashi, T.; Takezawa, Y.; Sakamoto, A.; Shionoya, M. *Chem. Commun.* **2016,** *52,* 3762–3765. doi:10.1039/C5CC10039A
21. Funai, T.; Nakamura, J.; Miyazaki, Y.; Kiriu, R.; Nakagawa, O.; Wada, S.-i.; Ono, A.; Urata, H. *Angew. Chem., Int. Ed.* **2014,** *53,* 6624–6627. doi:10.1002/anie.201311235
22. Kim, E.-K.; Switzer, C. *ChemBioChem* **2013,** *14,* 2403–2407. doi:10.1002/cbic.201300634
23. Kaul, C.; Müller, M.; Wagner, M.; Schneider, S.; Carell, T. *Nat. Chem.* **2011,** *3,* 794–800. doi:10.1038/nchem.1117
24. Park, K. S.; Jung, C.; Park, H. G. *Angew. Chem., Int. Ed.* **2010,** *49,* 9757–9760. doi:10.1002/anie.201004406
25. Dairaku, T.; Furuita, K.; Sato, H.; Šebera, J.; Nakashima, K.; Kondo, J.; Yamanaka, D.; Kondo, Y.; Okamoto, I.; Ono, A.; Sychrovský, V.; Kojima, C.; Tanaka, Y. *Chem. – Eur. J.* **2016,** *22,* 13028–13031. doi:10.1002/chem.201603048
26. Takezawa, Y.; Shionoya, M.; Müller, J. In *Comprehensive Supramolecular Chemistry II;* Atwood, J. L., Ed.; Elsevier Ltd.: Oxford, 2017; Vol. 4, pp 259–293. doi:10.1016/B978-0-12-409547-2.12556-9
27. Takezawa, Y.; Maeda, W.; Tanaka, K.; Shionoya, M. *Angew. Chem., Int. Ed.* **2009,** *48,* 1081–1084. doi:10.1002/anie.200804654
28. Tanaka, K.; Yamada, Y.; Shionoya, M. *J. Am. Chem. Soc.* **2002,** *124,* 8802–8803. doi:10.1021/ja020510o

29. Engelhard, D. M.; Nowack, J.; Clever, G. H. *Angew. Chem., Int. Ed.* **2017,** *56,* 11640–11644. doi:10.1002/anie.201705724
30. Engelhard, D. M.; Pievo, R.; Clever, G. H. *Angew. Chem., Int. Ed.* **2013,** *52,* 12843–12847. doi:10.1002/anie.201307594
31. Wang, E.; Feigon, J. In *Oxford Handbook of Nucleic Acid Structure;* Neidle, S., Ed.; Oxford University Press: Oxford, UK, 1999; pp 355–388.
32. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006,** *34,* 5402–5415. doi:10.1093/nar/gkl655
33. Germann, M. W.; Zhou, N.; van de Sande, J. H.; Vogel, H. J. *Methods Enzymol.* **1995,** *261,* 207–225. doi:10.1016/S0076-6879(95)61011-1
34. Pallan, P. S.; Lubini, P.; Bolli, M.; Egli, M. *Nucleic Acids Res.* **2007,** *35,* 6611–6624. doi:10.1093/nar/gkm612
35. Pattabiraman, N. *Biopolymers* **1986,** *25,* 1603–1606. doi:10.1002/bip.360250903
36. van de Sande, J. H.; Ramsing, N. B.; Germann, M. W.; Elhorst, W.; Kalisch, B. W.; von Kitzing, E.; Pon, R. T.; Clegg, R. C.; Jovin, T. M. *Science* **1988,** *241,* 551–557. doi:10.1126/science.3399890
37. Otto, C.; Thomas, G. A.; Rippe, K.; Jovin, T. M.; Peticolas, W. L. *Biochemistry* **1991,** *30,* 3062–3069. doi:10.1021/bi00226a012
38. Rentzeperis, D.; Rippe, K.; Jovin, T. M.; Marky, L. A. *J. Am. Chem. Soc.* **1992,** *114,* 5926–5928. doi:10.1021/ja00041a003
39. Liu, C.-Q.; Shi, X.-F.; Bai, C.-L.; Zhao, J.; Wang, Y. *J. Theor. Biol.* **1997,** *184,* 319–326. doi:10.1006/jtbi.1996.0274
40. Sugiyama, H.; Ikeda, S.; Saito, I. *J. Am. Chem. Soc.* **1996,** *118,* 9994–9995. doi:10.1021/ja961371b
41. Seela, F.; Wei, C. *Helv. Chim. Acta* **1997,** *80,* 73–85. doi:10.1002/hlca.19970800107
42. Domínguez-Martín, A.; Johannsen, S.; Sigel, A.; Operschall, B. P.; Song, B.; Sigel, H.; Okruszek, A.; González-Pérez, J. M.; Niclós-Gutiérrez, J.; Sigel, R. K. O. *Chem. – Eur. J.* **2013,** *19,* 8163–8181. doi:10.1002/chem.201203330
43. Leitner, D.; Schröder, W.; Weisz, K. *J. Am. Chem. Soc.* **1998,** *120,* 7123–7124. doi:10.1021/ja972694q
44. Raghunathan, G.; Miles, H. T.; Sasisekharan, V. *Biopolymers* **1994,** *34,* 1573–1581. doi:10.1002/bip.360341202
45. Liu, K.; Miles, H. T.; Frazier, J.; Sasisekharan, V. *Biochemistry* **1993,** *32,* 11802–11809. doi:10.1021/bi00095a008
46. Cheng, Y.-K.; Pettitt, B. M. *J. Am. Chem. Soc.* **1992,** *114,* 4465–4474. doi:10.1021/ja00038a004
47. Abrescia, N. G. A.; Thompson, A.; Huynh-Dinh, T.; Subirana, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 2806–2811. doi:10.1073/pnas.052675499
48. Abrescia, N. G. A.; González, C.; Gouyette, C.; Subirana, J. A. *Biochemistry* **2004,** *43,* 4092–4100. doi:10.1021/bi0355140
49. Acosta-Reyes, F. J.; Alechaga, E.; Subirana, J. A.; Campos, J. L. *PLoS One* **2015,** *10,* e0120241. doi:10.1371/journal.pone.0120241
50. Beal, P. A.; Dervan, P. B. *Science* **1991,** *251,* 1360–1363. doi:10.1126/science.2003222
51. Durland, R. H.; Kessler, D. J.; Gunnell, S.; Duvic, M.; Pettitt, B. M.; Hogan, M. E. *Biochemistry* **1991,** *30,* 9246–9255. doi:10.1021/bi00102a017
52. Westhof, E. *Nature* **1992,** *358,* 459–460. doi:10.1038/358459b0
53. Morvan, F.; Debart, F.; Vasseur, J.-J. *Chem. Biodiversity* **2010,** *7,* 494–535. doi:10.1002/cbdv.200900220
54. Aramini, J. M.; Kalisch, B. W.; Pon, R. T.; van de Sande, J. H.; Germann, M. W. *Biochemistry* **1996,** *35,* 9355–9365. doi:10.1021/bi960612p
55. Aramini, J. M.; van de Sande, J. H.; Germann, M. W. *Biochemistry* **1997,** *36,* 9715–9725. doi:10.1021/bi9706071
56. Aramini, J. M.; Cleaver, S. H.; Pon, R. T.; Cunningham, R. P.; Germann, M. W. *J. Mol. Biol.* **2004,** *338,* 77–91. doi:10.1016/j.jmb.2004.02.035
57. Megger, D. A.; Fonseca Guerra, C.; Bickelhaupt, F. M.; Müller, J. *J. Inorg. Biochem.* **2011,** *105,* 1398–1404. doi:10.1016/j.jinorgbio.2011.07.005
58. Fortino, M.; Marino, T.; Russo, N. *J. Phys. Chem. A* **2015,** *119,* 5153–5157. doi:10.1021/jp5096739
59. Ono, T.; Yoshida, K.; Saotome, Y.; Sakabe, R.; Okamoto, I.; Ono, A. *Chem. Commun.* **2011,** *47,* 1542–1544. doi:10.1039/C0CC02028A
60. Guo, X.; Seela, F. *Chem. – Eur. J.* **2017,** *23,* 11776–11779. doi:10.1002/chem.201703017
61. Park, K. S.; Lee, J. Y.; Park, H. G. *Chem. Commun.* **2012,** *48,* 4549–4551. doi:10.1039/c2cc17148a
62. Yang, H.; Mei, H.; Seela, F. *Chem. – Eur. J.* **2015,** *21,* 10207–10219. doi:10.1002/chem.201500582
63. Ihara, T.; Ishii, T.; Araki, N.; Wilson, A. W.; Jyo, A. *J. Am. Chem. Soc.* **2009,** *131,* 3826–3827. doi:10.1021/ja809702n
64. Megger, D. A.; Müller, J. *Nucleosides, Nucleotides Nucleic Acids* **2010,** *29,* 27–38. doi:10.1080/15257770903451579
65. Espinosa Leal, L. A.; Karpenko, A.; Swasey, S.; Gwinn, E. G.; Rojas-Cervellera, V.; Rovira, C.; Lopez-Acevedo, O. *J. Phys. Chem. Lett.* **2015,** *6,* 4061–4066. doi:10.1021/acs.jpclett.5b01864
66. Terrón, A.; Moreno-Vachiano, B.; Bauzá, A.; García-Raso, A.; Fiol, J. J.; Barceló-Oliver, M.; Molins, E.; Frontera, A. *Chem. – Eur. J.* **2017,** *23,* 2103–2108. doi:10.1002/chem.201604331
67. Berry, D. A.; Jung, K.-Y.; Wise, D. S.; Sercel, A. D.; Pearson, W. H.; Mackie, H.; Randolph, J. B.; Somers, R. L. *Tetrahedron Lett.* **2004,** *45,* 2457–2461. doi:10.1016/j.tetlet.2004.01.108
68. Sinha, I.; Kösters, J.; Hepp, A.; Müller, J. *Dalton Trans.* **2013,** *42,* 16080–16089. doi:10.1039/c3dt51691a
69. Sinha, I.; Fonseca Guerra, C.; Müller, J. *Angew. Chem., Int. Ed.* **2015,** *54,* 3603–3606. doi:10.1002/anie.201411931
70. Müller, J.; Drumm, M.; Boudvillain, M.; Leng, M.; Sletten, E.; Lippert, B. *J. Biol. Inorg. Chem.* **2000,** *5,* 603–611. doi:10.1007/s007750000143
71. Lippert, B. *Coord. Chem. Rev.* **2000,** *200–202,* 487–516. doi:10.1016/S0010-8545(00)00260-5
72. Jana, S. K.; Guo, X.; Mei, H.; Seela, F. *Chem. Commun.* **2015,** *51,* 17301–17304. doi:10.1039/C5CC06734K
73. Lobana, T. S.; Kumari, P.; Sharma, R.; Castineiras, A.; Butcher, R. J.; Akitsu, T.; Aritake, Y. *Dalton Trans.* **2011,** *40,* 3219–3228. doi:10.1039/c0dt01291b
74. Scheller, K. H.; Sigel, H. *J. Am. Chem. Soc.* **1983,** *105,* 3005–3014. doi:10.1021/ja00348a013
75. Mandal, S.; Wang, C.; Prajapati, R. K.; Kösters, J.; Verma, S.; Chi, L.; Müller, J. *Inorg. Chem.* **2016,** *55,* 7041–7050. doi:10.1021/acs.inorgchem.6b00927
76. Mandal, S.; Hepp, A.; Müller, J. *Dalton Trans.* **2015,** *44,* 3540–3543. doi:10.1039/C4DT02663B
77. Mandal, S.; Hebenbrock, M.; Müller, J. *Chem. – Eur. J.* **2017,** *23,* 5962–5965. doi:10.1002/chem.201605327
78. Schmidbaur, H.; Schier, A. *Angew. Chem., Int. Ed.* **2015,** *54,* 746–784. doi:10.1002/anie.201405936
79. Mandal, S.; Hebenbrock, M.; Müller, J. *Angew. Chem., Int. Ed.* **2016,** *55,* 15520–15523. doi:10.1002/anie.201608354
80. Grunenberg, J. *personal communication.*

81. Grunenberg, J. *Angew. Chem., Int. Ed.* **2017,** *56,* 7288–7291. doi:10.1002/anie.201701790

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Synthetic mRNA capping

Fabian Muttach[‡1], Nils Muthmann[‡1] and Andrea Rentmeister[*1,2]

Review

Open Access

Address:
[1]University of Münster, Department of Chemistry, Institute of Biochemistry, Wilhelm-Klemm-Str. 2, 48149 Münster, Germany and [2]Cells-in-Motion Cluster of Excellence (EXC1003-CiM), University of Münster, Germany

Email:
Andrea Rentmeister[*] - a.rentmeister@uni-muenster.de

* Corresponding author ‡ Equal contributors

Keywords:
cap analogue; cap synthesis; click chemistry; enzymatic capping; methyltransferase; RNA

Beilstein J. Org. Chem. **2017,** *13,* 2819–2832.
doi:10.3762/bjoc.13.274

Received: 14 September 2017
Accepted: 04 December 2017
Published: 20 December 2017

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

Eukaryotic mRNA with its 5′-cap is of central importance for the cell. Many studies involving mRNA require reliable preparation and modification of 5′-capped RNAs. Depending on the length of the desired capped RNA, chemical or enzymatic preparation – or a combination of both – can be advantageous. We review state-of-the art methods and give directions for choosing the appropriate approach. We also discuss the preparation and properties of mRNAs with non-natural caps providing novel features such as improved stability or enhanced translational efficiency.

Introduction

The 5′-cap is a hallmark of eukaryotic mRNA and involved in numerous interactions required for cellular functions. Chemically, the 5′-cap consists of an inverted 7-methylguanosine connected to the rest of the eukaryotic mRNA via a 5′–5′ triphosphate bridge. This so-called cap0 serves as quality control for correct mRNA processing and contributes to stabilization of eukaryotic mRNA [1,2], splicing [3,4], nuclear export [5], initiation of translation [6,7] and mRNA decay [8]. The most important direct interaction partners of the 5′-cap are the cap binding complex (CBC) [9,10] in the nucleus required for nuclear export and the eukaryotic translation initiation factor 4E (eIF4E) [11] in the cytoplasm which is indispensable for cap-dependent translation. Additionally, capped RNA serves as a marker for the innate immune system to distinguish triphosphorylated viral RNAs from cellular RNAs [12]. The antiviral response is among others mediated by the cytosolic receptor RIG-I which is activated by short single and double-stranded triphosphorylated RNAs and MDA-5. MDA-5 recognizes long triphosphorylated RNAs and RNAs lacking the 2′-OH methylation at the first nucleotide (cap1), a modification which is commonly observed in eukaryotes [13-15].

Besides cap0 and cap1, cap structures with further modifications exist. Additional methyl groups are often found at the second nucleotide (cap2) while in trypanosomes up to four methylated nucleotides are observed (termed cap4) [16,17].

Owing to the importance of different cap structures for recognition processes in the cell, it becomes clear that an uncapped transcript does not adequately represent a eukaryotic mRNA

and that preparation of correctly capped RNAs is essential to assess the function of mRNAs in the cellular context. Furthermore, altering the cap structure bears potential to increase mRNA stability and translational efficiency – two properties which may provide the key to therapeutic applications of mRNA in the near future [18-20]. Finally, investigations of structure and mechanism of 5′-cap/protein interactions are still hampered by the difficulty of producing large quantities of homogenously capped RNA.

In this review article, we present different synthetic routes to 5′-capped mRNAs based on enzymatic, chemical or chemoenzymatic methods. We will point out the difficulties and limitations of the different strategies and – if available – will show ways to circumvent them. This review focuses strictly on mRNA cap analogues (and some non-natural modifications); for preparation of other capped biomolecules such as capped siRNAs [21], peptidyl capped oligonucleotides [22], NAD-capped RNAs [23,24], 3'-dephospho-CoA linked RNA [25] or methylphosphate capping [26,27] we refer to the indicated articles.

Review

Enzymatic preparation of capped mRNA

Enzymatic preparation of capped mRNA is based on in vitro transcription (IVT) of a DNA template. While RNA synthesized via solid-phase synthesis is limited in its maximum length, RNAs with a length of several thousand nucleotides can easily be prepared through IVT. On the other hand, enzymatically produced RNA is often inhomogeneous in length and for short RNAs the yields obtained after purification may be low. This impedes the enzymatic production of short RNAs of a defined length for applications requiring defined and homogeneous RNA species. IVT produces uncapped, 5′-triphosphorylated RNA but there are two strategies to obtain mRNA with a cap, which will be discussed in detail in the following chapters.

Post-transcriptional capping

In post-transcriptional capping, the RNA from IVT is subjected to a dedicated enzymatic capping reaction. The enzymes used in vitro originate from capping apparatuses of different eukaryotic organisms or DNA viruses and can be produced recombinantly in *E. coli* [28,29]. Enzymatic formation of cap0 comprises three consecutive reactions targeted to nascent 5′-triphosphorylated pre-mRNAs (Figure 1). First, a 5′-triphosphatase (TPase) hydrolyzes the γ-phosphate of pre-mRNA. Next, the β-phosphate of the resulting 5′-diphosphate end is coupled to GMP to form 5′–5′-linked Gppp-RNA. The responsible guanylyltransferase uses GTP as substrate and forms a covalent enzyme-(lysyl-*N*)-GMP intermediate, reminiscent of DNA ligase-AMP intermediates [30,31]. Finally, the cap structure is methylated at the *N*7-position by an RNA(guanine-*N*7)methyltransferase using *S*-adenoysl-L-methionine (AdoMet) as a cosubstrate [31].

In nature, these capping enzymes act co-transcriptionally once the transcript has reached a length of 20–30 nucleotides [32], which is enabled by their recruitment to the C-terminal domain of the RNA polymerase II [33]. In higher eukaryotes, cap1 and cap2 structures are generated by subsequent methylation of the 2′-hydroxy group of the adjacent second and third ribose, respectively [34].

These capping enzymes – e.g., from Vaccinia virus – can be harnessed for the production of capped RNA in vitro by adding them and their respective cosubstrates to the IVT reaction, as described by pioneering work of the Rosenberg group [35]. To date, the capping enzymes from the Vaccinia virus are commercially available and most widely used for post-transcriptional in vitro capping. They consist of two viral proteins D1 and D12. The triphosphatase and guanylyltransferase activity are located in the N-terminal half and the methyltransferase in the C-terminal half of the large D1 protein, whereas the small D12 protein has no catalytic activity but activates D1 [36-38].

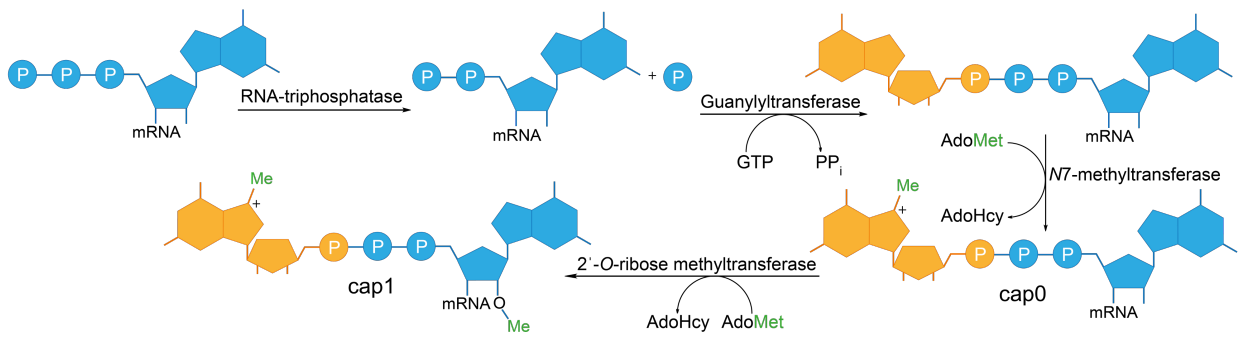
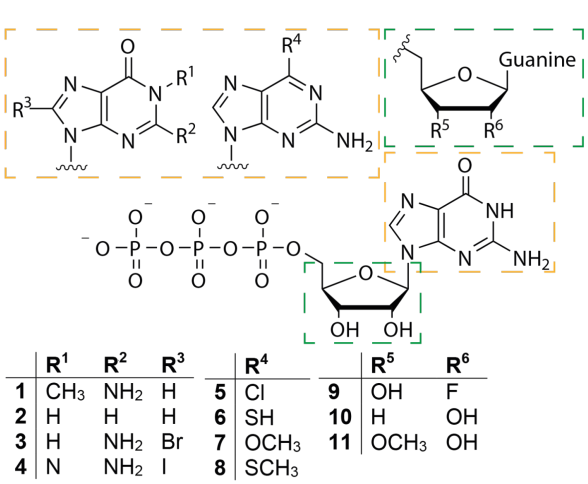


Figure 1: Schematic representation of enzymatic 5′-cap formation in eukaryotic mRNA. The 5′-triphosphate-end of the pre-mRNA is hydrolyzed to a diphosphate by an RNA 5′-triphosphatase. A guanylyltransferase transfers GMP onto the β-phosphate of the 5′-diphosphate to form a 5′ to 5′-triphosphate linkage. The guanine is methylated at the *N*7-position by an RNA (guanine-*N*7)methyltransferase, yielding the cap0 structure. Further methylation at the 2′-OH position of the first nucleotide results in formation of the cap1 structure.

Originally, the RNA capping with the Vaccinia capping apparatus was reported to be inefficient [35,37,39,40]. To date, the enzyme is commercially available, however, the amount of enzyme needed for the production of capped RNA in µmol scale prevents its general applicability [41]. For the application of the Vaccinia capping enzyme in the production of large-scale 5′-capped RNA, Fuchs et al. have recently reported an expression and purification protocol for the Vaccinia enzyme, allowing for capping in large quantities in a more cost-efficient manner compared to commercially available capping methods [41].

Post-transcriptional capping to obtain mRNA with a cap1 structure can be achieved using the Vaccinia mRNA cap 2′-*O*-methyltransferase which is commercially available [42,43]. Additionally, authentic mRNAs can be produced with the commercially available mScript™ system which combines a T7 RNA polymerase, a trifunctional capping enzyme, a 2′-*O*-methyltransferase and a poly(A) polymerase. Albeit expensive, this system allows for production of mRNAs in one pot with claimed quantitative yields and high translational activity.

Post-transcriptional preparation of non-natural cap analogues was achieved by capping enzymes with relaxed substrate specificity. For example, ribavirin is used as a substrate by the Vaccinia capping enzyme and can be transferred onto the diphosphate end of an RNA transcript to form a ribavirin-pppN structure. RNA transcripts blocked with ribavirin showed little translational efficiency, which might explain the antiviral activity of ribavirin [44]. Enzymatic formation of cap analogues from GTP analogues was achieved with the capping enzyme of the model organism *Paramecium bursaria Chlorella virus-1* (PBCV-1) by Bisaillon and co-workers [45]. Out of 22 nucleotide analogues tested in this study, 13 were found to form a covalent complex with the PBCV-1 guanylyltransferase (GTase) while 11 were actually transferred onto a 5′-diphosphate RNA (Figure 2). Moreover, RNAs capped with those nucleotide analogues were translated even in the absence of the *N*7-methyl group when alternative modifications enabled binding to eIF4E [45].



	R¹	R²	R³		R⁴		R⁵	R⁶
1	CH_3	NH_2	H	5	Cl	9	OH	F
2	H	H	H	6	SH	10	H	OH
3	H	NH_2	Br	7	OCH_3	11	OCH_3	OH
4	N	NH_2	I	8	SCH_3			

Figure 2: Nucleotide analogues **1**–**11** were converted by *Paramecium bursaria Chlorella virus-1* capping enzyme instead of GTP to generate RNAs with the respective caps [45].

Co-transcriptional capping

In co-transcriptional capping, cap analogues are added directly to the IVT. Their incorporation at the 5′-end by RNA polymerases with relaxed substrate specificity (e.g., T3, T7 or SP6 RNA polymerases) directly yields the respective 5′-capped mRNA (Figure 3). Internal incorporation of cap analogues during IVT does not occur, because cap analogues lack a free 5′-triphosphate.

The most commonly used cap analogue is m^7GpppG but several modified or alternative cap analogues are also accepted by RNA polymerases. Therefore, this route can be used to install non-natural dinucleotides at the 5′-end that are accessible for a further chemical reaction [46].

One often overlooked limitation of co-transcriptional capping is that not all mRNA obtained from IVT is capped, simply because the cap analogue competes with GTP as initiator nucleotide. Importantly, the ratio of capped/uncapped mRNA is usually not visible on a gel. This issue can be mitigated by lowering the GTP concentration or by digesting uncapped (i.e., triphosphorylated) RNA with a 5′-polyphosphatase which produces monophosphorylated RNA followed by 5′-phosphate-dependent exonuclease digestion.

Another problem encountered with m^7GpppG as initiator is elongation into the "wrong" direction, namely at the 3′-OH of m^7G, yielding mRNA with the cap in reverse orientation (Figure 3). Up to one half of the mRNA can contain the cap in its reverse orientation and will not be translated [47]. This problem was solved by developing anti-reverse cap analogues (ARCA) that are methylated or deoxygenated at the 3′-OH of the *N*7-methylguanosine ribose ($m_2^{7,3'\text{-}O}$GpppG or $m^{7,3'\text{-}d}$GpppG). This prevents elongation at the "wrong" 3′-OH and hence ARCA caps are exclusively incorporated in the correct orientation [48,49]. Interestingly, modifications at the 2′-position of m^7G also prevented reverse incorporation of the cap analogue [50]. The problem of orientation is circumvented when GpppA cap analogues are used in combination with the T7 class II promotor phi2.5 which allows initiating RNA synthesis with ATP. Hence, GpppA- or m^7GpppA-capped RNAs can be produced [51]. When the common GTP-initiating T7 class III promoter phi6.5 is used, GpppA is incorporated in its

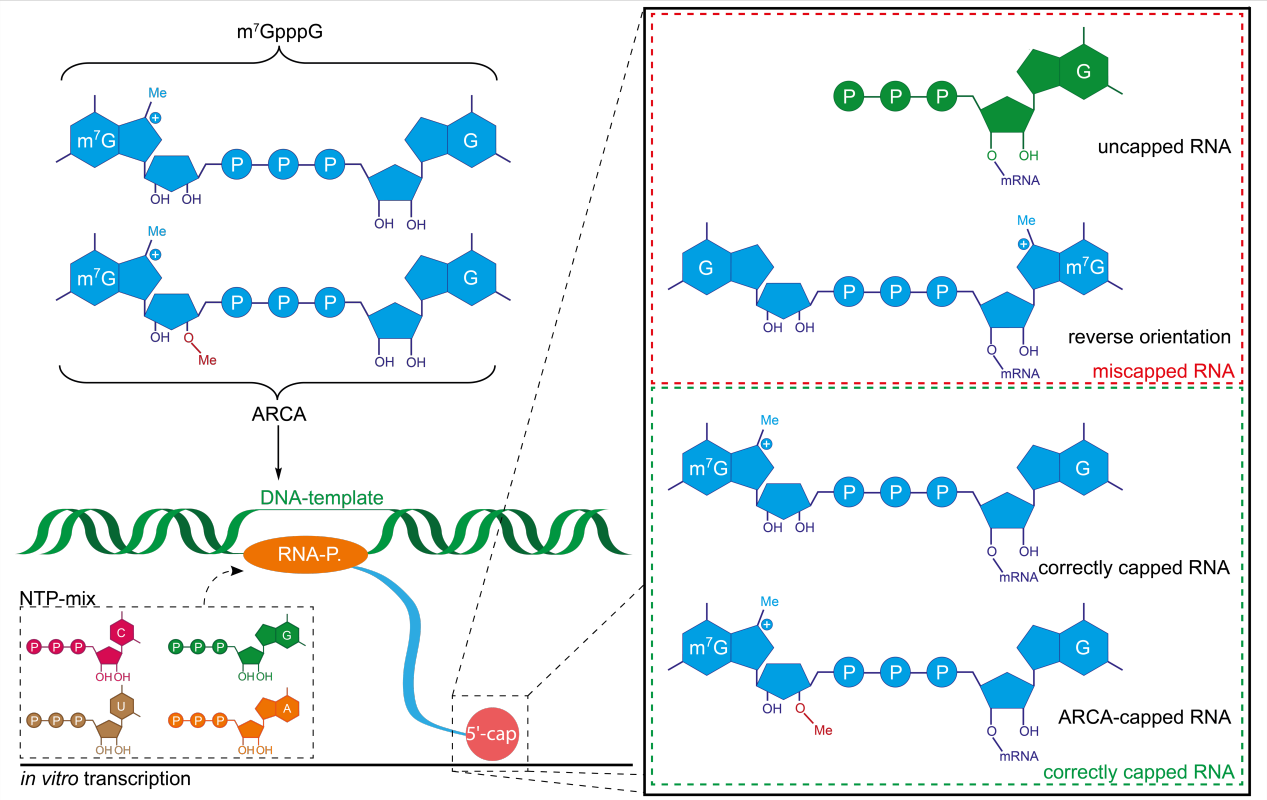


Figure 3: Schematic representation of co-transcriptional capping with different cap analogues. A DNA-dependent RNA polymerase initiates transcription from a DNA template by incorporation of a cap analogue, thereby producing capped RNA. When m^7GpppG is used as a cap analogue, miscapped RNA with the cap analogue incorporated in its reverse orientation is produced in addition to the correctly capped RNA. With the ARCA cap, reverse incorporation is excluded [48].

reverse orientation, yielding ApppG-capped mRNAs which are biologically not active.

Due to the strict preference of bacteriophage RNA polymerases for G or A, depending on the promotor, artificial mRNAs starting with a U or C at the 5′-end cannot be prepared using in vitro transcription which limits possible applications for example in structural analysis.

Co-transcriptional capping of short RNA fragments

For preparation of short, capped RNA with the sequence $GpppAN_n$ or $m^7GpppAN_n$ ($1 \leq n \leq 9$ nt), bacteriophage T7 gene 4 primase [52] or its active domain [53] can be used. Primase incorporates cap analogues exclusively in their correct orientation. Normally, gene 4 primase from the T7 phage produces short RNAs with the sequence pppAC from a DNA template. Matsuo et al. observed that GpppA or m^7GpppA can be incorporated as efficiently as ATP as the first nucleotide [52]. The substrate specificity of gene 4 primase for adenosine as the first nucleotide prevents incorporation of GpppA in its reverse orientation and incorporation of GpppG altogether. This method was used for the production of isotope-labeled capped RNA for cap-eIF4E NOESY-NMR studies [52]. Peyrane et al. demonstrated that using the N-terminal fragment bearing the primase activity resulted in comparable preparation yield for the RNA while expression and solubility of the fragment were improved [53].

mRNA cap analogues

Preparation of cap analogues

The co-transcriptional capping described above requires the preparation of cap analogues which are added to the transcription reaction. Ideally, these cap analogues should meet the following criteria: (i) high incorporation efficiencies when added to IVT, (ii) correct orientation when incorporated into RNA, (iii) strong binding to the cap-binding protein eIF4E, (iv) inhibitory potential when added as competitor in an in vitro translation assay and (v) high translation efficiency of resulting capped RNA. Figure 4A depicts the structure of the standard cap analogues m^7GpppG (**12**) and GpppG (**14**). The synthesis of m^7GpppG starts from guanosine diphosphate (GDP, **15**) and guanosine monophosphate (GMP, **16**, Figure 4B), which are both accessible by phosphorylation of guanosine [54]. Methylation of GDP gives m^7GDP (**17**) with high yield and regioselectivity [55]. The key step in cap analogue synthesis is the formation of the triphosphate linkage. Multiple strategies have been

reported which mostly rely on the same principle: One of the two nucleotides (typically the monophosphorylated nucleotide) is equipped with a good leaving group while the other one acts as a nucleophile. Different leaving groups have been exploited for the synthesis of cap analogues, comprising phenylthio [56], 5-chloro-8-quinolyl [57], morpholidate [48] and imidazolide moieties [58,59]. Imidazole activation in DMF with $ZnCl_2$ was first reported by Sekine et al. [60] and is the most often used method for the formation of triphosphates. *P*-Imidazoles are known to react with numerous nucleophiles such as nucleoside mono-, -di- or -triphosphates and are typically reacted in anhydrous DMF in the presence of zinc chloride. The GMP imidazolide (**18**) is reacted with m^7GDP (**17**) in the presence of $ZnCl_2$ as catalyst to yield m^7GpppG (**12**) [49].

In the past years, variations of this general synthetic strategy were used to obtain numerous cap analogues. Among the most interesting modifications is the above-mentioned anti-reverse cap analogue (ARCA) and GpppG which are commercially available. In addition, cap analogues with improved properties – namely binding to eIF4E, translational efficiency, and nuclease resistance – have been developed. Furthermore, cap analogues have therapeutic potential as demonstrated by a number of cap-derived translation inhibitors [61-63].

Applications of novel cap analogues

The search for novel – non-natural or modified – caps with improved properties has already yielded promising results. RNA capped with a locked nucleic acid (LNA)-modified dinucleotide cap analogue was translated 3-times more efficiently than regular m^7G-capped RNA [64]. Additionally, RNA capped with the LNA cap analogue was found to be ≈1.6-fold more stable in a luciferase assay in cultured cells than the respective RNA with the standard cap. However, in this study it was not

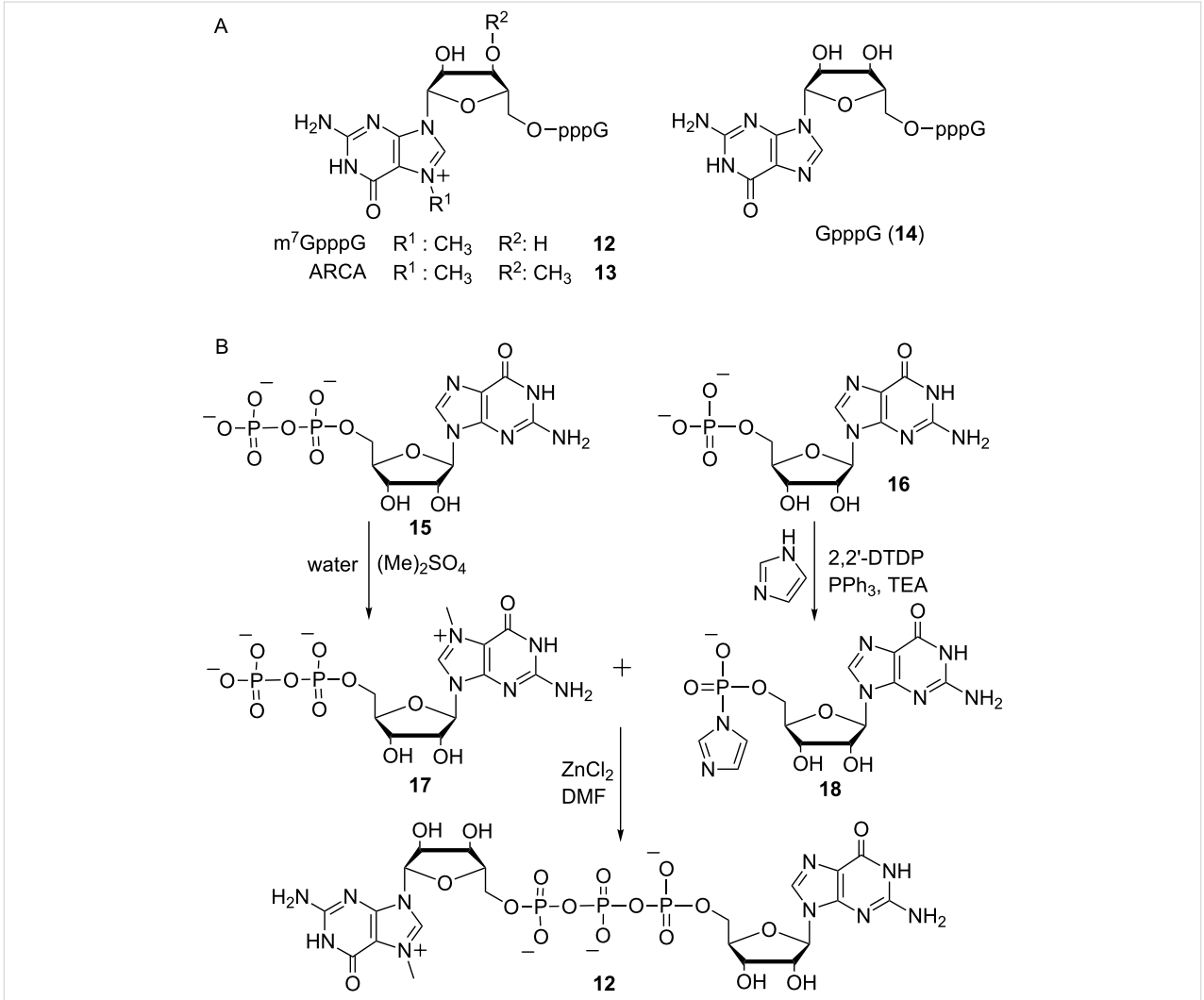


Figure 4: (A) Structures of commercially available mRNA cap analogues. (B) Synthetic route to cap analogues as exemplified by the synthesis of the m^7GpppG cap analogue. 2,2′-DTDP: 2,2′-dithiodipyridine.

assessed how this LNA cap analogue performs in comparison to the established ARCA cap. Interestingly, a 3′-*O*-propargyl containing m^7GpppG cap analogue also showed more than 3-fold higher translational efficiency compared to the standard cap. The cap analogue is exclusively incorporated in the correct (forward) orientation and molecular modelling studies pointed to a stronger binding of the propargyl-modified cap to eIF4E compared to the standard cap [65].

With regard to translational activity, several dinucleotide cap analogues containing a tetraphosphate were shown to be superior to the regular triphosphate in in vitro studies [66]. RNAs capped with $m^7Gp_4m^7G$ were translated with more than 3-fold higher efficiency. Interestingly, also benzyl-modified tetraphosphate cap analogues showed more than 2-fold higher translation in in vitro translation experiments. In a further step, tetraphosphates with methylene(bisphosphonate) moieties were prepared which improved binding to eIF4E and in some cases conferred enzymatic resistance against DcpS degradation [67]. N^2-Triazole-containing monophosphate cap analogues were shown to be as efficient as m^7GpppG in translation inhibition assays [68].

Further modifications can be placed in the phosphate moieties. ARCA-capped RNAs substituted with a sulphur atom at the β-position were shown to be resistant to the Dcp1/2 decapping complex from *S. pombe* while at the same time displaying high affinity to eIF4E and being translationally active when incorporated into RNA [69,70]. These properties were further improved with a range of 1,2-dithiodiphosphate cap analogues, some of which showed significantly improved stability when incorporated in an mRNA and applied in dendritic cells [71].

Furthermore, cap analogues providing additional functions were synthesized. A photo-crosslinking cap analogue containing a 6-thioguanosine was prepared which allowed for selective crosslinking [72]. Successful crosslinking was exemplified by the intrastrand crosslinking of histone H4 mRNA capped with a 6-thioguanosine cap analogue. Synthesis of biotin-labeled caps was achieved with a 2′-NH_2-modified cap analogue which was reacted with an *N*-hydroxysuccinimide biotin active ester [73]. The biotin-labeled cap analogue could be incorporated into mRNA during IVT and retained binding to eIF4E and translational activity in an in vitro translation assay.

Besides their use in the preparation of cap-modified RNAs via IVT, cap analogues have found alternative applications. Since cap-binding proteins (e.g., eIF4E and DcpS) have high affinity to cap analogues, resins functionalized with the cap analogue m^7GTP can be used to purify binding proteins from fractionated cell lysates [74-76]. Using m^7G-modified sepharose resins, novel cap-binding proteins such as gemin-5 could be identified [77]. The affinity resins can be stabilized via methylene moieties, preventing enzymatic degradation of the cap analogue [78].

In recent years, cap analogues started to be recognized as inhibitors of translation by interfering with the eIF4E-RNA cap interaction. In tumorigenesis, oncogenic activity of eIF4E was attributed to its ability to activate translation [79]. Besides standard cap analogues which have long been used for eIF4E inhibition in vitro [80], the pro-drug 4Ei-1 bearing an *N*7-benzyl moiety was shown to be a potent inhibitor of cap-dependent translation in zebrafish [81]. Poor cellular uptake of cap analogues could be circumvented by coupling to an adenovirus-like particle, resulting in inhibition of hepatocellular carcinoma growth in a rat model [82]. Recently, an artificially capped RNA was prepared bearing an orthosteric eIF4E inhibitor at its 5′-end [83,84]. RNA with this cap surrogate retained binding to eIF4E as measured by surface plasmon resonance (SPR). This work provides the basis for introduction of other artificial cap analogues at the 5′-end aiming to modulate biological activity of the resulting RNAs.

Enzymatic modification of chemically synthesized cap analogues

An alternative to the complete chemical synthesis of cap analogues is the use of enzymes to functionalize standard cap analogues. This approach benefits from the specificity of enzymes, hence the functional moieties are directly introduced at defined positions of the mRNA cap. In the past years, our group developed chemoenzymatic strategies for modification and functionalization at the *N*7- and N^2-position. Enzymatic modification is based on methyltransferases which naturally transfer a methyl group from their cosubstrate *S*-adenosyl-L-methionine (AdoMet) to the target molecule [85]. Functionalized side chains can be transferred from AdoMet analogues if an appropriate promiscuous methyltransferase is available [86]. Importantly, an unsaturated bond has to be present in β-position of the sulphonium center which stabilizes the transition state in the enzymatic transfer from the AdoMet analogue [87].

Engineering of the trimethylguanosine synthase GlaTgs2 from the protozoan *Giardia lamblia* resulted in a variant (V34A) which accommodated AdoMet analogues with bulkier sidechains and transferred various functional moieties including propargyl, pentenynyl, azidobut-2-enyl and 4-vinylbenzyl to the N^2-position of capped RNA or mRNA cap analogues such as m^7GpppA, m^7GpppG or m^7GTP (Figure 5A) [88-91]. Recently, we revealed that the *N*7-cap methyltransferase Ecm1 from *Encephalitozoon cuniculi* is highly promiscuous. Sterically very demanding AdoMet analogues bearing for example a

norbornene or 4-vinylbenzyl moiety were efficiently converted [92,93]. The pronounced promiscuity can be attributed to the structure of Ecm1 which forms a substrate binding cleft rather than a pocket [94].

Vinylbenzyl-modified cap analogues (bearing the modification at either the *N*7 or N^2-position) provided a platform for inverse electron-demand Diels–Alder reactions with tetrazine conjugates and for photo-click reactions using tetrazoles. Even photo-crosslinking moieties were enzymatically transferred to the *N*7-position of the mRNA cap from suitable AdoMet analogues. Notably, quantitative modification at the *N*7-position was achieved [96]. Diazirine and aryl–azide photo-crosslinker moieties were functional showing cross-linking to the cap-binding protein eIF4E. Microscale thermophoresis revealed that these crosslinker-modified caps still bound to eIF4E, albeit with

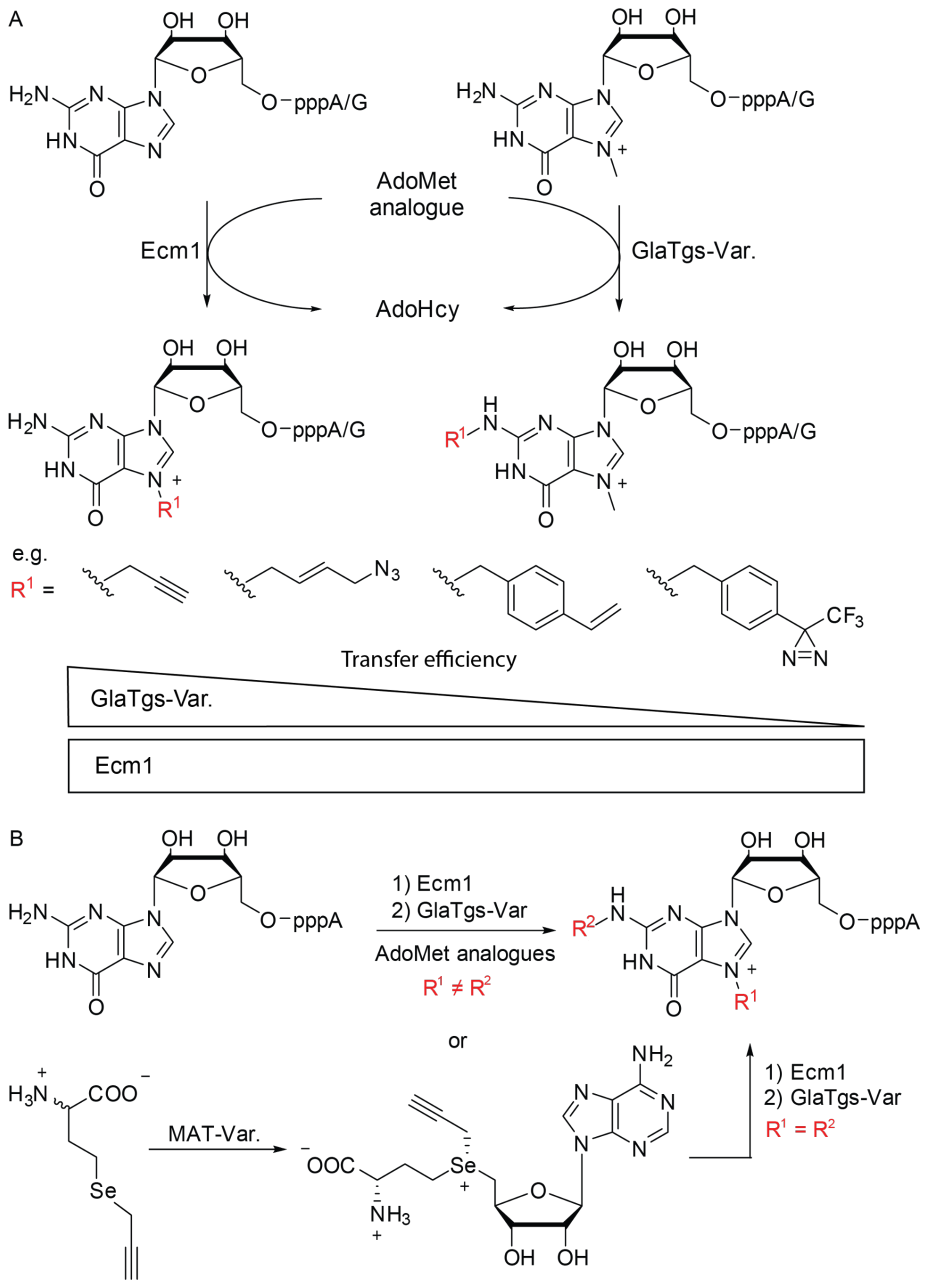


Figure 5: Enzymatic modification of cap analogues at their N^2- or *N*7-position or a combination of both. (A) Functional moieties such as alkynes and azides can be enzymatically transferred to the N^2-position using GlaTgs-Var. or *N*7-position using Ecm1 [88-91]. While transfer efficiencies decrease with increasing sterical demand when using GlaTgs-Var., transfer efficiencies are largely independent of size with Ecm1. (B) Both enzymes can be combined to yield dual or double modified cap analogues. The AdoMet analogue can also be prepared enzymatically starting from a (seleno)-methionine analogue and a MAT-Var [95].

strongly decreased affinity. Translation was highly susceptible to modifications at the *N*7-position of the mRNA cap. While for *N*7-allyl or *N*7-azidobutenyl modifications no translational activity was observed in vitro, the *N*7-benzyl modification showed residual activity. This may be attributed to a stacking of the benzyl-moiety between tryptophans in the eIF4E binding pocket [63,66,97].

Enzymatic modification at the N^2- and *N*7-position can also be combined to yield double and dual-modified cap analogues. Modification of the N^2-position by Tgs-enzymes is dependent on methylation at the guanine *N*7-position, which results in a positive charge. However, GlaTgs activity relies on the positive charge rather than the methyl group itself, as exemplified by studies showing that *N*7-ethyl and *N*7-benzyl-modified cap analogues are still substrates for GlaTgs [98]. This allowed us to enzymatically prepare cap analogues with different combinations of functional moieties (Figure 5B) [95]. A 4-vinylbenzyl/azido dual modification allowed appending two different fluorescent dyes which could be applied as FRET pair. In this case, labeling was achieved in two bioorthogonal reactions, an iEDDA and a SPAAC reaction. Furthermore, dual modification with an azido and an alkyne function enabled fluorophore/biotin labeling using a combination of SPAAC and CuAAC reaction. Efficient double labeling of the mRNA cap with alkyne moieties could also be achieved based on a recently reported enzymatic cascade reaction [99]. In this system a *Se*-propargyl-modified AdoMet analogue (SeAdoYn [100]) was prepared enzymatically from the respective methionine analogue and ATP by a methionine adenosyltransferase variant (MAT-Var.). The AdoMet analogue was directly converted by the methyltransferases, resulting in double alkyne modified cap analogues [95].

Chemical synthesis of capped mRNA

Solid-phase synthesis of capped RNA

Chemical synthesis of capped RNA is based on the solid-phase synthesis of RNA followed by chemical or enzymatic installation of the 5′-cap. The general principle of solid-phase RNA synthesis is beyond the scope of this review and has been described in excellent review articles [101-104]. The longest RNA synthesized via solid-phase chemistry to date has a length of 170 nucleotides and was prepared with the 2-cyanoethoxymethyl (CEM) as the 2′-OH protection group [105].

Chemical synthesis of 5′-capped RNA in solution was originally reported to be low yielding, slow (reaction times of 6–10 days), and not suitable for large-scale preparations [58,106-110]. Since then, several groups improved the chemical synthesis of capped RNA via solid-phase synthesis. The highly base-labile m^7G moiety turned out to be a limiting factor because it is not compatible with standard solid-phase deprotection protocols. Due to its positive charge, the m^7G moiety is hydrolytically less stable than other purine nucleosides. Under basic conditions which are commonly used for RNA deprotection and cleavage from the solid support, opening of the imidazole ring of the 7-methylguanine would occur [111]. Thus, for synthesis of the cap structure on the solid support, standard deprotection with ammonia is not possible.

An early example of capped RNA prepared by solid-phase synthesis was reported by the group of Sekine in 2001 [112]. A 2,2,7-trimethylguanosine (TMG)-capped trinucleotide block of U1 snRNA with the structure $m_3^{2,2,7}G^{5'}pppAm^{2'}Um^{2'}A$ was prepared, starting from a 5′-phosphorylated trimer synthesized by standard phosphoramidite chemistry. To address the problem of m^7G instability under basic conditions, the TMG-capping reaction was carried out upon deprotection of all base-labile groups. Utilization of a novel, acid labile linker to the solid support allowed for subsequent release of the RNA. However, due to overall low coupling efficiencies and isolated yields (the compound was isolated in 20% overall yield after anion-exchange chromatography), this method was not used for large scale synthesis of capped RNA (Figure 6A). As the low reaction yields are mainly caused by the multistep preparation of the triphosphate bridge, the Sekine group presented a synthetic route to RNA bearing a 5′-terminal TMG-capped pyrophosphate linkage on solid support. Since pyrophosphate formation is easier than triphosphate formation, this route resulted in higher coupling yields. Whether this RNA is still biologically active remains to be demonstrated [113]. Furthermore, these capping approaches can be used to produce biologically relevant RNA. U1snRNA was prepared via enzymatic ligation of a short RNA (10 nt long) containing a trimethylated $m_3^{2,2,7}G$ cap moiety to a 154 nt long RNA produced via IVT. The respective U1snRNAs with a pyrophosphate bridged TMG cap and a TMG cap containing an ethylene glycol linkage were also produced [114].

Unlike IVT, solid-phase synthesis offers the flexibility to introduce modified nucleotides at specific positions. Chemical synthesis of the intricate trypanosomatid cap4 structure, characterized by 2′-*O*-methylation of the first four nucleotides and additional methylation at the first adenosine and the fourth uridine, was reported in 2004 by the group of Darzynkiewicz. The preparation was achieved by reacting an imidazole activated m^7GDP with the 5′-phosphorylated tetramer [115]. This cap was successfully used for affinity purification of trypanosomatid cap4 interacting proteins [116,117].

Nagata et al. reported on the first preparation of mature mRNA based on a chemically synthesized RNA strand which was

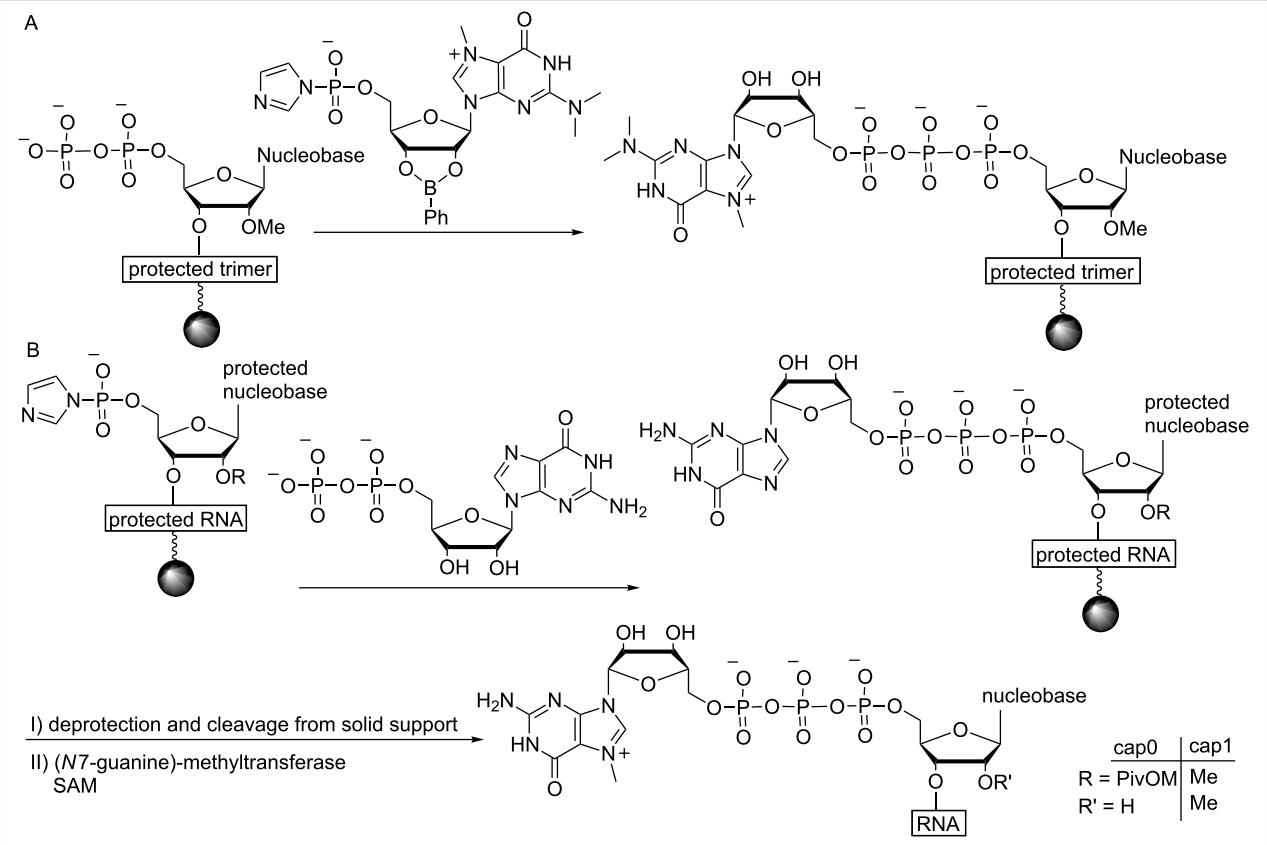


Figure 6: Synthesis of cap-containing RNA by solid-phase synthesis. (A) A TMG-capped mRNA was synthesized starting from an RNA tetramer which was subjected to 5′-terminal pyrophosphorylation followed by reaction with a 2,2,7-trimethylguanosine 5′-phosphorimidazolide derivative [112]. Subsequent cleavage from the solid support was achieved using 80% AcOH (rt, 24 h) and TBDMS protecting groups were removed with HCl (pH 2, rt, 12 h). (B) Large-scale production of RNAs with cap0 or cap1 by a combination of solid-phase synthesis and enzymatic methylation [111]. Deprotection conditions: DBU (1,8-diazadicyclo[5,4,0]undec-7-ene) in acetonitrile (rt, 3 min) followed by treatment with aqueous ammonia (rt, 3 h).

shown to be biologically active in cells [105]. This was achieved by combining solid-phase synthesis and enzymatic modification. Specifically, 5′-diphosphorylated RNAs (up to 170 nt long) were chemically synthesized, cleaved from the solid support, deprotected and purified. This was followed by enzymatic capping, 2′-*O*-methylation and polyadenylation.

A combination of chemical synthesis and enzymatic modification was also used by Thillier et al. for the large scale synthesis of capped RNA. Herein, to circumvent the problem of m^7G instability, non-methylated capped RNAs were first synthesized using the phosphoramidite 2′-*O*-pivaloyloxymethyl method, followed by enzymatic *N*7 methylation using the human (guanine-*N*7)-methyltransferase (Figure 6B). A cap1 structure could also be obtained via 2′-OH methylation of the terminal nucleotide [111]. This approach was applied in collaboration with other groups for the production and investigation of capped RNA [118].

In summary novel chemical capping strategies enable preparation of capped RNAs in high yield and independent of the sequence, providing access to RNAs that could not be prepared via IVT. However, preparation of biologically relevant mRNAs that are typically thousands of nucleotides long is not directly feasible, as the longest chemically prepared RNA to date comprises 170 nt. Methods combining chemical and enzymatic preparation of capped RNA bear potential to resolve these limitations and will be described in the following.

Combining chemical and enzymatic methods: primer extension

Engineering of the replicative DNA polymerase from *Thermococcus gorgonarius* (Tgo) into a DNA-dependent RNA polymerase (termed TGK) enabled production of up to 1,700 nt long RNAs from a ssDNA template and an RNA primer [119]. The primer-dependent RNA synthesis obviates the need to initiate RNA synthesis with pppG in contrast to most other RNA polymerases used for conventional IVT. TGK turned out to accept a number of variations at the 5′-end including an oligoribonucleotide primer containing the desired cap. This approach unites the flexibility of RNA synthesis and processivity of RNA polymerases for the preparation of long and cap modified RNAs.

Using this system, several biologically relevant RNAs such as GFP RNA, firefly luciferase RNA and $m^7Gpppm^6A_m$-RNA were produced [118,119].

Click chemistry for the preparation of capped RNAs and cap analogues

As an alternative to preparation of longer RNA via IVT, different hypermethylated cap analogues with a 2′-azido moiety allowed for reaction with an alkyne-modified RNA in a CuAAC reaction to yield cap modified RNA – albeit with a non-natural linkage (Figure 7A) [120]. This capping strategy also worked with an alkyne-modified triphosphorylated RNA and 5′-azido modified methylguanosine resulting in a capped RNA containing a triazole linkage after CuAAC reaction (Figure 7B) [121]. In a similar approach a 5′-azido-modified RNA was prepared by solid-phase synthesis and reacted with an alkyne-functionalized m^7G-cap analogue in a CuAAC reaction [122]. Besides its utility on long RNA, this click chemistry approach was also

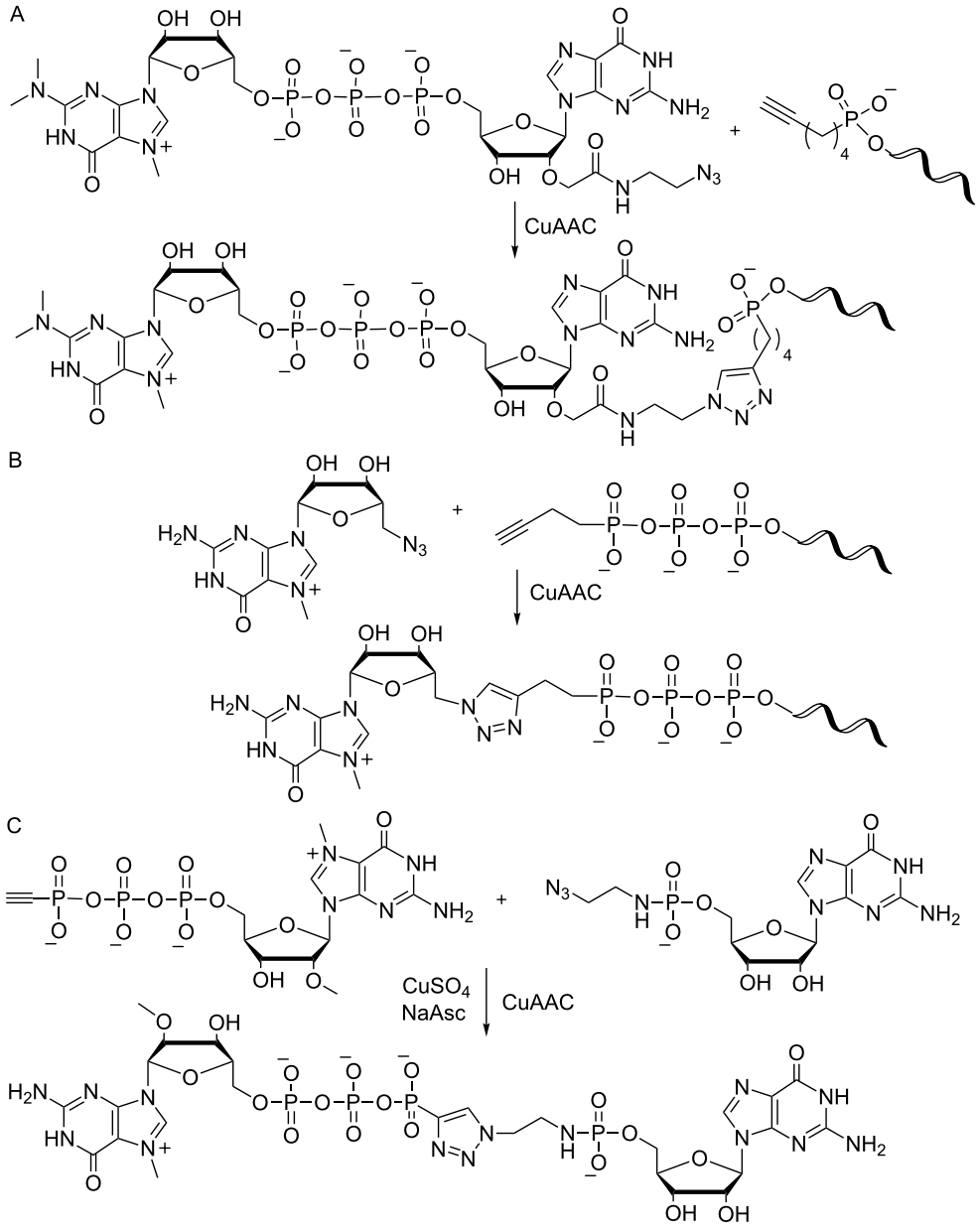


Figure 7: Click chemistry for the preparation of capped RNA and cap analogues. (A) Preparation of capped RNA via a copper-catalyzed azide–alkyne cycloaddition (CuAAC) of an azido-modified cap analogue with a 5'-alkyne bearing RNA [120]. (B) An alkyne-modified triphosphorylated RNA is reacted with 5'-azido-methylguanosine in a CuAAC [121]. (C) Alkyne- and azido-containing nucleotide building blocks are reacted in a CuAAC to give a functional cap analogue [121].

applied to the chemical synthesis of cap analogues, simplifying the typically laborious and time-consuming synthesis [121]. A plethora of cap analogues was synthesized replacing one phosphate bridge with a triazole linkage. Depending on their structure and the exact positioning of the triazole linkage, modified cap analogues varied largely with regard to their functionality in in vitro translational assays, binding affinity to eIF4E and resistance to the decapping enzyme DcpS. Best translational efficiencies similar to the standard cap were achieved with a tetraphosphate cap analogue containing a triazole bridge (Figure 7C).

Conclusion

The 5′-cap is the key modification of eukaryotic mRNAs and provides an interaction platform for proteins involved in fundamental processes like nuclear export and translation. Therefore, preparation of mRNAs with the canonical cap structure is indispensable for a comprehensive understanding of mRNA functions that go beyond the genetically encoded information, e.g., studies elucidating RNA-protein interactions [123] or structure analysis [124]. Moreover, artificially capped RNAs or RNAs with modified 5′-caps may provide a means to control or selectively block some of these functions, resulting in improved translational efficiency or higher stability.

Depending on the desired length of the capped RNA fully synthetic, enzymatic or a combination of both strategies is feasible and allows production of differently capped RNAs with a length ranging from several nucleotides to authentic mRNAs (>1000 nt). Novel strategies for the synthesis of cap analogues have led to the development of 5′-caps with tailored functionalities which are, for instance, resistant to enzymatic degradation or bear functional moieties for additional bioconjugation reactions. A combination of chemical 5′-cap analogue synthesis followed by enzymatic modifications has further allowed conferring novel functionalities (e.g., photo-crosslinking moieties) which were previously not easily accessible. Combining enzymatic modification at different positions (e.g., N^2 and $N7$-position) renders dual and double modifications possible, further diversifying mRNA 5′-cap modifications and leading to the highly regiospecific introduction of two different functionalities. Most recent developments focused on the development of completely artificial mRNA caps which conferred specific properties such as eIF4E binding and turned non-modified RNAs into strongly eIF4E-binding RNAs.

Acknowledgements

A.R. gratefully acknowledges financial support by the DFG (RE2796/2-1, RE2796/6-1) and the Fonds der Chemischen Industrie (Dozentenpreis).

ORCID® iDs

Nils Muthmann - https://orcid.org/0000-0003-0313-4242
Andrea Rentmeister - https://orcid.org/0000-0002-3107-4147

References

1. Furuichi, Y.; LaFiandra, A.; Shatkin, A. J. *Nature* **1977,** *266,* 235–239. doi:10.1038/266235a0
2. Shimotohno, K.; Kodama, Y.; Hashimoto, J.; Miura, K. I. *Proc. Natl. Acad. Sci. U. S. A.* **1977,** *74,* 2734–2738. doi:10.1073/pnas.74.7.2734
3. Konarska, M. M.; Padgett, R. A.; Sharp, P. A. *Cell* **1984,** *38,* 731–736. doi:10.1016/0092-8674(84)90268-X
4. Edery, I.; Sonenberg, N. *Proc. Natl. Acad. Sci. U. S. A.* **1985,** *82,* 7590–7594. doi:10.1073/pnas.82.22.7590
5. Köhler, A.; Hurt, E. *Nat. Rev. Mol. Cell Biol.* **2007,** *8,* 761–773. doi:10.1038/nrm2255
6. Sonenberg, N.; Hinnebusch, A. G. *Cell* **2009,** *136,* 731–745. doi:10.1016/j.cell.2009.01.042
7. Müller-McNicoll, M.; Neugebauer, K. M. *Nat. Rev. Genet.* **2013,** *14,* 275–287. doi:10.1038/nrg3434
8. Wang, Z.; Jiao, X.; Carr-Schmid, A.; Kiledjian, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 12663–12668. doi:10.1073/pnas.192445599
9. Izaurralde, E.; Lewis, J.; McGuigan, C.; Jankowska, M.; Darzynkiewicz, E.; Mattaj, I. W. *Cell* **1994,** *78,* 657–668. doi:10.1016/0092-8674(94)90530-4
10. Izaurralde, E.; Lewis, J.; Gamberi, C.; Jarmolowski, A.; McGuigan, C.; Mattaj, I. W. *Nature* **1995,** *376,* 709–712. doi:10.1038/376709a0
11. Marcotrigiano, J.; Gingras, A.-C.; Sonenberg, N.; Burley, S. K. *Cell* **1997,** *89,* 951–961. doi:10.1016/S0092-8674(00)80280-9
12. Pichlmair, A.; Reis e Sousa, C. *Immunity* **2007,** *27,* 370–383. doi:10.1016/j.immuni.2007.08.012
13. Sahin, U.; Karikó, K.; Türeci, Ö. *Nat. Rev. Drug Discovery* **2014,** *13,* 759–780. doi:10.1038/nrd4278
14. Pichlmair, A.; Schulz, O.; Tan, C. P.; Näslund, T. I.; Liljeström, P.; Weber, F.; Reis e Sousa, C. *Science* **2006,** *314,* 997–1001. doi:10.1126/science.1132998
15. Hornung, V.; Ellegast, J.; Kim, S.; Brzózka, K.; Jung, A.; Kato, H.; Poeck, H.; Akira, S.; Conzelmann, K.-K.; Schlee, M.; Endres, S.; Hartmann, G. *Science* **2006,** *314,* 994–997. doi:10.1126/science.1132505
16. Zamudio, J. R.; Mittra, B.; Campbell, D. A.; Sturm, N. R. *Mol. Microbiol.* **2009,** *72,* 1100–1110. doi:10.1111/j.1365-2958.2009.06696.x
17. Ullu, E.; Tschudi, C. *Proc. Natl. Acad. Sci. U. S. A.* **1991,** *88,* 10074–10078. doi:10.1073/pnas.88.22.10074
18. Ziemniak, M.; Strenkowska, M.; Kowalska, J.; Jemielity, J. *Future Med. Chem.* **2013,** *5,* 1141–1172. doi:10.4155/fmc.13.96
19. Warminski, M.; Sikorski, P. J.; Kowalska, J.; Jemielity, J. *Top. Curr. Chem.* **2017,** *375,* No. 16. doi:10.1007/s41061-017-0106-y
20. Weissman, D.; Karikó, K. *Mol. Ther.* **2015,** *23,* 1416–1417. doi:10.1038/mt.2015.138
21. Wei, L.; Cao, L.; Xi, Z. *Angew. Chem., Int. Ed.* **2013,** *52,* 6501–6503. doi:10.1002/anie.201301122
22. Egetenmeyer, S.; Richert, C. *Chem. – Eur. J.* **2011,** *17,* 11813–11827. doi:10.1002/chem.201101828
23. Höfer, K.; Abele, F.; Schlotthauer, J.; Jäschke, A. *Bioconjugate Chem.* **2016,** *27,* 874–877. doi:10.1021/acs.bioconjchem.6b00072

24. Cahová, H.; Winz, M.-L.; Höfer, K.; Nübel, G.; Jäschke, A. *Nature* **2015,** *519,* 374–377. doi:10.1038/nature14020
25. Kowtoniuk, W. E.; Shen, Y.; Heemstra, J. M.; Agarwal, I.; Liu, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106,* 7768–7773. doi:10.1073/pnas.0900528106
26. Shumyatsky, G.; Wright, D.; Reddy, R. *Nucleic Acids Res.* **1993,** *21,* 4756–4761. doi:10.1093/nar/21.20.4756
27. Singh, R.; Reddy, R. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86,* 8280–8283. doi:10.1073/pnas.86.21.8280
28. Martin, S. A.; Moss, B. *J. Biol. Chem.* **1975,** *250,* 9330–9335.
29. Furuichi, Y.; Muthukrishnan, S.; Tomasz, J.; Shatkin, A. J. *J. Biol. Chem.* **1976,** *251,* 5043–5053.
30. Shuman, S. *Prog. Nucleic Acid Res. Mol. Biol.* **2001,** *66,* 1–40.
31. Fabrega, C.; Hausmann, S.; Shen, V.; Shuman, S.; Lima, C. D. *Mol. Cell* **2004,** *13,* 77–89. doi:10.1016/S1097-2765(03)00522-7
32. Salditt-Georgieff, M.; Harpold, M.; Chen-Kiang, S.; Darnell, J. E., Jr. *Cell* **1980,** *19,* 69–78. doi:10.1016/0092-8674(80)90389-X
33. Cho, E. J.; Takagi, T.; Moore, C. R.; Buratowski, S. *Genes Dev.* **1997,** *11,* 3319–3326. doi:10.1101/gad.11.24.3319
34. Kuge, H.; Brownlee, G. G.; Gershon, P. D.; Richter, J. D. *Nucleic Acids Res.* **1998,** *26,* 3208–3214. doi:10.1093/nar/26.13.3208
35. Paterson, B. M.; Rosenberg, M. *Nature* **1979,** *279,* 692–696. doi:10.1038/279692a0
36. Myette, J. R.; Niles, E. G. *J. Biol. Chem.* **1996,** *271,* 11936–11944. doi:10.1074/jbc.271.20.11936
37. Shuman, S.; Surks, M.; Furneaux, H.; Hurwitz, J. *J. Biol. Chem.* **1980,** *255,* 11588–11598.
38. Martin, S. A.; Paoletti, E.; Moss, B. *J. Biol. Chem.* **1975,** *250,* 9322–9329.
39. Contreras, R.; Cheroutre, H.; Degrave, W.; Fiers, W. *Nucleic Acids Res.* **1982,** *10,* 6353–6362. doi:10.1093/nar/10.20.6353
40. Pelletier, J.; Sonenberg, N. *Cell* **1985,** *40,* 515–526. doi:10.1016/0092-8674(85)90200-4
41. Fuchs, A.-L.; Neu, A.; Sprangers, R. *RNA* **2016,** *22,* 1454–1466. doi:10.1261/rna.056614.116
42. Lockless, S. W.; Cheng, H.-T.; Hodel, A. E.; Quiocho, F. A.; Gershon, P. D. *Biochemistry* **1998,** *37,* 8564–8574. doi:10.1021/bi980178m
43. Barbosa, E.; Moss, B. *J. Biol. Chem.* **1978,** *253,* 7698–7702.
44. Bougie, I.; Bisaillon, M. *J. Biol. Chem.* **2004,** *279,* 22124–22130. doi:10.1074/jbc.M400908200
45. Issur, M.; Bougie, I.; Despins, S.; Bisaillon, M. *PLoS One* **2013,** *8,* e75310. doi:10.1371/journal.pone.0075310
46. Samanta, A.; Krause, A.; Jäschke, A. *Chem. Commun.* **2014,** *50,* 1313–1316. doi:10.1039/C3CC46132G
47. Pasquinelli, A. E.; Dahlberg, J. E.; Lund, E. *RNA* **1995,** *1,* 957–967.
48. Peng, Z.-H.; Sharma, V.; Singleton, S. F.; Gershon, P. D. *Org. Lett.* **2002,** *4,* 161–164. doi:10.1021/ol0167715
49. Stepinski, J.; Waddell, C.; Stolarski, R.; Darzynkiewicz, E.; Rhoads, R. E. *RNA* **2001,** *7,* 1486–1495.
50. Jemielity, J.; Fowler, T.; Zuberek, J.; Stepinski, J.; Lewdorowicz, M.; Niedzwiecka, A.; Stolarski, R.; Darzynkiewicz, E.; Rhoads, R. E. *RNA* **2003,** *9,* 1108–1122. doi:10.1261/rna.5430403
51. Coleman, T. M.; Wang, G.; Huang, F. *Nucleic Acids Res.* **2004,** *32,* e14. doi:10.1093/nar/gnh007
52. Matsuo, H.; Moriguchi, T.; Takagi, T.; Kusakabe, T.; Buratowski, S.; Sekine, M.; Kyogoku, Y.; Wagner, G. *J. Am. Chem. Soc.* **2000,** *122,* 2417–2421. doi:10.1021/ja9926820
53. Peyrane, F.; Selisko, B.; Decroly, E.; Vasseur, J. J.; Benarroch, D.; Canard, B.; Alvarez, K. *Nucleic Acids Res.* **2007,** *35,* e26. doi:10.1093/nar/gkl1119
54. Hall, R. H.; Khorana, H. G. *J. Am. Chem. Soc.* **1955,** *77,* 1871–1875. doi:10.1021/ja01612a049
55. Kore, A. R.; Parmar, G. *Nucleosides, Nucleotides Nucleic Acids* **2006,** *25,* 337–340. doi:10.1080/15257770500544552
56. Nakagawa, I.; Konya, S.; Ohtani, S.; Hata, T. *Synthesis* **1980,** 556–557. doi:10.1055/s-1980-29119
57. Fukuoka, K.; Suda, F.; Suzuki, R.; Takaku, H.; Ishikawa, M.; Hata, T. *Tetrahedron Lett.* **1994,** *35,* 1063–1066. doi:10.1016/S0040-4039(00)79966-7
58. Sawai, H.; Wakai, H.; Nakamura-Ozaki, A. *J. Org. Chem.* **1999,** *64,* 5836–5840. doi:10.1021/jo990286u
59. Sawai, H.; Wakai, H.; Shimazu, M. *Tetrahedron Lett.* **1991,** *32,* 6905–6906. doi:10.1016/0040-4039(91)80440-H
60. Kadokura, M.; Wada, T.; Urashima, C.; Sekine, M. *Tetrahedron Lett.* **1997,** *38,* 8359–8362. doi:10.1016/S0040-4039(97)10263-5
61. Jia, Y.; Chiu, T.-L.; Amin, E. A.; Polunovsky, V.; Bitterman, P. B.; Wagner, C. R. *Eur. J. Med. Chem.* **2010,** *45,* 1304–1313. doi:10.1016/j.ejmech.2009.11.054
62. Cai, A.; Jankowska-Anyszka, M.; Centers, A.; Chlebicka, L.; Stepinski, J.; Stolarski, R.; Darzynkiewicz, E.; Rhoads, R. E. *Biochemistry* **1999,** *38,* 8538–8547. doi:10.1021/bi9830213
63. Darzynkiewicz, E.; Stepinski, J.; Ekiel, I.; Goyer, C.; Sonenberg, N.; Temeriusz, A.; Jin, Y.; Sijuwade, T.; Haber, D.; Tahara, S. M. *Biochemistry* **1989,** *28,* 4771–4778. doi:10.1021/bi00437a038
64. Kore, A. R.; Shanmugasundaram, M.; Charles, I.; Vlassov, A. V.; Barta, T. J. *J. Am. Chem. Soc.* **2009,** *131,* 6364–6365. doi:10.1021/ja901655p
65. Shanmugasundaram, M.; Charles, I.; Kore, A. R. *Bioorg. Med. Chem.* **2016,** *24,* 1204–1208. doi:10.1016/j.bmc.2016.01.048
66. Grudzien, E.; Stepinski, J.; Jankowska-Anyszka, M.; Stolarski, R.; Darzynkiewicz, E.; Rhoads, R. E. *RNA* **2004,** *10,* 1479–1487. doi:10.1261/rna.7380904
67. Rydzik, A. M.; Lukaszewicz, M.; Zuberek, J.; Kowalska, J.; Darzynkiewicz, Z. M.; Darzynkiewicz, E.; Jemielity, J. *Org. Biomol. Chem.* **2009,** *7,* 4763–4776. doi:10.1039/b911347a
68. Piecyk, K.; Lukaszewicz, M.; Darzynkiewicz, E.; Jankowska-Anyszka, M. *RNA* **2014,** *20,* 1539–1547. doi:10.1261/rna.046193.114
69. Grudzien-Nogalska, E.; Jemielity, J.; Kowalska, J.; Darzynkiewicz, E.; Rhoads, R. E. *RNA* **2007,** *13,* 1745–1755. doi:10.1261/rna.701307
70. Kowalska, J.; Lewdorowicz, M.; Zuberek, J.; Grudzien-Nogalska, E.; Bojarska, E.; Stepinski, J.; Rhoads, R. E.; Darzynkiewicz, E.; Davis, R. E.; Jemielity, J. *RNA* **2008,** *14,* 1119–1131. doi:10.1261/rna.990208
71. Strenkowska, M.; Grzela, R.; Majewski, M.; Wnek, K.; Kowalska, J.; Lukaszewicz, M.; Zuberek, J.; Darzynkiewicz, E.; Kuhn, A. N.; Sahin, U.; Jemielity, J. *Nucleic Acids Res.* **2016,** *44,* 9578–9590. doi:10.1093/nar/gkw896
72. Nowakowska, M.; Kowalska, J.; Martin, F.; d'Orchymont, A.; Zuberek, J.; Lukaszewicz, M.; Darzynkiewicz, E.; Jemielity, J. *Org. Biomol. Chem.* **2014,** *12,* 4841–4847. doi:10.1039/C4OB00059E
73. Jemielity, J.; Lukaszewicz, M.; Kowalska, J.; Czarnecki, J.; Zuberek, J.; Darzynkiewicz, E. *Org. Biomol. Chem.* **2012,** *10,* 8570–8574. doi:10.1039/c2ob26060c
74. Sonenberg, N.; Rupprecht, K. M.; Hecht, S. M.; Shatkin, A. J. *Proc. Natl. Acad. Sci. U. S. A.* **1979,** *76,* 4345–4349. doi:10.1073/pnas.76.9.4345

75. Altmann, M.; Edery, I.; Sonenberg, N.; Trachsel, H. *Biochemistry* **1985,** *24,* 6085–6089. doi:10.1021/bi00343a009
76. Edery, I.; Altmann, M.; Sonenberg, N. *Gene* **1988,** *74,* 517–525. doi:10.1016/0378-1119(88)90184-9
77. Bradrick, S. S.; Gromeier, M. *PLoS One* **2009,** *4,* e7030. doi:10.1371/journal.pone.0007030
78. Szczepaniak, S. A.; Zuberek, J.; Darzynkiewicz, E.; Kufel, J.; Jemielity, J. *RNA* **2012,** *18,* 1421–1432. doi:10.1261/rna.032078.111
79. Wendel, H.-G.; Silva, R. L. A.; Malina, A.; Mills, J. R.; Zhu, H.; Ueda, T.; Watanabe-Fukunaga, R.; Fukunaga, R.; Teruya-Feldstein, J.; Pelletier, J.; Lowe, S. W. *Genes Dev.* **2007,** *21,* 3232–3237. doi:10.1101/gad.1604407
80. Darzynkiewicz, E.; Ekiel, I.; Lassota, P.; Tahara, S. M. *Biochemistry* **1987,** *26,* 4372–4380. doi:10.1021/bi00388a028
81. Ghosh, B.; Benyumov, A. O.; Ghosh, P.; Jia, Y.; Avdulov, S.; Dahlberg, P. S.; Peterson, M.; Smith, K.; Polunovsky, V. A.; Bitterman, P. B.; Wagner, C. R. *ACS Chem. Biol.* **2009,** *4,* 367–377. doi:10.1021/cb9000475
82. Zochowska, M.; Piguet, A.-C.; Jemielity, J.; Kowalska, J.; Szolajska, E.; Dufour, J.-F.; Chroboczek, J. *Nanomedicine* **2015,** *11,* 67–76. doi:10.1016/j.nano.2014.07.009
83. Gampe, C. M.; Hollis-Symynkywicz, M.; Zécri, F. *Angew. Chem., Int. Ed.* **2016,** *55,* 10283–10286. doi:10.1002/anie.201604385
84. Chen, X.; Kopecky, D. J.; Mihalic, J.; Jeffries, S.; Min, X.; Heath, J.; Deignan, J.; Lai, S.; Fu, Z.; Guimaraes, C.; Shen, S.; Li, S.; Johnstone, S.; Thibault, S.; Xu, H.; Cardozo, M.; Shen, W.; Walker, N.; Kayser, F.; Wang, Z. *J. Med. Chem.* **2012,** *55,* 3837–3851. doi:10.1021/jm300037x
85. Cheng, X.; Roberts, R. J. *Nucleic Acids Res.* **2001,** *29,* 3784–3795. doi:10.1093/nar/29.18.3784
86. Klimašauskas, S.; Weinhold, E. *Trends Biotechnol.* **2007,** *25,* 99–104. doi:10.1016/j.tibtech.2007.01.006
87. Dalhoff, C.; Lukinavičius, G.; Klimašauskas, S.; Weinhold, E. *Nat. Chem. Biol.* **2006,** *2,* 31–32. doi:10.1038/nchembio754
88. Schulz, D.; Holstein, J. M.; Rentmeister, A. *Angew. Chem., Int. Ed.* **2013,** *52,* 7874–7878. doi:10.1002/anie.201302874
89. Holstein, J. M.; Schulz, D.; Rentmeister, A. *Chem. Commun.* **2014,** *50,* 4478–4481. doi:10.1039/C4CC01549E
90. Holstein, J. M.; Stummer, D.; Rentmeister, A. *Chem. Sci.* **2015,** *6,* 1362–1369. doi:10.1039/C4SC03182B
91. Holstein, J. M.; Stummer, D.; Rentmeister, A. *Protein Eng., Des. Sel.* **2015,** *28,* 179–186. doi:10.1093/protein/gzv011
92. Muttach, F.; Muthmann, N.; Reichert, D.; Anhäuser, L.; Rentmeister, A. *Chem. Sci.* **2017,** *8,* 7947–7953. doi:10.1039/C7SC03631K
93. Holstein, J. M.; Anhäuser, L.; Rentmeister, A. *Angew. Chem., Int. Ed.* **2016,** *55,* 10899–10903. doi:10.1002/anie.201604107
94. Hausmann, S.; Zheng, S.; Fabrega, C.; Schneller, S. W.; Lima, C. D.; Shuman, S. *J. Biol. Chem.* **2005,** *280,* 20404–20412. doi:10.1074/jbc.M501073200
95. Holstein, J. M.; Muttach, F.; Schiefelbein, S. H. H.; Rentmeister, A. *Chem. – Eur. J.* **2017,** *23,* 6165–6173. doi:10.1002/chem.201604816
96. Muttach, F.; Mäsing, F.; Studer, A.; Rentmeister, A. *Chem. – Eur. J.* **2017,** *23,* 5988–5993. doi:10.1002/chem.201605663
97. Brown, C. J.; McNae, I.; Fischer, P. M.; Walkinshaw, M. D. *J. Mol. Biol.* **2007,** *372,* 7–15. doi:10.1016/j.jmb.2007.06.033
98. Benarroch, D.; Jankowska-Anyszka, M.; Stepinski, J.; Darzynkiewicz, E.; Shuman, S. *RNA* **2010,** *16,* 211–220. doi:10.1261/rna.1872110
99. Muttach, F.; Rentmeister, A. *Angew. Chem., Int. Ed.* **2016,** *55,* 1917–1920. doi:10.1002/anie.201507577
100. Willnow, S.; Martin, M.; Lüscher, B.; Weinhold, E. *ChemBioChem* **2012,** *13,* 1167–1173. doi:10.1002/cbic.201100781
101. Reese, C. B. *Org. Biomol. Chem.* **2005,** *3,* 3851–3868. doi:10.1039/b510458k
102. Caruthers, M. H. *J. Biol. Chem.* **2013,** *288,* 1420–1427. doi:10.1074/jbc.X112.442855
103. Roy, S.; Caruthers, M. *Molecules* **2013,** *18,* 14268–14284. doi:10.3390/molecules181114268
104. Sproat, B. S. RNA Synthesis Using 2′-O-(*tert*-Butyldimethylsilyl) Protection. In *Oligonucleotide Synthesis;* Herdewijn, P., Ed.; Humana Press: Totowa, NJ, 2005; pp 17–31.
105. Nagata, S.; Hamasaki, T.; Uetake, K.; Masuda, H.; Takagaki, K.; Oka, N.; Wada, T.; Ohgi, T.; Yano, J. *Nucleic Acids Res.* **2010,** *38,* 7845–7857. doi:10.1093/nar/gkq638
106. Sekine, M.; Kadokura, M.; Satoh, T.; Seio, K.; Wada, T.; Fischer, U.; Sumpter, V.; Lührmann, R. *J. Org. Chem.* **1996,** *61,* 4412–4422. doi:10.1021/jo952263v
107. Iwase, R.; Sekine, M.; Hata, T.; Miura, K. *Tetrahedron Lett.* **1988,** *29,* 2969–2972. doi:10.1016/0040-4039(88)85060-3
108. Zuberek, J.; Wyslouch-Cieszynska, A.; Niedzwiecka, A.; Dadlez, M.; Stepinski, J.; Augustyniak, W.; Gingras, A.-C.; Zhang, Z.; Burley, S. K.; Sonenberg, N.; Stolarski, R.; Darzynkiewicz, E. *RNA* **2003,** *9,* 52–61. doi:10.1261/rna.2133403
109. Koukhareva, I. I.; Lebedev, A. V. *Nucleosides, Nucleotides Nucleic Acids* **2004,** *23,* 1667–1680. doi:10.1081/NCN-200031492
110. Mikkola, S.; Salomaki, S.; Zhang, Z.; Maki, E.; Lonnberg, H. *Curr. Org. Chem.* **2005,** *9,* 999–1022. doi:10.2174/1385272054368402
111. Thillier, Y.; Decroly, E.; Morvan, F.; Canard, B.; Vasseur, J.-J.; Debart, F. *RNA* **2012,** *18,* 856–868. doi:10.1261/rna.030932.111
112. Kadokura, M.; Wada, T.; Seio, K.; Moriguchi, T.; Huber, J.; Lührmann, R.; Sekine, M. *Tetrahedron Lett.* **2001,** *42,* 8853–8856. doi:10.1016/S0040-4039(01)01941-4
113. Ohkubo, A.; Sasaki, K.; Noma, Y.; Tsunoda, H.; Seio, K.; Sekine, M. *Bioorg. Med. Chem.* **2009,** *17,* 4819–4824. doi:10.1016/j.bmc.2009.04.073
114. Ohkubo, A.; Kondo, Y.; Suzuki, M.; Kobayashi, H.; Kanamori, T.; Masaki, Y.; Seio, K.; Nagai, K.; Sekine, M. *Org. Lett.* **2013,** *15,* 4386–4389. doi:10.1021/ol401917r
115. Lewdorowicz, M.; Yoffe, Y.; Zuberek, J.; Jemielity, J.; Stepinski, J.; Kierzek, R.; Stolarski, R.; Shapira, M.; Darzynkiewicz, E. *RNA* **2004,** *10,* 1469–1478. doi:10.1261/rna.7510504
116. Jemielity, J.; Heinonen, P.; Lönnberg, H.; Darzynkiewicz, E. *Nucleosides, Nucleotides Nucleic Acids* **2005,** *24,* 601–605. doi:10.1081/NCN-200061922
117. Lewdorowicz, M.; Stepinski, J.; Kierzek, R.; Jemielity, J.; Zuberek, J.; Yoffe, Y.; Shapira, M.; Stolarski, R.; Darzynkiewicz, E. *Nucleosides, Nucleotides Nucleic Acids* **2007,** *26,* 1339–1348. doi:10.1080/15257770701533446
118. Mauer, J.; Luo, X.; Blanjoie, A.; Jiao, X.; Grozhik, A. V.; Patil, D. P.; Linder, B.; Pickering, B. F.; Vasseur, J.-J.; Chen, Q.; Gross, S. S.; Elemento, O.; Debart, F.; Kiledjian, M.; Jaffrey, S. R. *Nature* **2017,** *541,* 371–375. doi:10.1038/nature21022
119. Cozens, C.; Pinheiro, V. B.; Vaisman, A.; Woodgate, R.; Holliger, P. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109,* 8067–8072. doi:10.1073/pnas.1120964109

120. Wojtczak, B. A.; Warminski, M.; Kowalska, J.; Lukaszewicz, M.; Honcharenko, M.; Smith, C. I. E.; Strömberg, R.; Darzynkiewicz, E.; Jemielity, J. *RSC Adv.* **2016,** *6,* 8317–8328. doi:10.1039/C5RA25684D
121. Walczak, S.; Nowicka, A.; Kubacka, D.; Fac, K.; Wanat, P.; Mroczek, S.; Kowalska, J.; Jemielity, J. *Chem. Sci.* **2017,** *8,* 260–267. doi:10.1039/C6SC02437H
122. Warminski, M.; Kowalska, J.; Jemielity, J. *Org. Lett.* **2017,** *19,* 3624–3627. doi:10.1021/acs.orglett.7b01591
123. Gebhardt, A.; Habjan, M.; Benda, C.; Meiler, A.; Haas, D. A.; Hein, M. Y.; Mann, A.; Mann, M.; Habermann, B.; Pichlmair, A. *Nat. Commun.* **2015,** *6,* No. 8192. doi:10.1038/ncomms9192
124. Smietanski, M.; Werner, M.; Purta, E.; Kaminska, K. H.; Stepinski, J.; Darzynkiewicz, E.; Nowotny, M.; Bujnicki, J. M. *Nat. Commun.* **2014,** *5,* No. 3004. doi:10.1038/ncomms4004

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Polarization spectroscopy methods in the determination of interactions of small molecules with nucleic acids – tutorial

Tamara Šmidlehner[1], Ivo Piantanida[*1] and Gennaro Pescitelli[*2]

Review

Open Access

Address:
[1]Division of Organic Chemistry and Biochemistry, Ruđer Bošković Institute; P. O. Box 180, 10002 Zagreb, Croatia and [2]Department of Chemistry, University of Pisa, via Moruzzi 13, Pisa, Italy

Email:
Ivo Piantanida[*] - pianta@irb.hr; Gennaro Pescitelli[*] - gennaro.pescitelli@unipi.it

* Corresponding author

Keywords:
circular dichroism; emission-based dichroism; groove binding; intercalation; linear dichroism; non-covalent interactions; nucleic acids recognition; vibrational circular dichroism

Beilstein J. Org. Chem. **2018,** *14,* 84–105.
doi:10.3762/bjoc.14.5

Received: 28 September 2017
Accepted: 13 December 2017
Published: 08 January 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

The structural characterization of non-covalent complexes between nucleic acids and small molecules (ligands) is of a paramount significance to bioorganic research. Highly informative methods about nucleic acid/ligand complexes such as single crystal X-ray diffraction or NMR spectroscopy cannot be performed under biologically compatible conditions and are extensively time consuming. Therefore, in search for faster methods which can be applied to conditions that are at least similar to the naturally occurring ones, a set of polarization spectroscopy methods has shown highly promising results. Electronic circular dichroism (ECD) is the most commonly used method for the characterization of the helical structure of DNA and RNA and their complexes with ligands. Less common but complementary to ECD, is flow-oriented linear dichroism (LD). Other methods such as vibrational CD (VCD) and emission-based methods (FDCD, CPL), can also be used for suitable samples. Despite the popularity of polarization spectroscopy in biophysics, aside several highly focused reviews on the application of these methods to DNA/RNA research, there is no systematic tutorial covering all mentioned methods as a tool for the characterization of adducts between nucleic acids and small ligands. This tutorial aims to help researchers entering the research field to organize experiments accurately and to interpret the obtained data reliably.

Review

1. Introduction

Many biological molecules are chiral and chromophoric among which the most important examples include proteins and nucleic acids. Moreover, the chiral constituents of natural biopolymers are homochiral, e.g., (almost) exclusively L-amino acids and D-sugars are found. Among other properties, chiral chromophoric molecules absorb left circularly polarized light

(L-CP light) differently from right circularly polarized light (R-CP light). L-CP and R-CP light can be seen as the two components, rotating in opposite directions (anti-clockwise vs clockwise), of plane polarized light (PP, Figure 1). Thus, the use of circularly polarized light has led to the development of several spectroscopical methods for the study of chiral non-racemic molecules, including biopolymers [1-4]. Particularly important applications of these methods are found in structural studies of biomacromolecules [3]. For instance electronic circular dichroism (ECD), the most commonly used method, is indispensable in the structural studies of proteins and also intensively used in the characterization of the helical structure of DNA and RNA [5]. As nucleic acids are characterized by a dominant helical chirality and exhibit a rather small set of secondary structures, each characterized by a different polarization spectroscopy signature, they are convenient targets to monitor structural changes induced by outer stimuli. Linear dichroism (LD) is another type of polarization spectroscopy which does not require a chiral sample but rather an oriented one. It is based on the differential absorption of light polarized either parallel or perpendicular to a certain axis of orientation. This is also applicable to biomacromolecules such as nucleic acids, whose helices are normally elongated in a single direction [2]. In this context one of the most common approaches is to monitor the changes in the ECD or LD spectrum upon binding of a ligand to DNA or RNA [6].

Non-covalent interactions of small molecules (ligands) with DNA and RNA are of paramount interest because many biological processes, drugs and biochemical tools/probes rely on them [7,8]. Due to the possibly multifaceted nature of the complex formed between a small-molecule ligand and a large receptor (DNA/RNA), several complementary methods are needed for the accurate characterization of their interactions. Although the design and performance of experiments used for DNA/ligand interaction studies are common to the various techniques described here (and also to standard UV–vis, IR or fluorescence spectroscopy), there are several important differences related to the sensitivity of the methods, possible artifacts and the interpretation of the obtained results. Since some of the significant

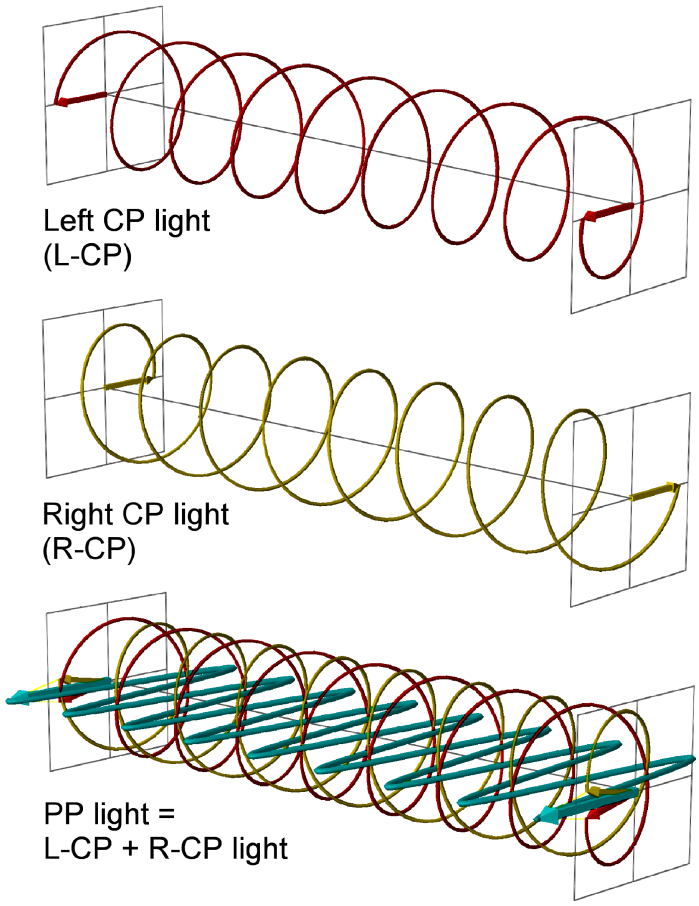


Figure 1: Plane polarized light (PP, cyan) as the sum of the left and right circularly polarized light (L-CP, red, and R-CP, yellow). L-CP is defined as such as an observer far from the source (on the right side of the figure) will see a vector describing an anti-clockwise circle.

facts are often neglected, the obtained results may fail to give reliable information or are misinterpreted.

With the hope that a detailed description of the experimental conditions and general rules for the interpretation of results will lead to more extensive applications and higher accuracy of the results, several reviews and tutorials were written in the past. For instance, a tutorial focused on circular dichroism as a tool for studies of non-covalent ligand/DNA interactions [9], provided a step-by-step protocol to the measurement. Very recently, a similar tutorial was published for LD experiments [10], and also a general outline on how to interpret CD and LD results with respect to the most common ligand/DNA binding modes was summarized on a more comparative basis [6]. Within the last decades, complementary methods to ECD also were developed, for instance, vibrational CD (VCD) was successfully applied to investigate DNA/ligand interactions [11]. Furthermore, fluorescence detected circular dichroism (FDCD) combines the advantages of both CD and fluorescence emission technique, which is ideal for the selective study of DNA ligands that strongly change fluorescence upon binding [3]. In a sense complementary to FDCD, also circularly polarized luminescence (CPL) is a chiroptical emission technique which has been employed in the same context for the first time recently [12].

The aim of this review is to summarize in one tutorial all required information and practical advice for performing experiments with the most common polarized spectroscopy methods: ECD, LD, VCD, FDCD and CPL with the goal of an accurate determination of interactions of small molecules with nucleic acids. Among the methods discussed, primary focus will be centered on the most frequently employed techniques, ECD and LD. Moreover, we will elaborate the interpretation of results not only for the most common DNA, RNA/ligand binding modes (intercalation, groove binding) but also for increasingly appearing ligand aggregates binding to polynucleotides (Figure 2). Many naturally occurring small molecules owe different biological activities due to aggregation, for instance, the close analogs netropsin (the single molecule in the DNA minor groove) and distamycin (the dimer in DNA minor groove) [13]. In addition, we will discuss the newest possibilities of computational analyses of the results as an outreach from the currently used empirical rules [6] for the determination of the ligand binding mode.

2. General aspects

ECD and LD are based on the phenomenon of light absorption by one or more chromophores in the UV–vis range of the electromagnetic spectrum, where electronic transitions occur. The application of these polarization spectroscopy methods for the study of a complex between a small molecule (ligand) and DNA or RNA can generally be divided into two wavelength ranges:

a) monitoring changes in the wavelength range where DNA and RNA absorb light (λ = 200–300 nm) and thus possess intrinsic spectra. Changes in the intrinsic spectral properties of DNA or RNA can often be correlated to a specific change in the secondary structure of the polynucleotide (see chapter 2.1.). However, if a ligand's chromophore also absorbs in this range, the deconvolution of all contributions is not trivial (see interpretation of results in the chapters of the corresponding methods);

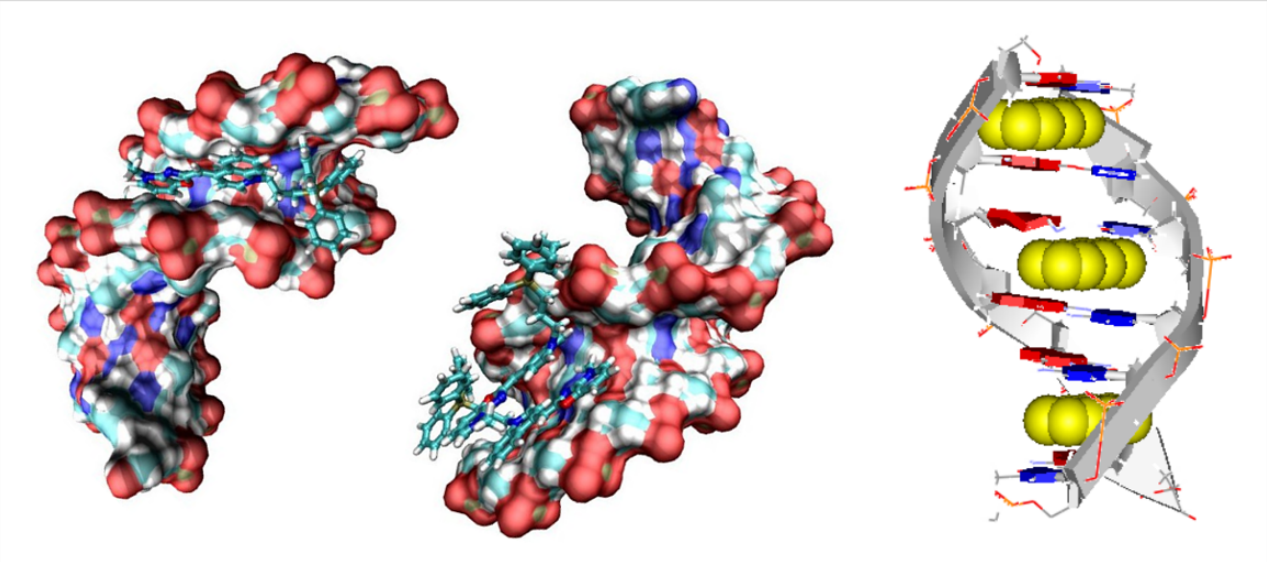


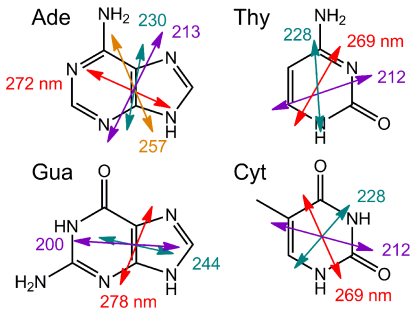
Figure 2: Schematic representation of ligand binding modes to DNA or RNA (left to right: DNA minor groove binding, minor groove aggregation, intercalation), which can be distinguished by polarization spectroscopy methods.

b) monitoring changes in polarization spectra at λ > 300 nm for ligands that have corresponding chromophores. This is the most common used approach as it monitors changes of only one species involved in the complex formation and thus the interpretation of results is simple and straightforward.

However, to design the polarization spectroscopy experiment accurately, it is essential that solutions of both the polynucleotide and the ligand are adequately prepared and characterized. For that reason, a short description of the most important issues in sample preparation is summarized in chapters 2.2 and 2.3.

2.1. Relation between DNA or RNA secondary structure and polarization spectroscopy

Nucleobases are achiral but nucleoside and nucleotide derivatives are optically active. The π→π* transitions of the bases (Scheme 1) contribute to the electronic circular dichroism mostly as a result of the chiral perturbation exerted by the sugar moiety. The ECD signals of single nucleosides/nucleotides are, accordingly, quite small. Coupling into polynucleotides, particularly in the double-stranded helix of DNA or RNA, introduces helical chirality, whereby the helical axis is almost perpendicular to the aromatic base-pair plane. In this situation, the ECD changes dramatically as a consequence of the so-called coupled oscillator or exciton coupling mechanism between the various π→π* transitions of regularly arranged chromophores [14] (Figure 3, top).



Scheme 1: Principal electronic transitions of nucleobases (uracil is similar to thymine, Thy). The arrows depict the polarization direction of each transition, and the length is roughly proportional to the relative intensity.

Chiroptical properties and ECD spectra of particular DNA or RNA sequences are therefore strongly dependent on the polynucleotide secondary structure [15], at variance to the common UV–vis spectra of the same samples (Figure 4). Of course, this fact has very important practical applications in monitoring the polynucleotide structure change caused by outer stimuli like ligand binding, pH changes, melting, etc. Intriguingly, most of

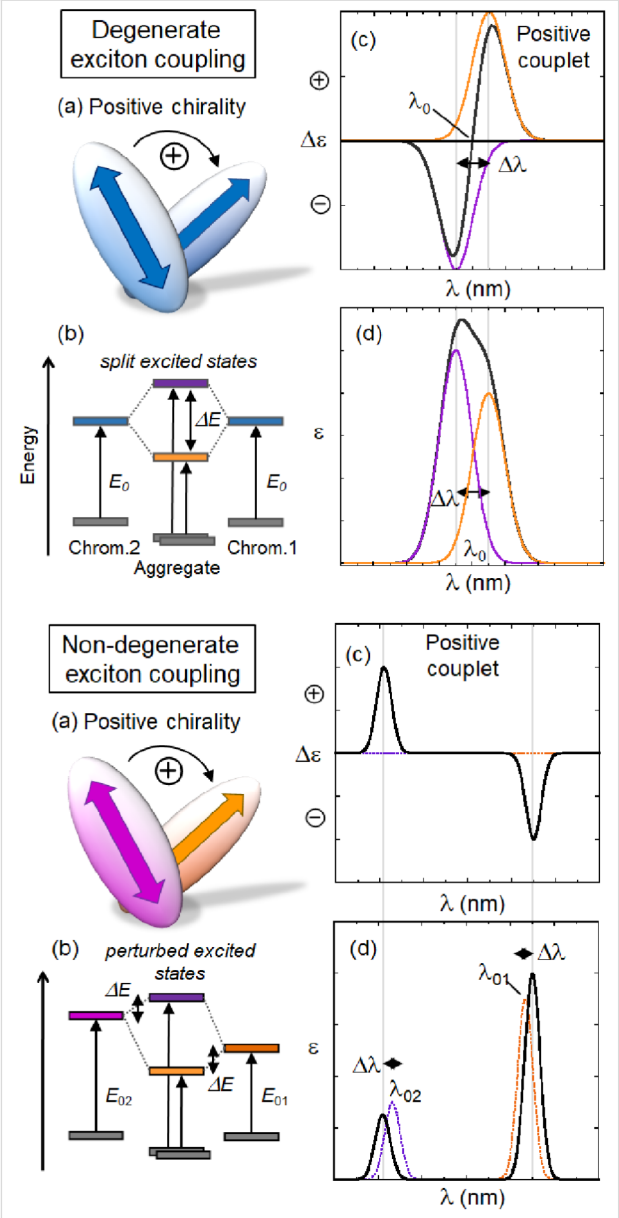


Figure 3: Exciton coupling mechanism in its degenerate (top) and non-degenerate version (bottom). a) Chirality defined by transition moments located on the two chromophores. A positive chirality corresponds to a clockwise direction when rotating the transition dipole in the front onto that in the back. b) Splitting of excited states due to exciton coupling. Notice the stronger effect in the degenerate case. c) ECD couplet originating from exciton coupling. The sign of the couplet (i.e., of its long-wavelength branch) is the same of the chirality (exciton chirality rule). In the degenerate case, the couplet is centered in correspondence of the chromophore transition at λ_0. In the non-degenerate case, each band is localized close to one chromophore transition (λ_{01} and λ_{02}). d) Absorption spectra originating from exciton coupling. The wavelength splitting Δλ is related to the energy splitting Δ*E*.

the single-stranded (ss) polynucleotides, if long enough, also show some extent of secondary chiral organization, thus also allowing a monitoring by ECD spectroscopy [16,17]. The non-degenerate coupled oscillator mechanism is also responsible for

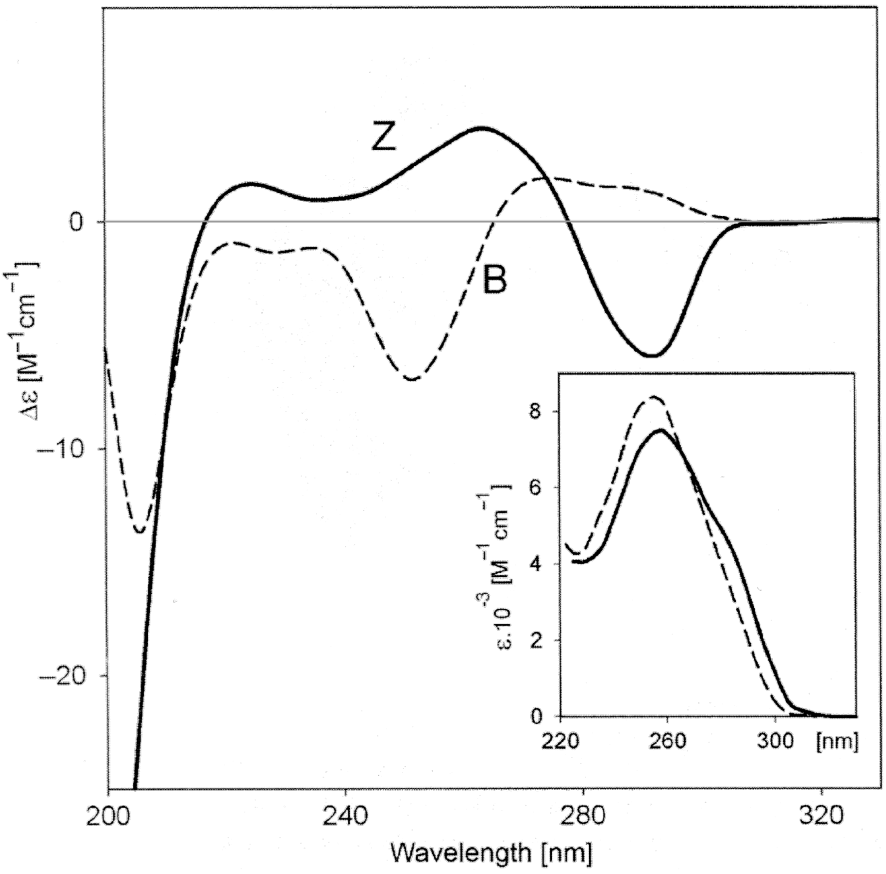


Figure 4: ECD and UV (inset) spectra of poly(dG-dC)–poly(dG-dC) in the B and Z-form, obtained in 59% and 67% aqueous trifluoroethanol, respectively. Adapted from [15].

generating ECD signals in correspondence with electronic transitions of achiral chromophores surrounded by a chiral environment. The most relevant example here is that of an achiral ligand bound to a nucleic acid, yielding a so-called induced circular dichroism (ICD) whose sign and magnitude are determined by the binding geometry. In this case, the exciton coupling is said to be non-degenerate (Figure 3, bottom). It must be stressed that the binding also determines a change in the standard absorption bands of the ligand, which can be similarly monitored to study the interaction [18].

Detailed information about various DNA/RNA structures and their ECD spectra is out of the scope of this review. A specific example concerning guanine quadruplexes is discussed in chapter 3.

2.2. DNA or RNA stock solution preparation

Here we discuss in detail only polynucleotides with lengths of more than 100 base pairs because oligonucleotides with known composition are much easier to dissolve, to check their structural properties and to determine exact concentrations. Also, short oligonucleotides can fail in representing a biologically significant structural model, because of heterogeneous binding sites due to the "capping" effect, whereby a ligand can bind similarly to the end base pairs and to a binding site along the helix [19].

Commercially available nucleic acids are usually sold as lyophilized white fibers and should be stored as defined by the producer before dissolving. Most of the nucleic acids are available as sodium salts and a variety of different sequences is available. *Calf thymus* DNA is the most commonly used DNA extracted from *calf thymus* tissue which consists of 41.9 mol % of G–C and 58.1 mol % A–T base pairs. Other natural DNAs with different base-pair ratios are not that easily available. However, a diversity of synthetic polynucleotides is commercially available, such as double-stranded nucleic acids with alternating or homo-base sequences as well as single-stranded homo-polynucleotides. Such synthetic DNA or RNA have well-defined structural properties and therefore are recommended for studies of small molecules targeting structural DNA or RNA selectivity.

All polynucleotides should be dissolved exactly as defined by the producer, usually in a buffer of defined ionic strength. A common mistake is disregarding the proposed procedure by using too low ionic strengths, which severely impacts the double-stranded helix folding and stability, and consequently strongly influences the corresponding ECD spectrum. Another common mistake with a significant impact on further experiments is preparing a too highly concentrated polynucleotide stock solution. Particularly, guanine-rich polynucleotides are prone to gelation and consequently to formation of non-homogeneous solutions. Thus, an optimal concentration for the stock solution of polynucleotides would be about 0.01 M. It is calculated as nucleobase/mol by using the molar extinction coefficient at the maximum of the wavelength absorbance (as defined by the supplier). In addition, dissolving polynucleotides is not an instantaneous process and usually takes several hours at room temperature. When DNA or RNA stock solutions are prepared, collecting the absorption and ECD spectra is the best way to check their quality.

Commonly, solutions of synthetic polynucleotides are stored at −20 °C and used without any further purification. However, DNAs isolated from natural resources (e.g., *ct*-DNA as the most common) are quite often exceptionally long. Therefore, upon soaking the dry fiber in the buffer, a very viscous solution of approximate 0.01 M concentration is obtained, that is hardly applicable for accurate titration experiments. In this case, it is advisable to sonicate the DNA solutions by sonication tips that are common in the biological laboratory, and a treatment for several times in 5-second periods affords a much less viscous solution and much shorter (about 100 base pairs) rod-like B-helical DNA fragments. There are several reasons why sonication of DNA isolated from natural resources is essential: the titrations with very viscous or even gelating non-sonicated DNA solutions are not accurate for practical reasons. The accuracy of automatic pipets and standard tips is designed for aqueous solutions of viscosities similar to that of pure water and, in addition, not all potential small-molecule binding sites on DNA-supercoiled fragments are accessible to the small molecule, therefore leading to erroneous site-size evaluation.

2.3. Ligand solution preparation and characterization

Preferably, the ligand should be dissolved in aqueous solution at a concentration of about 0.001 M. The prepared solution should be clear and homogeneous with no visible precipitation or opalescence. If the ligand is poorly soluble in water, other solvents can be used, but some of them may interfere by absorbing light and raising the wavelength cut-off, as well as by interacting to some extent with DNA or RNA structure. The most common and efficient solvent for this purpose is DMSO. It does not interfere significantly with DNA or RNA properties up to 0.1% v/v [20], while at higher DMSO quantities, its impact on ECD spectra of DNA or RNA should be corrected for. However, even at this small ratio mentioned above, DMSO significantly absorbs light at <240 nm, hampering an accurate collection of data in the short-wavelength region.

It is essential to check the dependence of the ligand's UV–vis spectrum on its concentration in the experimental conditions foreseen for further experiments. If the light absorbance is proportional to concentration, i.e., the Beer–Lambert law is fully respected, there is no intramolecular aggregation of the ligand. However, a non-linear response, usually hypochromic (i.e., a negative deviation from the Beer–Lambert law), supports the formation of ligand aggregates. If it is not possible to further dilute the sample to be used for polarization spectroscopy experiments, ligand solutions should be treated as a mixture of ligand plus its aggregates in thermodynamic equilibrium and analyzed accordingly. Similar to the absorption spectra for chiral ligands, ECD spectra at different concentrations can be collected, taking into account that the total absorbance of the sample is compatible with the sensitivity of the instrument. Again, aggregating ligands could form chiral aggregates with a characteristic exciton-coupled CD spectrum, which is by the way diagnostic for the supramolecular chirality [21,22].

Particular attention should be a paid to the UV–vis and ECD spectrum shapes and baselines as well as to an opalescence of the solution in the cuvette. An apparent positive drift of the baseline during ligand concentration increase in the cuvette, usually best seen in the long-wavelength range of the spectrum, is a clear evidence of precipitation or colloid/aggregate formation in the cuvette, which completely hampers further experiments.

The characterization of the thermal stability of the ligand solution by collecting it's UV–vis spectrum and if applicable, ECD spectrum at various temperatures, can give useful information about intra- or intermolecular interactions of a ligand. However, an ECD spectrum depends on the temperature in several ways other than affecting aggregation, for example by changing conformational populations. Therefore, a variable-temperature ECD spectrum is often not easy to interpret. Only after a detailed characterization of the ligand solution, experiments with polynucleotides can be performed by a series of different techniques described in following sections.

The electronic circular dichroism (ECD) will be discussed first as the most common technique with the least pitfalls. Thus it is most often appropriate for the characterization of the systems of interest. Then, LD follows as a complementary method used to

reveal the mutual orientation of the ligand chromophore with respect to the DNA/RNA helical axis. The other techniques mentioned in the Introduction (VCD, FDCD, and CPL) will be treated more shortly because of their relatively less widespread use in the context of interest. The particular aspects of data analysis and interpretation will be discussed for each method, while general aspects including computational approaches and comparison of the obtained data from several methods will be summarized in the last chapter.

3. Electronic circular dichroism (ECD)

Electronic circular dichroism (ECD) is the difference in absorption between left and right circularly polarized light (Figure 5). It provides information about the chiral species in solution which absorb light in the UV–vis range, due to transitions from the electronic ground state to one or more excited states. ECD is one of the most sensitive spectroscopic techniques for probing changes in the DNA or RNA binding mode of a ligand as a function of concentration and/or mixing ratios [3,5,6,9,18]. However, ECD relies on the difference in absorption between left and right circularly polarized light which makes it at least two orders of magnitude less sensitive than absorption spectroscopy [1,9].

ECD spectroscopy can give significant structural information based on electric and magnetic transition moments of adjacent chromophores and their mutual orientation. This is especially true in the presence of strong chromophores interacting through the coupled oscillator mechanism mentioned above. Consequently, based on the organization of DNA/RNA chromophores (base pairs) in a helical structure, ECD is primarily sensitive to the secondary structure of various nucleic acids and provides characteristic ECD spectra of nucleic acids as a result of base sequence and experimental conditions (see chapter 2.1.). As an example of how the exciton coupling mechanism determines the ECD spectra of polynucleotides, Figure 6 reports the case of guanine quadruplexes (G-quadruplexes) which are easy to analyze because of the presence of a single type of nucleobase and a rigid structure [23]. The ECD spectrum is dominated by the degenerate exciton coupling between the major transition of guanosine around 245 nm; the transition is long-axis polarized (Scheme 1) and, for a right-handed quadruplex arrangement, it defines a positive chirality between stacked guanines, which yields a positive couplet. Even in this sample case, ECD is, however, also affected by other contributions, e.g., the non-degenerate coupling between different guanosine transitions, long-range couplings between distant guanosines, etc.

More interestingly for our purpose, ECD is a useful method for probing conformational changes of nucleic acid upon ligand binding. Moreover, as also mentioned in chapter 2.1., achiral ligands upon binding to DNA or RNA can eventually acquire an induced CD (ICD) spectrum, especially when their transition moments are uniformly oriented with respect to the DNA/RNA binding site, which could give useful information about modes of interaction and binding geometry, as will be discussed below.

ECD spectroscopy shows several advantages for investigating DNA, RNA and their complexes with ligands with respect to other spectroscopic techniques (non-polarized spectroscopies, NMR, X-ray single crystal diffraction) [6,9]:

- The experiments are technically easy and comparatively quick to perform.
- The equipment for sample holding is broadly available, common to any other solution spectroscopy methods. Standard 1 cm quartz cuvettes for absorption and/or fluorimetric titrations can be used.
- The experiments are performed in solution (easily mimicking physiological conditions) with a concentration lower than that needed for NMR measurements. Moreover, ECD is a “fast” spectroscopy so that the typical problems relative to the NMR timescale are not encountered.

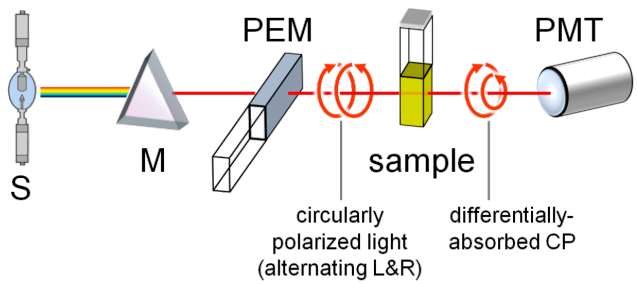


Figure 5: Schematic illustration of an ECD instrument. Legend: S, source; M, monochromator (wavelength selector); PEM, photoelastic modulator producing L-CP and R-CP light alternatively; PMT, photomultiplier (detector). A basic VCD instrument has a similar design, except that the monochromator is replaced by an interferometer for Fourier transform.

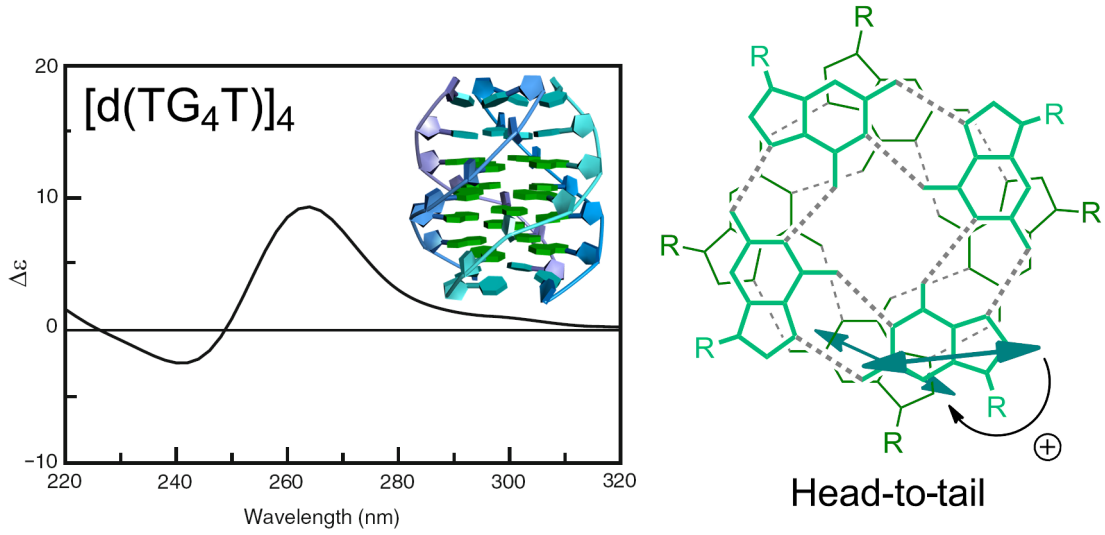


Figure 6: Exciton coupling in the G-quadruplex $[d(TG_4T)]_4$. For this sequence, the quadruplex is formed by four parallel strands and the quartets face each other in a head-to-tail arrangement. Reprinted in parts with permission from [22] to G.P. as an author of the article, copyright 2014, The Royal Society of Chemistry.

- Crystalline material is not required as for the X-ray single crystal diffraction technique.
- ECD spectroscopy is sensitive to a mutual orientation between the DNA/RNA axis and the orientation of the ligand transition moments. Thus it quite often yields detailed structural information about the studied complex. This is the most important advantage with respect to the use of non-polarization spectroscopy, e.g., absorption spectra.

However, some characteristics of the ECD instruments limit the experimental conditions. The most important limitations are: a) the total absorbance of samples at a chosen signal through the complete experiment should ideally be below 1 a.u. (higher absorbance values up to 1.8 a.u. may still lead to sufficiently accurate spectra, especially with last-generation instruments, but should be checked for reproducibility); b) induced (I)CD bands could have poor signal-to-noise ratios for, e.g., intercalators or for any weak ICD signals (the sensitivity of standard ECD instruments is $g \approx 10^{-5}$, where $g = \Delta Abs/Abs = \Delta\varepsilon/\varepsilon$); the common approach of collecting a number of scans to improve the signal-to-noise ratio is time consuming. Thus, for the collection of titration data necessary to obtain an accurate binding isotherm (10–90% of complex formed, an excess of DNA binding sites over ligand), the application of the ECD method may be limited by available material and long times. This is particularly valid if the studied system and its binding constant are unknown and repetition/redesign of the experiment is necessary.

Recently, Garbett, Ragazzon and Chaires [9] summarized a general protocol for the use of ECD for the simultaneous determination of the binding mode and binding affinity of ligand/DNA complexes. In the light of the limitations listed above, here we would present an alternative approach, whereby an unknown DNA (or RNA)/ligand system is characterized primarily by a set of common methods (UV–vis spectroscopy or fluorimetric titrations processed by non-linear fitting procedures [9,24], thermal denaturation experiments [25]). According to these results, a set of several ECD spectra is designed particularly for the target of interest. For instance, a) characterization of single molecule binding would be studied at large excess (10 to 50-fold) of DNA/RNA over ligand; b) ligand aggregation within DNA/RNA would be studied at excess of ligand over DNA/RNA; c) kinetics of binding would be studied at different temperatures and instrument response times; d) competition experiments between two ligands aiming for the same binding site would require a specific design; and so on.

3.1. Practical information

Cuvettes: quartz, preferably high-quality manufactured with precisely parallel walls (e.g., fluorimetric cuvettes) and minimal residual strain. Cylindrical cells are usually recommended for

ECD, but the standard cuvettes may also be used, if they pass the following test. Check of quality: collect the ECD spectrum of the buffer, rotate the cuvette 180° about its vertical axis and collect the spectrum again. The obtained two spectra should overlap well. If not, and especially in the presence of bands with intensities > 1 mdeg and opposite sign for the two measurements, the cuvette should be replaced. A broad range of path lengths is available (0.1 mm–5 cm) to ensure the optimal conditions of the total sample absorbance (about 0.8 Abs is recommended, with 1.5 Abs as upper limit). Commonly 1 cm, 3 mL cuvettes are used, filled with a sufficient amount of the sample in such way that the light beam does not pass close to the meniscus. Narrow cuvettes (widths < 0.5 cm) should have black-masked side-walls to prevent light beam reflection and absorption flattening artifacts [26]. The same cuvette should be used for all the measurements of one system (titration). A good cleaning of the cuvettes is essential (see producer manual or [9]), particularly for ligands that adhere to glass (e.g., porphyrins). For particularly adhering ligands, whose ICD spectra are measured in the vis range (>360 nm), the application of disposable plastic cuvettes can solve or at least minimize the problem. However, for dyes that strongly adhere to cuvette walls the collected chiroptical spectrum does not correspond to conditions in homogeneous solution and thus cannot be interpreted according to here given rules.

Buffers: DNA and RNA require specific buffers and ionic strengths to be reliably folded into their native secondary structure. Significantly lowering the ionic strength before or during the experiment (i.e., upon dilution by ligand solution addition) will not only impair buffering capacity but also will influence the DNA/RNA secondary structure and consequently change the corresponding ECD bands. Commonly, the experiments should be performed at pH 5–8 and ionic strength >1 mM.

Instrumental conditions: Instrument parameters differ among the instrument producers; here we will suggest those of Jasco J-810, but other instruments have corresponding ones. The parameters to control are: wavelength range, scan rate, bandwidth, averaging time and number of scans.

Wavelength range: Start: 220 nm (at lower wavelengths the total absorption of samples is high and requires specific conditions). End: maximum absorption of sample + 50 nm (the additional wavelength range is required to monitor if the baseline beyond the absorption region of the sample is flat after spectrum subtraction).

It is compulsory to monitor in parallel the two channels corresponding to ECD and HT (related to absorbance) signals. The ECD channel is often reported in mdeg, related to ellipticity. To convert mdeg in absorbance units, the following formula holds: θ (mdeg) = $\Delta A/33000$.

In biopolymers fields, the molar ellipticity [θ] is still used, defined as $[\theta] = (\theta/lc)$ where l is the path length in cm and c the molar concentration. We, however, recommend its conversion in molar circular dichroism: $\Delta\varepsilon = \Delta A/lc = [\theta]/3300$.

On the other hand, if the ECD data are used to monitor ligand binding, they can be directly plotted and interpolated in mdeg. In general, the slower the collection of data and the larger the number of scans are, the better is the quality of final data. However, it is better to average a number of faster scans than collect one scan very slowly. A more detailed analysis of parameters, which are necessary for the collection of high-quality data for binding isotherms is given in a specialized review [9]. Here we propose parameters allowing for the collection of one spectrum within a reasonable time and an acceptable quality. The following settings are recommended: Scanning speed: 50–100 nm/min (a faster speed is applicable for ECD spectra with wide bands); response (time constant): 1 s; bandwidth: 1 nm; accumulation of scans, which are averaged in one spectrum: 4, 8 or 16.

Of course, temperature accuracy is crucial in experiments involving DNA/RNA. Modern ECD instruments are equipped with a dedicated Peltier apparatus, which needs to be adjusted to the desired temperature.

3.2. Practical binding experiment/step by step procedure

- Put 2 mL of the buffer solution into a 1 cm path length cuvette (with a total volume of 3 mL) and record the spectrum of the buffer. The ECD spectrum of the buffer will not be zero, so the buffer background spectrum should be subtracted after each addition of DNA and sample. Remember the orientation of the cuvette in the holder and maintain it throughout the experiment.

- Add an aliquot of DNA stock solution to the cuvette to get c(polynucleotide) = 10–40 μM in the cuvette. Subtract the buffer spectrum from the DNA spectrum. The baseline in the range λ > 300 nm should be zero. If not, check for turbidity of the solution or other causes referred in chapter 2.

- Then, add aliquots of ligand stock solution (preferably at mM concentration) into the cuvette to cover the ratio $r_{[ligand]/[polynucleotide]}$ = 0.1–1 in 0.1 step size (within this range all major binding events shown in Figure 2 are usually detectable and can be subjected to detailed analysis). The incubation time prior to the collection of the spectrum depends on

the kinetics of binding, determined previously by other methods (UV–vis or fluorimetric titration). After each addition, the buffer background spectrum should be subtracted.

It takes about 10–20 min to collect one spectrum by multiple accumulations. During that time, the previously recorded spectra can be processed and compared to follow in real time the evolution of spectra. Any deviation from the baseline in the range where the studied ligand/DNA do not absorb light should be immediately inspected for turbidity (chapter 2). Furthermore, the Abs channel should be continuously monitored, taking care that the total absorbance of the sample does not exceed 1.5 or the limit suggested by the vendor.

3.3. Interpretation of results

3.3.1. Wavelength range < 300 nm: If the ligand does not absorb light within this range, the changes in the intrinsic ECD spectrum of the DNA and RNA can be correlated with a change in the secondary structure due to ligand binding. For instance, a significant decrease in the ECD spectrum over the whole 200–300 nm range indicates a disruption of helical chirality by intercalation or severe kinking of the helix by sterically demanding groove binders. At variance, if the ECD spectrum of the DNA and RNA does not change significantly, the biopolymer helical structure is preserved, suggesting a ligand groove binding, outer surface binding, or no/weak binding.

If a ligand does absorb light within this range, the observed changes cannot be unambiguously attributed to the DNA/RNA, because the ICD of the ligand (which is in principle unpredictable in intensity and sign), will combine with the intrinsic ECD spectrum of the polynucleotide. In special cases of poorly organized ss-polynucleotides, the intensity of the polynucleotide ECD bands may significantly increase while preserving the ECD band fingerprint. If so, one can presume a strong increase of polynucleotide helicity [27].

3.3.2. Wavelength range > 300 nm: In this range DNA and RNA do not absorb light. Thus, all ECD signals can be attributed to the ligand solely. The most straightforward case occurs with achiral ligands, whereby ligand binding to DNA/RNA results in an induced (I)CD spectrum.

If the ligand is chiral and has an intrinsic ECD spectrum (see chapter 2.3., characterization), the difference in the ECD spectrum caused by ligand binding is obtained by subtracting the intrinsic spectrum. However, it should be taken into account that the observed change in the ECD spectrum can have several origins: a) a change of the ligand's inherent chirality due to the structural changes caused by binding; b) an induced (I)CD as a result of ligand insertion into chiral binding sites; c) exciton coupling between multiple aggregated ligand molecules.

The appearance of one or more isodichroic points (the CD equivalent of isosbestic points) during ECD titration suggests the formation of one dominant type of DNA or RNA/ligand complex. Usually, a shift of isodichroic points can be observed during the titration, if it is pushed up to a large excess of ligand over DNA or RNA. This indicates that another, usually less preferred binding mode takes place. Isodichroic points can be found in the DNA absorbing range as well as in a range >300 nm where the ligand absorbs.

In the experimental conditions of one dominant binding mode (excess of DNA/RNA binding sites, clear-cut isodichroic points), there is a restricted number of possible outcomes for an achiral ligand which can be interpreted in a qualitative way to afford information about the preferred binding mode (Figure 7). The following conclusions stem from a large collection of experimental data and must be regarded as empirical [2,9]. However, they have also been substantiated by some of the theoretical approaches which will be mentioned in chapter 7.2.

3.3.3. Weak ICD (intensity of ICD band several times lower than the ECD band of DNA/RNA): a) Negative sign, non-linear relation of ICD intensity to ratio $r_{[ligand]/[DNA]}$ approaching saturation at about r = 0.2–0.3. This is a strong indication of intercalative binding, with the transition moment of the ligand oriented "parallel" to the long axis of adjacent base pairs (Figure 7, brown hue and bottom-right panel). It should be additionally supported by: i) a red-shift of the ligand absorption band; ii) at least moderate thermal stabilization of ds-polynucleotide; and iii) at least 10 μM binding constant (at common conditions, pH 5–8, I = 0.05–0.1, rt).

b) Positive sign, non-linear relation of ICD intensity to ratio $r_{[ligand]/[DNA]}$ approaching saturation at about r = 0.2–0.3. This indicates either groove binding with a loose orientation of the ligand with respect to the DNA axis or intercalative binding with the transition moment of the ligand oriented perpendicular to the long axis of adjacent base pairs (Figure 7, green hue). Additional experiments, preferably NMR are needed.

c) Negligible ICD intensity within signal-to-noise ratio, although other methods indicate a strong binding. Most likely intercalation takes place with the transition moment of ligand oriented at an angle to the long axis of adjacent base pairs, which happens to cancel positive and negative contribution. This could also be an indication of a ligand binding on the outer DNA/RNA surface through electrostatic interactions with the phosphate backbone. However, that is plausible only for highly

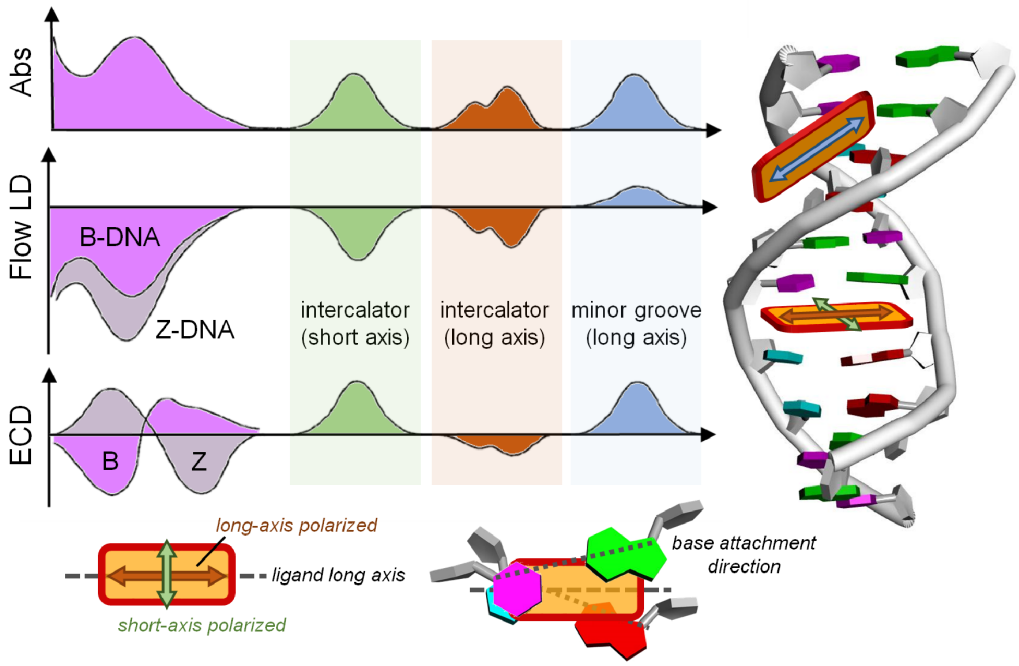


Figure 7: Empirical analysis of ECD and flow LD spectra to establish the preferred orientation of a ligand bound to DNA. We define the transition moment of the ligand oriented "parallel" to the long axis of adjacent base pairs, when it is halfway between the closest base attachment directions (bottom right panel). The LD is sensitive to the arrangement of the dye transition moment with respect to the helix axis, and to the overall orientation of the DNA double helix with respect to the flow direction (see Figure 9, below); Z-DNA is easier to orient than B-DNA because of a more extended elongation. ECD in the region <300 nm reflects the DNA conformation, in particular the mutual arrangement between nucleobases, which is different for right-handed B and left-handed Z-DNA. In the region >300 nm, the ligand ICD depends on the arrangement of the ligand chromophore with respect to the nucleobases. The sign of ICD is opposite for mutually orthogonal transition moments (long vs short axis) allied with an intercalated chromophore.

positive-charged ligands (at least four net positive charges present).

3.3.4. Strong ICD (intensity of ICD band similar or stronger than the CD bands of DNA/RNA): Usually of positive sign, strongly supports minor groove binding to DNA or major groove binding to ds-RNA (Figure 7, blue hue).

Exciton-coupled bisignate ICD bands: The appearance of exciton-coupled bisignate ICD bands (see chapter 2.1. and Figure 3, top) strongly support the aggregate binding along the polynucleotide, in which the ligand chromophores form an aggregate, with a well-defined supramolecular chirality. In this case, the ECD can still be said to be "induced" by the polynucleotide because it acts as a chiral template. An unambiguous interpretation of exciton-coupled bisignate ICD bands is not simple due to many different aggregation types which could take place upon DNA/RNA binding. For instance, at an excess of ligand over dominant DNA binding site (minor groove), surplus ligand molecules can form simple dimers within the DNA minor groove [28], which at even higher excesses over DNA can change to different, larger aggregates of H- or J-type [29]. In other cases the ligand can immediately form large helical arrays along the DNA or RNA as a dominant binding mode. For instance, a ligand upon binding to a polynucleotide can show ICD bands for several binding modes (Figure 8, left), whereby the first binding mode is dominant at $r < 0.2$, while the second binding mode (aggregation) is characterized by a new ICD band at 452 nm at $r > 0.2$. However, for another polynucleotide, the same ligand can give only one binding mode based on the uniform aggregation at excess of ligand over polynucleotide bases at $r > 0.4$ (Figure 8, right) [30].

3.3.5. CD bands out of expected wavelength range: In rare cases a ligand and DNA can form specific aggregates of very large sizes, which scatter light in the wavelength range where the components do not absorb light. In these conditions, a quite strong, well-defined and reproducible ECD band outside the expected range is observed. One example is the exceptionally strong CD spectrum of ψ-DNA [31,32], which is caused upon ligand addition (spermine) and by far exceeds the wavelength range at which the ligand and DNA absorb light (>300 nm). Such phenomenon is not related to standard chiroptical properties and will not be discussed here, but if a similar signal is obtained, it is advisable to refer to a set of methods dealing with DNA condensation (AFM, DLS).

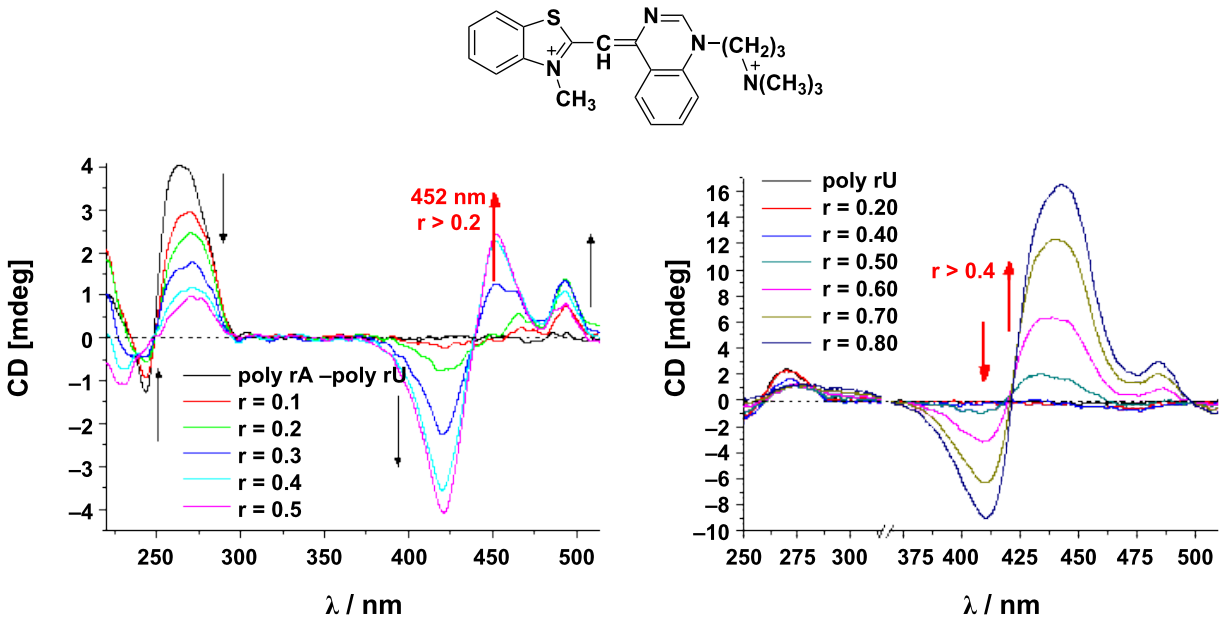


Figure 8: CD titration of poly rA-poly rU (left) and poly rU (right) (c(polynucleotide) = 2.0 × 10^{-5} mol dm^{-3}) with ligand L at molar ratios r = [L]/[polynucleotide] (pH 7.0, sodium cacodylate buffer, I = 0.05 mol dm^{-3}). Adapted from [30].

4. Linear dichroism, LD

Circular and linear dichroism spectroscopy (ECD and LD) are often used as complementary tools for the investigation of DNA or RNA structure, as well as for DNA/RNA interactions with various ligands [2,6,9,10]. Figure 7 nicely summarizes the complementarity of the methods.

As mentioned in the Introduction, the measurement of LD requires the sample to be oriented in a known way with respect to the polarization of radiation. Different orienting methods exist for different kinds of samples [2]. For macromolecules like polynucleotides which adopt a secondary structure with a preferred direction of elongation (that is, the polynucleotide helical axis), flow linear dichroism is the most suitable technique. In this case, DNA/RNA alignment is achieved by means of a special flow cell described below. Flow linear dichroism is defined as a difference in absorption of light polarized parallel and perpendicular to some reference axes taken by convention to be the flow direction along which the polynucleotide is at least partially aligned (Figure 9a). Flow LD is therefore a very useful method for the characterization of ds-DNA or ds-RNA conformation (base inclination), and of the flexibility and binding geometry of ds-DNA (RNA)/small molecule complexes [33,34].

The aforementioned flow cell is a cylindrical rotating Couette cell, where the liquid is subjected to a constant gradient over the annular gap between the rotating (inner cylinder) and the fixed coaxial outer cylinder (Figure 9a and c). The speed of the rotation should be adjusted in such way as to cause the orientation of the molecule and not a turbulent flow and 5000 rpm is an indicative figure. A necessary prerequisite for a macromolecule such as DNA is a minimum length of at least 1000 base pairs to be successfully oriented [10]. LD probes the orientation of base transition moments relative to the DNA helical axis. Thus, standard B-DNA whose base pairs are perpendicular to the helical axis will have the same spectral shape as normal absorption spectra but with a negative sign in the 240–280 nm region of absorption of DNA (Figure 10). The negative sign stems from the definition of LD: LD = $A_{\parallel} - A_{\perp}$, where $A_{\parallel}$ is the absorption for plane-polarized light parallel to the orientation axis, and $A_{\perp}$ is the absorption for plane-polarized light perpendicular to the orientation axis. For a given ligand, according to induced LD (ILD), its orientation with respect to the DNA can be determined by flow LD as long as the direction of the transition dipole moment within the ligand is known or can be established, and the orientation parameter S (Figure 9b) of the DNA is also known.

4.1. Practical information

A direct comparison of the experimental conditions and materials applicable in ECD and LD reveals that the LD method is more limited. Here are some reasons:

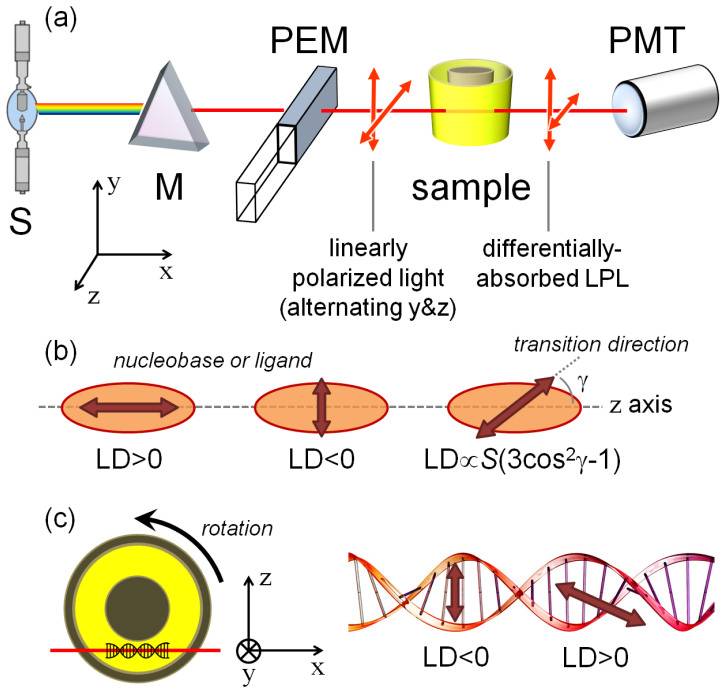


Figure 9: a) Schematic diagram of a flow LD instrument. b) Definition of the optical factor relating the direction of sample transition moment and the macroscopic sample orientation. The formula refers to a uniaxial sample; *S* is the orientation parameter, which quantifies the extent of effective sample orientation. c) Schematic illustration of a Couette flow LD experiment applied to a DNA/ligand complex.

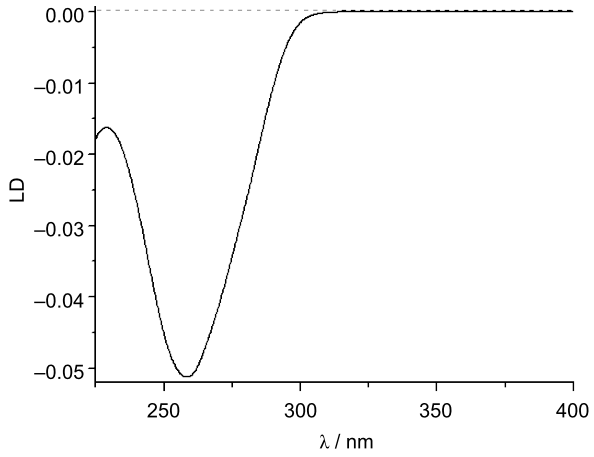


Figure 10: LD spectrum of *calf thymus* (*ct*) DNA (c = 2 × 10^{-4} M), pH 7, sodium cacodylate buffer, I = 0.05 M.

Sample absorbance: LD measures the absorbance only of the oriented parts of a sample (which can only be estimated), whereas most quantitative analyses require knowledge of the total absorbance.

Polynucleotide length and sample concentration: An efficient orientation along the flow requires a minimum length of the polynucleotide. It is estimated empirically that lengths over 1000 base pairs have a sufficiently high percentage of oriented DNA molecules to be measured at 100 µM concentrations (still 10 times higher than CD conditions). That precludes an LD application for any shorter polynucleotides, thus most of the synthetic DNA and RNA cannot be measured at c < 1 mM.

Light scattering: Non-flat baseline outside absorbing regions is a clear evidence of light scattering.

LD titrations are also collected as a function of the concentration of the compound. For that purpose, two types of flow cells are available: a cell with a total volume of 4 mL that allows adding stock solutions of the compound into the same cell, and 80 µL cuvettes for which every addition needs to be prepared separately.

For LD experiments, an additional spinning device is required. The baseline for LD experiments is measured on the solution in the non-spinning cuvette and should be recorded before each separate addition. All measurements should be done in the same cuvette which requires washing and careful drying

between the measurements, often implying a long experimental time.

DNA should be at least 1000 base pairs long and therefore while preparing the DNA sample solution, sonicating is not recommended. The concentration of the DNA in the cuvette should be 100 μM. For the experiments described in the next section, it is relevant to understand that elongated cylinder-like polynucleotide fragments are spatially oriented along the liquid flow obtained by rotation of the solid quartz cylinder within the fixed outer quartz cylinder, separated by 0.5 mm.

Instrument settings: If the LD device is attached to the ECD spectrometer, for LD experiments the same settings can be adjusted as for the ECD experiments. Make sure to choose the LD option while choosing the channel mode in Data mode and adjust the bandwidth to 2 nm. The LD unit should be connected to the outer spinning motor.

4.2. Practical/step by step

Steps, if using an LD unit with a total volume of 4 mL:

- Put 1 mL of a buffered solution of DNA (c = 200 μM) into the cell. To mix the solution in the cell, spin it by switching on the motor for a few seconds. Turn off the spinning and record the spectrum. Then, turn on the spinning up to 5000 rpm and record another spectrum. Subtract the non-spinning spectrum from the spinning spectrum to get the LD spectrum of DNA.
- Add an aliquot of the sample solution at a concentration which corresponds to the desired ratio $r_{[ligand]/[DNA]}$ followed by the aliquot of DNA stock solution that will compensate for dilution. Mix the solution in the cell for a few seconds, turn off the spinning device and record the spectrum. After completion of the non-spinning spectrum, turn on the motor and record the spinning spectrum. Subtract the non-spinning spectrum from the spinning spectrum to get the LD spectrum at the first ratio r. Repeat this for all desired ratios r. While the spinning mode is on, check if there is any bubbling of the solution and monitor all the channels for scattering.

Steps, if using an LD cuvette with a total volume of 80 μL:

For this cuvette, it is advisable to prepare each solution containing free DNA and the desired set of ratios $r_{[ligand]/[DNA]}$ separately in vials (with a total sample volume of 100 μL) and consequently transfer 80 μL of each sample into the cuvette, starting with free DNA, followed by samples from lowest to highest ratios r.

- Put 80 μL of a solution of DNA (c = 200 μM) into the cell. Mix the solution for a few seconds with spinning. Turn off the spinning motor and record the spectrum. Then, turn on the spinning up to 5000 rpm and record another spectrum. Subtract the non-spinning spectrum from the spinning spectrum to get the LD spectrum.
- After recording the first spectrum, the cuvette should be washed properly with redistilled water (3 times) and with ethanol (3 times, spectroscopic grade) followed by drying it using the air outlet.
- Into the dried cuvette, transfer 80 μL of a ligand/DNA mixture at a particular ratio r. Mix the solution, record the non-spinning spectrum and then the spinning spectrum. Repeat this for all samples.

4.3. Interpretation of the results

If a targeted biomacromolecule is spatially well-oriented in the sample, a small molecule which binds uniformly to identical, mutually independent binding sites of the biomacromolecules, will acquire an LD signal with an intensity proportional to the quantity of bound small molecules [33,34]. However, the intensity of the LD signal depends on several other factors such as: a) the orientation parameter of the DNA (S in Figure 9); b) the orientation of the bound molecule with respect to the macromolecule (Figure 7); c) the local direction of the transition dipole moment allied with the observed absorption band (Figure 7). Provided that at least factor S is known, one may in principle use LD measurements to estimate the angle by which a transition of a bound molecule is oriented relative to the axis of the DNA helix.

4.3.1. Changes in DNA absorbing region (<300 nm): The LD spectrum of ds-DNA/RNA can change upon binding of small molecules due to structural changes of the double helix. For instance, if upon small molecule binding the DNA/RNA double helix is shortened or kinked, it will orient less effectively and the LD signal will be reduced and/or changed in its shape [33,34]. Alternatively, a marked increase of the negative LD amplitude at 260 nm with the addition of a small molecule could imply that the DNA/RNA becomes better-oriented in flow, due to a stiffening of the ds-DNA/RNA structure [33,34].

4.3.2. Induced LD (>300 nm): When the compound is added to the DNA, and if the compound's absorbing region shows LD signals, the compound is binding in one or more specific orientations (rather than randomly along the backbone) [6,10,35]. The sign of the induced LD then can give information on the geometrical orientation of a bound compound to the DNA (Figure 7):

1. Negative ILD: Negative ILD in the compound's absorbing region indicates that the transition moment of the compound, allied with the observed absorption band, is perpendicular to the DNA axis (90°), which is consistent with intercalation [2,6].
2. Positive ILD: When a positive ILD is observed for the polarized band of the compound, this implies that at least the chromophoric portions that are associated with the observed transitions of the compound are not intercalated (inserted between the base pairs, thus being perpendicular to the DNA chiral axis). Instead, the direction of the chromophore transition moment (e.g., the chromophore long axis indicated by the blue double-arrow in Figure 7) is most probably oriented at about 45° to the DNA chiral axis, thus matching the minor groove of the DNA [6,10]. For a planar molecule, this implies non-intercalative binding, whereas, for many potentially nonplanar or flexible molecules, a partial insertion between base pairs cannot be excluded on this basis.
3. No or very small ILD: The compound is non-specifically oriented, either due to a mixed binding mode or externally bound to the phosphate backbone.

5. Vibrational circular dichroism, VCD

VCD is analogous to ECD in the IR region of the electromagnetic spectrum, where molecular vibrational transitions occur (between 4000 and 750 cm^{-1}) [3,36]. Recently, VCD has developed as a reliable spectroscopic method for the determination of the absolute configuration and conformational distribution of chiral molecules in solution [1,4,36]. The main drawback of VCD is the inherently small signal intensity, which is around 100-fold less intense than for ECD. Despite recent technical improvements, for accurate results highly concentrated samples are still needed. The insufficient solubility of DNA or RNA samples, as well as aggregation properties often hamper an application of VCD. Another severe problem is the need for an IR-transparent solvent, which hinders measurements in purely aqueous solutions, if not done in D_2O. The main advantage of VCD over ECD is that it does not need the presence of a conjugated aromatic chromophore, making VCD spectra much richer in bands than a typical ECD spectrum (just like IR spectra vs UV–vis spectra) [1,4,36].

Nowadays, density functional theory (DFT) allows a calculation of the VCD spectrum of a chiral molecule, which then can be compared with the measured VCD spectrum and by this the absolute configuration (AC) of a chiral molecule can be determined [37]. Furthermore, an oligonucleotide VCD spectrum calculation was described recently [38].

5.1. DNA VCD

In VCD spectroscopy, two major spectral regions can be monitored to probe the structures of nucleic acids. In the region between 1700–1600 cm^{-1}, the stretching modes of C=O, C=N and C=C of the nucleic acid bases occur. The vibrational bands appearing above ≈1650 cm^{-1} are associated with the C=O (due to different number of these groups found in different bases, nucleic acids with different base compositions will provide different VCD spectra) and those appearing below ≈1650 cm^{-1} are associated with the C=C stretching modes (with some contribution from C=N stretching as well) [39,40]. The VCD in this region is highly sensitive to the secondary structure of DNA/RNA and may be employed to assess secondary structures and to monitor their changes upon various stimuli [41]. Similar to the ECD case, such a correlation is often made on an empirical basis [39,40] but may be substantiated by theoretical approaches [38].

The other spectral region interesting for nucleic acids is 1250–1000 cm^{-1}, where the stretching modes of the phosphate group are found plus some vibrational modes of the sugar moieties. The VCD appearing in this region is less dependent on the base composition [42].

5.2. Practical information

Choice of solvent and sample cell: Although the analysis of biomolecule interactions is best performed in aqueous media, water is not the most suitable solvent for VCD due to its strong band ("scissoring" mode) found at 1650 cm^{-1}, overlapping with the C=O stretching from nucleic acids [35]. Therefore, highly pure deuterated water (D_2O) is used for VCD DNA measurements. The frequency range that is not accessible in D_2O is 1150–1450 cm^{-1} [39]. In that case, ordinary NaCl and KBr sample cells cannot be used due to their solubility in water. Instead, CaF_2 or BaF_2 cells are used in VCD measurements. CaF_2 allows measurements down to 1000 cm^{-1} while BaF_2 is used for measurements for which a wider spectral range down to ≈750 cm^{-1} is needed. Special care must be taken regarding storing and handling of the cell because of the slight hygroscopicity of BaF_2. Standard cells for VCD comprise two CaF_2 or BaF_2 windows separated by a Teflon spacer and sealed, with path lengths around 50–100 μm [36]. They have small volumes and must be filled carefully using syringes, making the addition of aliquots not practical for VCD.

Sample preparation and instrument adjustments: Sample concentration at different ratios r in pure D_2O should be prepared possibly under anhydrous conditions and inserted into the cells. IR absorbance adjusted to around 0.4 is optimal for VCD measurements [36]. This corresponds to a high sample concen-

tration, which is unfavorable for the measurement of rare biomolecules.

Contrary to ECD, VCD instruments use Fourier transform (FT) collection of spectra. The standard resolution is 4 cm^{-1} [37]. Because of the small signal-to-noise ratio, several thousand spectra need to be collected and averaged (at least 2000–4000). This requires long measurement times, usually 1 h or longer. VCD spectra must be baseline corrected by subtracting the solvent spectra from those of the samples. However, a baseline drift is expected to occur over long acquisition times.

VCD applications to study DNA or RNA/ligand binding: Because of several reasons which should be clear from the paragraphs above, VCD is much less employed than ECD in the study of adducts between DNA/RNA and ligands. Another reason is that, due to the frequently very complex pattern of bands seen in VCD spectra, a clear-cut division into regions where only the polynucleotide or only the ligand contribute is impossible, contrary to what happens for ECD. Therefore, we will briefly mention here a few illustrative applications. An extensive review has been published in 2009 by Urbanová [11].

By far the best investigated ligands by VCD are porphyrins. Studies with both natural and synthetic DNA allowed the authors to establish the dominant binding modes, helix distortion, and base-pair stabilization [43,44]. In these studies, the diagnostic signals were those associated with C=O stretching and interestingly enough, VCD was always used in combination with ECD. We wish to mention that the interaction of nucleic acids and metal porphyrins can also be studied by magnetic circular dichroism (MCD), a chiroptical spectroscopy based on the differential absorption of L-CP and R-CP light in the presence of a strong magnetic field oriented parallel to the direction of light [45]. Another well-studied ligand is the anticancer drug daunomycin, proving its preferred intercalation site [46]. Finally, we mention a study on cisplatin bound to a model DNA octamer, whose VCD spectra were simulated by density functional theory (DFT) calculations [47].

6. Emission-based polarized spectroscopies: Fluorescence detected circular dichroism (FDCD) and circularly polarized luminescence (CPL)

6.1. General

Emission-based spectroscopy methods such as fluorescence-detected circular dichroism (FDCD) and circularly polarized luminescence (CPL) combine the advantages of both chiroptical and fluorescence techniques, therefore, being sensitive to molecules which are both chiral and fluorescent [3,48].

The FDCD method is based on the collection of the differential fluorescence emission of a sample excited alternatively by left and right circularly polarized light (Figure 11a). Combining the conformational sensitivity and chiral specificity of ECD with the detection sensitivity and specificity of fluorescence, FDCD is far more specific and sensitive than standard transmission ECD [3,49]. CPL is the differential emission of left and right circularly polarized radiation from a chiral fluorescent sample irradiated with non-polarized light (Figure 11b). FDCD and CPL represent the chiroptical counterparts of fluorescence spectroscopy recorded in the excitation and emission mode, respectively. Therefore, while FDCD senses the geometry and properties of the ground state, CPL senses those of the lowest excited state [49].

FDCD was used for measuring the mixtures of a fluorophore and one or more chromophoric but non-fluorophoric species, whereby it was possible to determine the contribution of only the fluorescent probe in the system [49], like for instance, a fluorescent ligand associated with a biopolymer or a fluorophore inserted into non-fluorescent chiral biomolecules (DNA, RNA, proteins) [3]. In such cases, FDCD monitors specifically the fluorescent ligand, provided that no energy transfer occurs. Additional advantages are that it can be used on samples that are optically dense and/or highly scattering [50]. The latest applications of FDCD to detect ligand/nucleic acid interaction date back to the ’80 [51], mostly because of the technical difficulties explained below. On the contrary, CPL is still in its infancy regarding the same context. The first paper describing the interaction of *ct*-DNA with two fluorescent ligands has recently appeared [12]. CPL may complement ECD by providing specific information on the emission behavior of a fluorescent achiral ligand bound to DNA.

6.2. Some practical advice and guidelines for FDCD and CPL experiments

FDCD experiments can be done in a fluorescence cuvette with a 1 cm path length in an ECD instrument with a photomultiplier placed at 90° to the excitation beam to collect ECD and fluorescence from the sample simultaneously. The emission signal should be filtered by using a long-pass filter, or, alternatively, by an emission monochromator. The most important drawback of FDCD measurements is the presence of so-called polarization artifacts, related to the presence of fluorescence polarization (photoselection) [52]. These are expected to be especially important for nucleic acids and their adducts because of the size of the system. An ellipsoidal mirror surrounding the cell can be

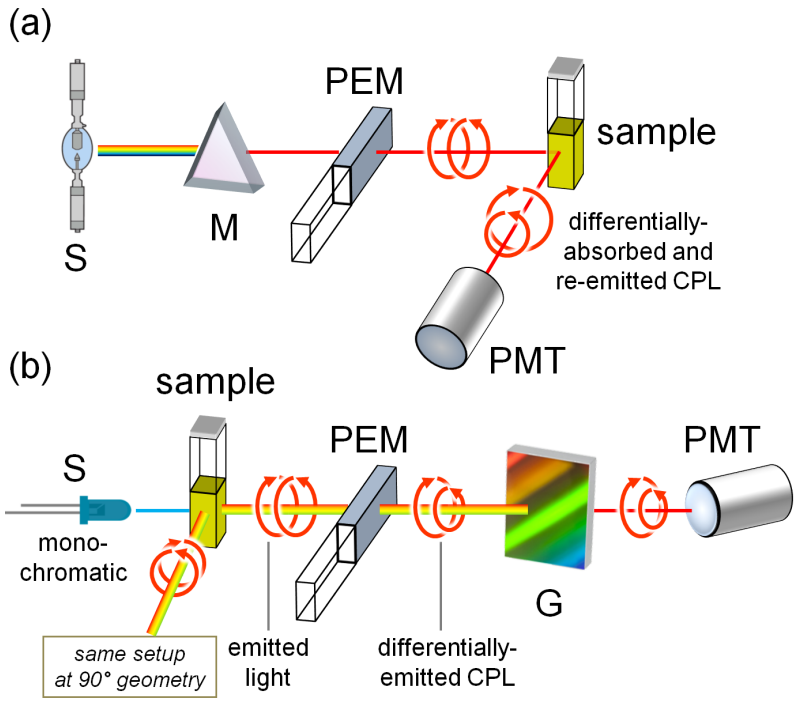


Figure 11: Schematic illustration of an FDCD a) and a CPL b) instrument. Legend: S, source; M, monochromator (excitation wavelength selector); G, diffraction grating (emission wavelength selector); PEM, photoelastic modulator; PMT, photomultiplier (detector). In CPL, both 0° and 90° setups are possible; in FDCD, the 90° setup is used.

used to improve the collection of fluorescence emission and to avoid artefacts [50].

CPL is also subjected to various kinds of artifacts, which need to be properly considered [12]. It must be stressed that the first commercial CPL instrument has only recently appeared on the market, sold by Jasco, Inc., but several home-made instruments are active.

7. Analysis of the obtained results

It is essential to remember that the here presented methods are often not sufficient by themselves for the accurate characterization of a ligand/polynucleotide complex, but usually are part of a broader set of methods (including fluorimetric and/or UV–vis titrations, thermal denaturation experiments, isothermal titration calorimetry (ITC) experiments, gel electrophoresis, etc.) [6]. Thus, the interpretation of the results should take into account all data.

7.1. Binding constant and binding ratio $n_{[ligand]/[polynucleotide]}$ determination

Traditionally, the affinity of the ligand to DNA or RNA is calculated from the Scatchard equation (McGhee, von Hippel formalism) [24], advisably by non-linear fitting of the experimental data (to avoid numerous problems, like large data-weighing errors, associated with linear Scatchard transformations) [53], and presuming that a single dominant binding site occurs for each ligand. An excellent protocol on how to organize CD experiments and data processing is given in reference [9]. One of the intriguing new approaches is GlobalFit processing [54,55], whereby the data from all titration experiments (e.g., CD, fluorescence, ITC, etc.), done at approximately similar conditions (concentration range, buffer), are processed simultaneously. The advantage is not only related to the use of a broad set of mutually independent methods, which can hardly have the same artifacts, but also the different sensitivity of each method for a particular complex formation response. For instance, a weak ICD band showing a non-linear change in titration experiments could be an error of the method, but if the non-linear fitting in GlobalFit procedure agrees with ITC or fluorimetric titration, then the ICD band can safely be attributed to a bound ligand.

However, if the ligand binds to DNA/RNA by multiple different binding modes within the same titration experiment (which is also detectable by other methods such as fluorescence and ITC), then Scatchard-based non-linear fitting cannot provide binding constants for all binding modes, but only for the preva-

lent binding at high excess of DNA/RNA over ligand, at which each ligand is bound independently. One of the possible approaches was demonstrated on the model (distamycin A and netropsin), showing how ECD data can be used in deconvolution of complex systems [56]. Also, singular value decomposition (SVD) analysis could help, since it allows a model-free determination of the number of linearly independent components in a given matrix of data [57]. Furthermore, several commercially available programs (e.g., HyperQuad [58] or Specfit [59]) offer versatile approaches to multicomponent spectra analysis, whereby the introduction of at least some known parameters (binding constants for dominant binding sites derived from GlobalFit or Scatchard calculations) allows the deconvolution of ECD or LD data to give ICD (ILD) bands for each type of complex. Unfortunately, the intensity and the sign of ICD are hardly predictable, and in such cases, the computational approach given below can help.

7.2. Computational approach – a focus on ECD spectra

The use of computational approaches in the study of polynucleotides and their adducts with ligands may appear a formidable task if one looks at the complexity of the system. In fact, the calculation approach must rely on relatively small models and ad hoc calculation strategies. In the past, the first theoretical descriptions of ECD spectra of polynucleotide/ligand complexes were based on the coupled-oscillator model [60-62] or on the so-called matrix-method approach (which also includes magnetic-allowed n–π* transitions) [63,64]. These theoretical works were essential to substantiate the relation between ECD spectra and the mode of binding (see chapter 2, interpretation of results), first established empirically. The coupled-oscillator and matrix methods are hybrid in the sense that they require the knowledge of the chromophore transitions either from the literature or from quantum mechanical (QM) methods. The transitions are then described by a set of parameters (position of the point dipoles, dipole orientation, dipolar strength) and are allowed to "interact" through classical electrostatic equations, for example, the Coulombic dipole/dipole potential. Although the basic fragmentation or many-body approximation is retained in some modern calculations methods for DNA/RNA [65-67], these latter rely explicitly on QM computations such as time-dependent density functional theory (TDDFT) [68]. Alternatively, more time-efficient QM schemes like the so-called sTDDFT (simplified TDDFT) and sTDA-xTB (simplified Tamm–Dancoff approximation with extended tight binding) can be directly employed on relatively large molecular systems like DNA [69].

The main aim of QM ECD calculations of complexes between DNA/RNA and ligands is to simulate an ECD spectrum to be compared with the experimental one. Of course, the absolute configuration is not an issue here, while the overall geometry is. Ultimately, then, the aim of such calculations is to substantiate the occurrence of a certain binding mode which has been (empirically) assessed by ECD and/or LD, or other spectroscopic techniques [70]. A typical ECD calculation will require three major steps: a) the selection of an appropriate model, b) the generation of an input structure, and c) the actual ECD calculation.

a) Selection of the model: It is obvious that a full QM calculation of a "real" system made of a large portion of DNA/RNA helices, counterions, solvating and surrounding water molecules, plus one or more bound ligand molecules, is not affordable even with state-of-the-art computers. Therefore, one must select an appropriate system to be handled at least with the simplified TDDFT or TDA approaches mentioned above, or even with standard TDDFT. This necessarily requires the DNA/RNA to be replaced with a short oligomer, which in turn means that one must focus only on the ICD in the region >300 nm (see chapter 3). ECD, and especially exciton coupling, is dominated by first-neighbor interactions [21], therefore one needs to consider the nucleobases closest to the ligand. For example, for an intercalated ligand one may consider the two base pairs involved in the intercalation and their immediate neighbors, thus focusing on a basepair tetramer (Figure 12). For groove binders, the number of base pairs will be necessarily higher. Counterions are often neglected while water molecules may be important especially if the ligand has the option to act as hydrogen-bond acceptor or donor.

b) Input structure: This is probably the most intriguing step of the process and can be divided into three sub-steps. In the first sub-step, one must obtain a molecular model of the DNA/RNA in the conformation evidenced by previous experiments. The most obvious source thereof would be PDB database. Then, a molecular model of the ligand molecule can be generated by using a molecular-modeling software. At this point, one must model the binding. A naive approach which can be tried as first choice would be to manually dock the ligand to the desired site (intercalation, minor or major groove) and then to relax its structure and that of the closest base pairs by molecular mechanics calculations with a good force field [71], by keeping the remaining structure fixed (first step in Figure 13). Most often, however, this approach is not accurate enough, and one must resort to the more typical methods used for modeling of polynucleotides. These include docking simulations and/or molecular dynamics (MD) simulations, possibly within a solvent cavity (second step in Figure 13). The description of these methods is outside the scope of the present tutorial and specialized books should be consulted [72,73].

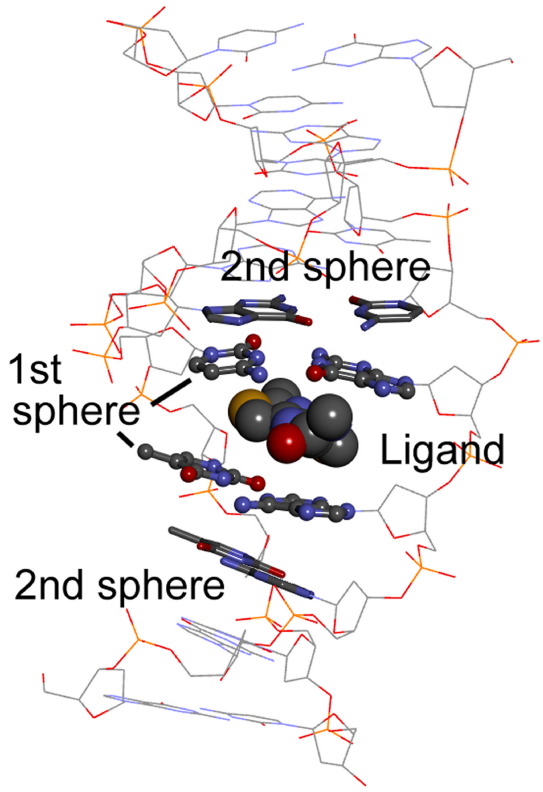


Figure 12: Minimal model to be considered to simulate the ICD spectrum of an intercalated ligand. ICD will be dominated by the intrinsic chirality of the ligand plus the nondegenerate exciton coupling with the most immediate nucleobase pairs (1st sphere). The following nucleobase pairs (2nd sphere) may possibly be neglected in the ECD calculations, but not in the input geometry generation.

In the second sub-step, one must simplify the structure obtained in the previous step (which would be by far too large for any QM approach) by cutting all unnecessary parts which are not expected to make a sizable contribution to the ligand ICD (see point a) above), neither to change much the structure of the ligand and of the surrounding nucleobases. Looking at Figure 12, one will preserve the ligand, the "first sphere" of nucleobases closest to the ligand, and the "second sphere" of nucleobases which are necessary to keep the ligand plus the first sphere in their position. Of course, two or more ligand molecules will be needed if multiple binding occurs (Figure 13) and relevant water molecules, hydrogen-bonded to the ligand, must be included.

In the third sub-step, the geometry of ligand(s) and the first sphere must be optimized with an accurate QM level, for example, DFT with a good functional such as M06-2X or ωB97X-D (third step in Figure 13). The second sphere can be kept frozen, or at least a part of it can be treated at a lower level of theory using the ONIOM approach. This optimization step is necessary to obtain a more accurate geometry of the chromophores which will be used for the ECD calculations, however, it shouldn't alter too much the overall geometry of the restricted model with respect to the whole system obtained after the first sub-step.

c) ECD calculations: Finally, ECD calculations will be run (last step in Figure 13) with one of the methods discussed above, in order of decreasing machine time usage: full TDDFT, sTDDFT, sTDA-xTB, TDDFT-based fragmentation approaches. For details, we refer the reader to the literature cited above. In the actual calculations, one may entirely neglect the second sphere or include it with some embedding approach like ONIOM. The QM system should be as small as possible, including the ligand(s) plus the closest nucleobases (bottom structure in Figure 13) [30], but even only a small ligand aggregate after the templating effect of DNA has been considered to build its geometry [74].

The calculated ECD spectrum resulting from the above sequence will be then compared with the experimental one, focusing only on the portion of the spectrum allied with the ligand transitions. It is important to compare not only the sign but the overall shape including the intensity. ECD is extremely sensitive to geometry, especially to the reciprocal arrangement of multiple chromophores. Therefore, in case of a good match,

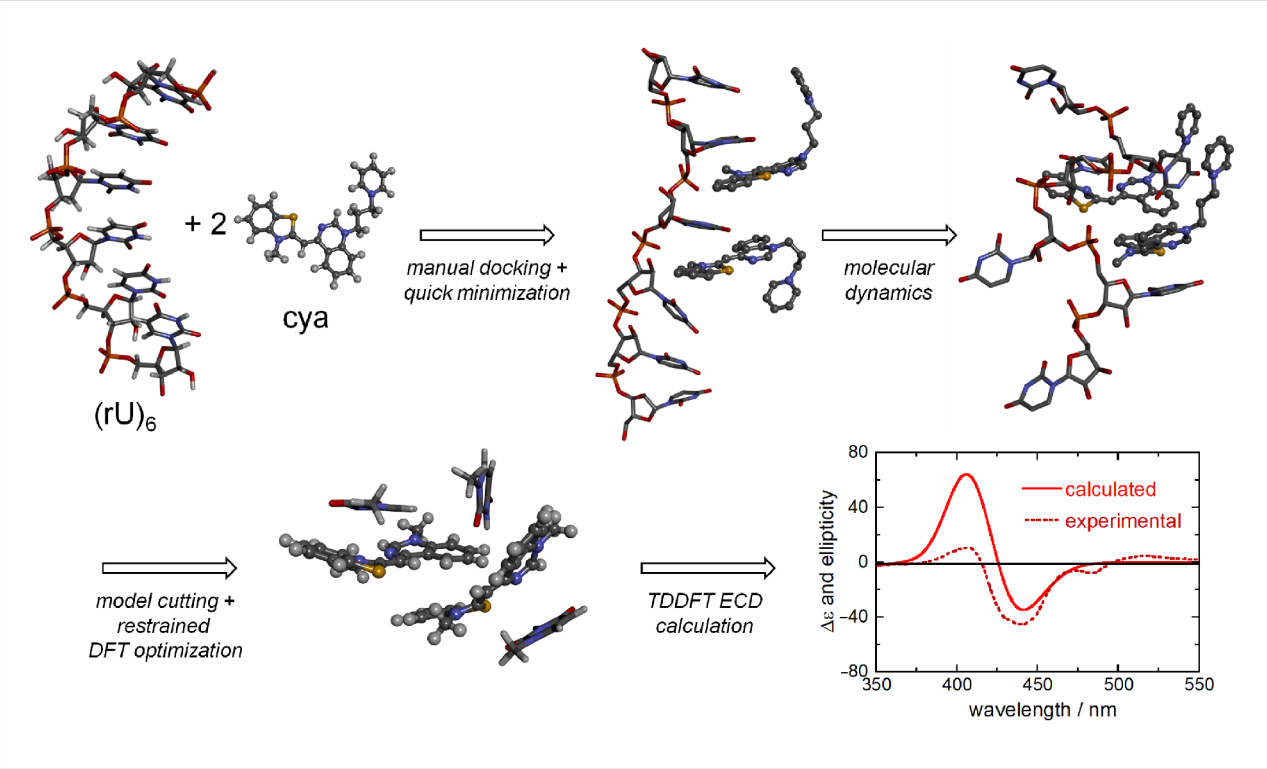


Figure 13: Schematic representation of the main steps for calculating the ICD spectrum of a ligand bound to a single-stranded RNA. In this case, a double binding of cyanine ligands (cya) was expected to occur in two consecutive binding sites. In the final model for ECD calculations, two “cya” molecules and the three closest nucleobases found after MD were considered. The ICD is exciton-like and is dominated by the exciton coupling between the cyanine ligands. Adapted from [30].

this will be a strong indication that the tested binding mode is the correct one. Otherwise, one can try if a different binding mode leads to a better agreement, with a trial-and-error tactic. On the other hand, it must be stressed that the described approach is based on several approximations, especially in the generation of the input structure. This means that a perfect agreement between the experimental and the calculated ECD may not necessarily be obtained even for the correct binding mode.

Conclusion

The family of chiroptical spectroscopy methods offers a toolbox of techniques which are very useful to characterize not only biomacromolecules such as nucleic acids, but also their interactions with small molecules. These techniques are based on the differential absorption or emission of circularly polarized light, thus they suffer from intrinsic lower sensitivity than their counterparts not based on polarized radiation. At the same time, however, they feature greatly enhanced structural sensitivity and selectivity toward chiral species. Therefore, they lend themselves as very practical tools to detect adducts between nucleic acids and small molecules, follow their evolution upon external stimuli, quantify their thermodynamics and kinetics, and provide information about the mode and geometry of binding. With the further development of computational methods, moreover, it is expected that spectra-to-structure relationships will be analyzed with greater accuracy and offer a detailed snapshot of the biomacromolecule/drug interaction at the molecular level.

Acknowledgements

IP and TS thank the Croatian Science Foundation project 1477 for support in several here cited results. G.P. thanks University of Pisa for PRA_2017_25 grant "Composti di metalli di transizione come possibili agenti antitumorali".

ORCID® iDs

Gennaro Pescitelli - https://orcid.org/0000-0002-0869-5076

References

1. Berova, N.; Polavarapu, P. L.; Nakanishi, K.; Woody, R. W., Eds. *Comprehensive Chiroptical Spectroscopy: Applications in Stereochemical Analysis of Synthetic Compounds, Natural Products, and Biomolecules;* Wiley: Hoboken, NJ, 2012; Vol. 2. doi:10.1002/9781118120392
2. Rodger, A.; Norden, B. *Circular Dichroism and Linear Dichroism, Circular Dichroism and Linear Dichroism;* Oxford University Press: New York, 1997.

3. Ranjbar, B.; Gill, P. *Chem. Biol. Drug Des.* **2009,** *74,* 101–120. doi:10.1111/j.1747-0285.2009.00847.x
4. Polavarapu, P. L. *Chiroptical Spectroscopy: Fundamentals and Applications;* CRC Press: Boca Raton, FL, 2016. doi:10.1201/9781315374888
5. Fasman, G. D. *Circular Dichroism and the Conformational Analysis of Biomolecules;* Plenum Press: New York, 1996. doi:10.1007/978-1-4757-2508-7
6. Eriksson, M.; Nordén, B. Linear and Circular Dichroism of Drug-Nucleic Acid Complexes. In *Methods in Enzymology;* Chaires, J. B.; Waring, M. J., Eds.; Academic Press: San Diego, 2001; Vol. 340, pp 68–98. doi:10.1016/S0076-6879(01)40418-6
7. Demeunynck, M.; Bailly, C.; Wilson, W. D. *Small Molecule DNA and RNA Binders: From Synthesis to Nucleic Acid Complexes;* Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2004.
8. Schneider, H. J., Ed. *Nucleic Acids as Supramolecular Targets;* Royal Society of Chemistry: London, 2013.
9. Garbett, N. C.; Ragazzon, P. A.; Chaires, J. B. *Nat. Protoc.* **2007,** *2,* 3166–3172. doi:10.1038/nprot.2007.475
10. Rodger, A.; Dorrington, G.; Ang, D. L. *Analyst* **2016,** *141,* 6490–6498. doi:10.1039/C6AN01771A
11. Urbanová, M. *Chirality* **2009,** *21* (Suppl. 1), E215–E230. doi:10.1002/chir.20803
12. Górecki, M.; Zinna, F.; Biver, T.; Di Bari, L. *J. Pharm. Biomed. Anal.* **2017,** *144,* 6–11. doi:10.1016/j.jpba.2017.02.010
13. Dervan, P. B. *Bioorg. Med. Chem.* **2001,** *9,* 2215–2235. doi:10.1016/S0968-0896(01)00262-0
14. Harada, N.; Nakanishi, K. *Circular Dichroic Spectroscopy - Exciton Coupling in Organic Stereochemistry;* University Science Books: Mill Valley, CA, 1983.
15. Kypr, J.; Kejnovská, I.; Bednářová, K.; Vorlíčková, M. Circular dichroism spectroscopy of nucleic acids. In *Comprehensive Chiroptical Spectroscopy: Applications in Stereochemical Analysis of Synthetic Compounds, Natural Products, and Biomolecules;* Berova,, N.; Polavarapu, P. L.; Nakanishi, K.; Woody, R. W., Eds.; Wiley: Hoboken, NJ, 2012. doi:10.1002/9781118120392.ch17
16. Egli, M.; Saenger, W. *Principles of Nucleic Acid Structure;* Springer Verlag: New York, 1983.
17. Cantor,, C. R.; Schimmel, P. R. *Biophysical Chemistry;* WH Freeman and Co.: San Francisco, 1980.
18. Biver, T. *Appl. Spectrosc. Rev.* **2012,** *47,* 272–325. doi:10.1080/05704928.2011.641044
19. Casagrande, V.; Alvino, A.; Bianco, A.; Ortaggi, G.; Franceschin, M. *J. Mass Spectrom.* **2009,** *44,* 530–540. doi:10.1002/jms.1529
20. Escara, J. F.; Hutton, J. R. *Biopolymers* **1980,** *19,* 1315–1327. doi:10.1002/bip.1980.360190708
21. Berova, N.; Di Bari, L.; Pescitelli, G. *Chem. Soc. Rev.* **2007,** *36,* 914–931. doi:10.1039/b515476f
22. Pescitelli, G.; Di Bari, L.; Berova, N. *Chem. Soc. Rev.* **2014,** *43,* 5211–5233. doi:10.1039/C4CS00104D
23. Randazzo, A.; Spada, G. P.; da Silva, M. W. Circular Dichroism of Quadruplex Structures. In *Quadruplex Nucleic Acids;* Chaires, J.; Graves, D., Eds.; Topics in Current Chemistry, Vol. 330; Springer: Berlin, Heidelberg, 2012. doi:10.1007/128_2012_331
24. Mc Ghee, J. D.; von Hippel, P. H. *J. Mol. Biol.* **1974,** *86,* 469–489. doi:10.1016/0022-2836(74)90031-X
25. Mergny, J.-L.; Lacroix, L. *Oligonucleotides* **2004,** *13,* 515–537. doi:10.1089/154545703322860825
26. Castiglioni, E.; Lebon, F.; Longhi, G.; Gangemi, R.; Abbate, S. *Chirality* **2008,** *20,* 1047–1052. doi:10.1002/chir.20639
27. Piantanida, I.; Palm, B. S.; Zinic, M.; Schneider, H.-J. *J. Chem. Soc., Perkin Trans. 2* **2001,** 1808–1816. doi:10.1039/b103214n
28. Tumir, L.-M.; Crnolatac, I.; Deligeorgiev, T.; Vasilev, A.; Kaloyanova, S.; Branilović, M. G.; Tomić, S.; Piantanida, I. *Chem. – Eur. J.* **2012,** *18,* 3859–3864. doi:10.1002/chem.201102968
29. Armitage, B. A. *Top. Curr. Chem.* **2005,** *253,* 55–76. doi:10.1007/b100442
30. Crnolatac, I.; Rogan, I.; Majić, B.; Tomić, S.; Deligeorgiev, T.; Horvat, G.; Makuc, D.; Plavec, J.; Pescitelli, G.; Piantanida, I. *Anal. Chim. Acta* **2016,** *940,* 128–135. doi:10.1016/j.aca.2016.08.021
31. Shin, Y. A.; Eichhorn, G. L. *Biopolymers* **1984,** *23,* 325–335. doi:10.1002/bip.360230211
32. Thomas, T. J.; Thomas, T. *Nucleic Acids Res.* **1989,** *17,* 3795–3810. doi:10.1093/nar/17.10.3795
33. Nordén, B.; Kubista, M.; Kurucsev, T. *Q. Rev. Biophys.* **1992,** *25,* 51–170. doi:10.1017/S0033583500004728
34. Nordén, B.; Kurucsev, T. *J. Mol. Recognit.* **1994,** *7,* 141–155. doi:10.1002/jmr.300070211
35. Patel, K. K.; Plummer, E. A.; Darwish, M.; Rodger, A.; Hannon, M. J. *J. Inorg. Biochem.* **2002,** *91,* 220–229. doi:10.1016/S0162-0134(01)00423-8
36. Nafie, L. A. *Vibrational Optical Activity;* Wiley: Chichester, 2011. doi:10.1002/9781119976516
37. He, Y.; Wang, B.; Dukor, R. K.; Nafie, L. A. *Appl. Spectrosc.* **2011,** *65,* 699–723. doi:10.1366/11-06321
38. Andrushchenko, V.; Bouř, P. *Chirality* **2010,** *22* (Suppl. 1), E96–E114. doi:10.1002/chir.20872
39. Polavarapu, P. L.; Zhao, C. *Fresenius' J. Anal. Chem.* **2000,** *366,* 727–734. doi:10.1007/s002160051566
40. Keiderling, T. A.; Lakhani, A. Conformational studies of biopolymers, peptides, proteins, and nucleic acids. A role for vibrational circular dichroism. In *Comprehensive Chiroptical Spectroscopy;* Berova, N.; Polavarapu, P. L.; Nakanishi, K.; Woody, R. W., Eds.; Wiley: Hoboken, NJ, 2012. doi:10.1002/9781118120392.ch22
41. Andrushenko, V.; Van De Sande, J. H.; Wieser, H. *Biopolymers* **2003,** *72,* 374–390. doi:10.1002/bip.10439
42. Wang, L.; Yang, L.; Keiderling, T. A. *Biophys. J.* **1994,** *67,* 2460–2467. doi:10.1016/S0006-3495(94)80734-9
43. Nový, J.; Urbanová, M.; Volka, K. *J. Mol. Struct.* **2005,** *748,* 17–25. doi:10.1016/j.molstruc.2005.03.011
44. Nový, J.; Urbanová, M. *Biopolymers* **2007,** *85,* 349–358. doi:10.1002/bip.20654
45. Mason, W. R. *A practical guide to magnetic circular dichroism spectroscopy;* John Wiley & Sons: Hoboken, NJ, 2007. doi:10.1002/9780470139233
46. Tsankov, D.; Maharaj, V.; Van de Sande, J. H.; Wieser, H. *Vib. Spectrosc.* **2007,** *43,* 152–164. doi:10.1016/j.vibspec.2006.06.018
47. Andrushchenko, V.; Wieser, H.; Bouř, P. *J. Phys. Chem. A* **2007,** *111,* 9714–9723. doi:10.1021/jp074572i
48. Riehl, J. P.; Muller, G. Circularly polarized luminescence spectroscopy and emission-detected circular dichroism. In *Comprehensive Chiroptical Spectroscopy;* Berova, N.; Polavarapu, P. L.; Nakanishi, K.; Woody, R. W., Eds.; Wiley: Hoboken, NJ, 2012.
49. Turner, D. H.; Tinoco, I., Jr.; Maestre, M. *J. Am. Chem. Soc.* **1974,** *96,* 4340–4342. doi:10.1021/ja00820a057
50. Maestre, M. F.; Reich, C. *Biochemistry* **1980,** *19,* 5214–5223. doi:10.1021/bi00564a010
51. Lamos, M. L.; Lobenstine, E. W.; Turner, D. H. *J. Am. Chem. Soc.* **1986,** *108,* 4278–4284. doi:10.1021/ja00275a007

52. Nehira, T.; Tanaka, K.; Takakuwa, T.; Ohshima, C.; Masago, H.; Pescitelli, G.; Wada, A.; Berova, N. *Appl. Spectrosc.* **2005,** *59,* 121–125. doi:10.1366/0003702052940459
53. Klotz, I. R. *Ligand–Receptor Energetics;* John Wiley & Sons, Inc: New York, 1997.
54. Zhao, H.; Schuck, P. *Anal. Chem.* **2012,** *4,* 9513–9519. doi:10.1021/ac302357w
55. Zhao, H.; Piszczek, G.; Schuck, P. *Methods* **2015,** *76,* 137–148. doi:10.1016/j.ymeth.2014.11.012
56. Lah, J.; Vesnaver, G. *J. Mol. Biol.* **2004,** *342,* 73–89. doi:10.1016/j.jmb.2004.07.005
57. Caesar, C. E. B.; Johnsson, R.; Ellervik, U.; Fox, K. R.; Lincoln, P.; Nordén, B. *Biophys. J.* **2006,** *91,* 904–911. doi:10.1529/biophysj.105.080564
58. Gans, P.; Sabatini, A.; Vacca, A. *Talanta* **1996,** *43,* 1739–1753. doi:10.1016/0039-9140(96)01958-3
59. Maeder, M.; Zuberbuehler, A. D. *Anal. Chem.* **1990,** *62,* 2220–2224. doi:10.1021/ac00219a013
60. Schipper, P. E.; Rodger, A. *J. Am. Chem. Soc.* **1983,** *105,* 4541–4550. doi:10.1021/ja00352a007
61. Schipper, P. E.; Nordén, B.; Tjerneld, F. *Chem. Phys. Lett.* **1980,** *70,* 17–21. doi:10.1016/0009-2614(80)80051-0
62. Kubista, M.; Åkerman, B.; Nordén, B. *J. Phys. Chem.* **1988,** *92,* 2352–2356. doi:10.1021/j100319a049
63. Lying, R.; Rodger, A.; Nordén, B. *Biopolymers* **1991,** *31,* 1709–1720. doi:10.1002/bip.360311405
64. Lyng, R.; Rodger, A.; Nordén, B. *Biopolymers* **1992,** *32,* 1201–1214. doi:10.1002/bip.360320910
65. Di Meo, F.; Pedersen, M. N.; Rubio-Magnieto, J.; Surin, M.; Linares, M.; Norman, P. *J. Phys. Chem. Lett.* **2015,** *6,* 355–359. doi:10.1021/jz502696t
66. Padula, D.; Jurinovich, S.; Di Bari, L.; Mennucci, B. *Chem. – Eur. J.* **2016,** *22,* 17011–17019. doi:10.1002/chem.201602777
67. Loco, D.; Jurinovich, S.; Di Bari, L.; Mennucci, B. *Phys. Chem. Chem. Phys.* **2016,** *18,* 866–877. doi:10.1039/C5CP06341H
68. Srebro-Hooper, M.; Autschbach, J. *Annu. Rev. Phys. Chem.* **2017,** *68,* 399–420. doi:10.1146/annurev-physchem-052516-044827
69. Seibert, J.; Bannwarth, C.; Grimme, S. *J. Am. Chem. Soc.* **2017,** *139,* 11682–11685. doi:10.1021/jacs.7b05833
70. Holmgaard List, N.; Knoops, J.; Rubio-Magnieto, J.; Idé, J.; Beljonne, D.; Norman, P.; Surin, M.; Linares, M. *J. Am. Chem. Soc.* **2017,** *139,* 14947–14953. doi:10.1021/jacs.7b05994
71. Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portells, G.; Battistini, F.; Gelpí, J. L.; Gonzáles, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. *Nat. Methods* **2016,** *13,* 55–58. doi:10.1038/nmeth.3658
72. Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide;* Springer: New York, 2002. doi:10.1007/978-0-387-22464-0
73. Šponer, J.; Lankaš, F. *Computational studies of RNA and DNA;* Springer: New York, 2006. doi:10.1007/978-1-4020-4851-3
74. Maj, M.; Jeon, J.; Góra, R. W.; Cho, M. *J. Phys. Chem. A* **2013,** *117,* 5909–5918. doi:10.1021/jp309807y

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Fluorescent nucleobase analogues for base–base FRET in nucleic acids: synthesis, photophysics and applications

Mattias Bood[‡1], Sangamesh Sarangamath[‡2], Moa S. Wranne[2], Morten Grøtli[1] and L. Marcus Wilhelmsson[*2]

Review

Open Access

Address:
[1]Department of Chemistry and Molecular Biology, University of Gothenburg, SE-412 96 Gothenburg, Sweden and [2]Department of Chemistry and Chemical Engineering, Chemistry and Biochemistry, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

Email:
L. Marcus Wilhelmsson[*] - marcus.wilhelmsson@chalmers.se

* Corresponding author ‡ Equal contributors

Keywords:
B-to-Z-DNA transition; fluorescent base analogues; FRET; netropsin; nucleic acid structure and dynamics; quadracyclic adenines; tricyclic cytosines; Z-DNA

Beilstein J. Org. Chem. **2018,** *14,* 114–129.
doi:10.3762/bjoc.14.7

Received: 08 November 2017
Accepted: 22 December 2017
Published: 10 January 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

Förster resonance energy transfer (FRET) between a donor nucleobase analogue and an acceptor nucleobase analogue, base–base FRET, works as a spectroscopic ruler and protractor. With their firm stacking and ability to replace the natural nucleic acid bases inside the base-stack, base analogue donor and acceptor molecules complement external fluorophores like the Cy-, Alexa- and ATTO-dyes and enable detailed investigations of structure and dynamics of nucleic acid containing systems. The first base–base FRET pair, tC^{O}–tC_{nitro}, has recently been complemented with among others the adenine analogue FRET pair, qAN1–qA_{nitro}, increasing the flexibility of the methodology. Here we present the design, synthesis, photophysical characterization and use of such base analogues. They enable a higher control of the FRET orientation factor, κ^2, have a different distance window of opportunity than external fluorophores, and, thus, have the potential to facilitate better structure resolution. Netropsin DNA binding and the B-to-Z-DNA transition are examples of structure investigations that recently have been performed using base–base FRET and that are described here. Base–base FRET has been around for less than a decade, only in 2017 expanded beyond one FRET pair, and represents a highly promising structure and dynamics methodology for the field of nucleic acids. Here we bring up its advantages as well as disadvantages and touch upon potential future applications.

Review

Introduction

The importance of nucleic acid structure and dynamics in the understanding of vital processes in living organisms has led to the development of a large number of techniques for such investigations. Among the most significant ones are NMR [1] and X-ray crystallography [2]. Both techniques offer a high structure resolution and NMR can also provide information on dynamics. However, there are occasions where NMR and X-ray crystallography suffer from drawbacks: the sample amount

requirement and biomolecular size restriction for NMR and the difficulties in obtaining crystals and the obvious lack of solution dynamics for X-ray crystallography. An important method for biomolecular structure and dynamics investigations that complements NMR and X-ray, normally at lower resolution, is Förster resonance energy transfer (FRET) [3,4]. FRET and especially single-molecule FRET (as an effect of a low number of biomolecules under study) has the advantage of enabling structure and dynamics investigations in living cells [3,5,6]. FRET is a process that depends on the radiationless energy transfer between a donor and an acceptor molecule [7]. The reason that it can be used as a structure and dynamics technique is that it depends heavily on the distance and relative orientation between the donor and acceptor. Typical distances that can be monitored range between 15–90 Å which well match the dimensions of biomolecules. The efficiency of an energy-transfer process (E, between 0 and 100%) can be established using either steady-state or time-resolved fluorescence spectroscopy by comparing fluorescence properties with and without the acceptor molecule present. This efficiency (E) depends on the distance (R_{DA}) between the donor and acceptor as described in Equation 1:

$$E = {R_0}^6 \big/ ({R_0}^6 + {R_{DA}}^6) \quad (1)$$

where R_0 is the Förster distance (Equation 2), a characteristic distance of the donor–acceptor pair at which the energy transfer efficiency (E) is 50%.

$$R_0 = 0.211\left(J_{DA}\kappa^2 n^{-4}\Phi_D\right)^{1/6} \text{ in Å} \quad (2)$$

As can be seen in Equation 2 the Förster distance depends on the quantum yield of the donor (Φ_D), the donor/acceptor spectral overlap integral (J_{DA}, overlap between energies of donor emission and acceptor absorption envelope), the refractive index of the medium (n), and importantly the geometric factor (κ, Equation 3). This factor takes the relative orientation of the transition dipole moments of the donor and acceptor into account and, thus, introduces an orientation dependence to R_0 and consequently also to the energy transfer efficiency, E. The orientation factor, which ranges between 0 and 4, is governed by Equation 3:

$$\kappa = \mathbf{e}_1 \cdot \mathbf{e}_2 - 3(\mathbf{e}_1 \cdot \mathbf{e}_{12})(\mathbf{e}_{12} \cdot \mathbf{e}_2) \quad (3)$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the unit vectors of the donor and acceptor transition dipole moments and $\mathbf{e}_{12}$ the unit vector between their centers (see Figure 1a; κ can also be expressed using the angles in Figure 1a: $\kappa = cos\ \varphi - 3cos\ \theta\ cos\ \omega$).

In most FRET applications, an orientation factor κ^2 of 2/3 is used. This is the correct value for freely rotating, isotropic donor and acceptor molecule orientations (Figure 1b). With an isotropic orientation of the donor and acceptor throughout the experiment the energy transfer efficiency (E) is directly related to the distance and the technique becomes a spectroscopic ruler. Such an assumption is often made, both correctly and incorrectly [8-10], for covalently attached external nucleic acid fluorophores like Cy-, Alexa- and ATTO-dyes. This provides a powerful means for measuring long distance ranges (typically 35–90 Å) in nucleic acid-containing systems. However, with the free rotation of the donor and the acceptor the ability of FRET to monitor changes in orientation between them is also lost. With virtually static donor and acceptor molecules (Figure 1c) κ^2 can be used to improve the structure resolution via the introduction of orientation information, i.e., FRET will also work as a spectroscopic protractor. Several investigations, including the ones by Tor et al. [11], Lewis et al. [12] and Lilley et al. [13], have taken significant steps in the direction of introducing orientational information into nucleic acid FRET. Recently, our group took this progress one step further and introduced base–base FRET [14], where the donor and acceptor molecules are nucleobase analogues [15,16]. With the donor and acceptor molecules rigidly stacked in the base-stack of the nucleic acid (Figure 1c) this approach provides highly accurate orientation information and has the potential to increase the structure and dynamics information obtained in a nucleic acid FRET experiment. Later Asanuma et al. introduced base-stacked aromatic moieties [17], not working as nucleobase analogues, which also can be used to provide information about orientation.

In this review we will focus on the FRET between fluorescent base analogues, i.e., base–base FRET, the theory behind it, the increased accuracy in orientation factor κ^2 as an effect of their position inside the base-stack, other advantages and disadvantages compared to FRET in nucleic acids using external fluorophores like Cy-, Alexa- and ATTO-dyes as well as finally summarize some of its recent applications. The field started less than a decade ago with the introduction of the first fluorescent nucleobase analogue FRET pair, tC^O–tC_{nitro}, and we divide this review into three parts: the first one dealing with the synthesis of the key players of base–base FRET, i.e., the base analogue donor and acceptor molecules, the second one dealing with their photophysical properties and the third one dealing with their application in studying nucleic acid-containing systems.

Synthesis of fluorescent base analogues

The development of synthesis methods of nucleobase analogues remains a challenge. This is mainly due to the presence of multiple reactive functional groups both on the nucleobase as well as the sugar moiety and requires the introduction of orthog-

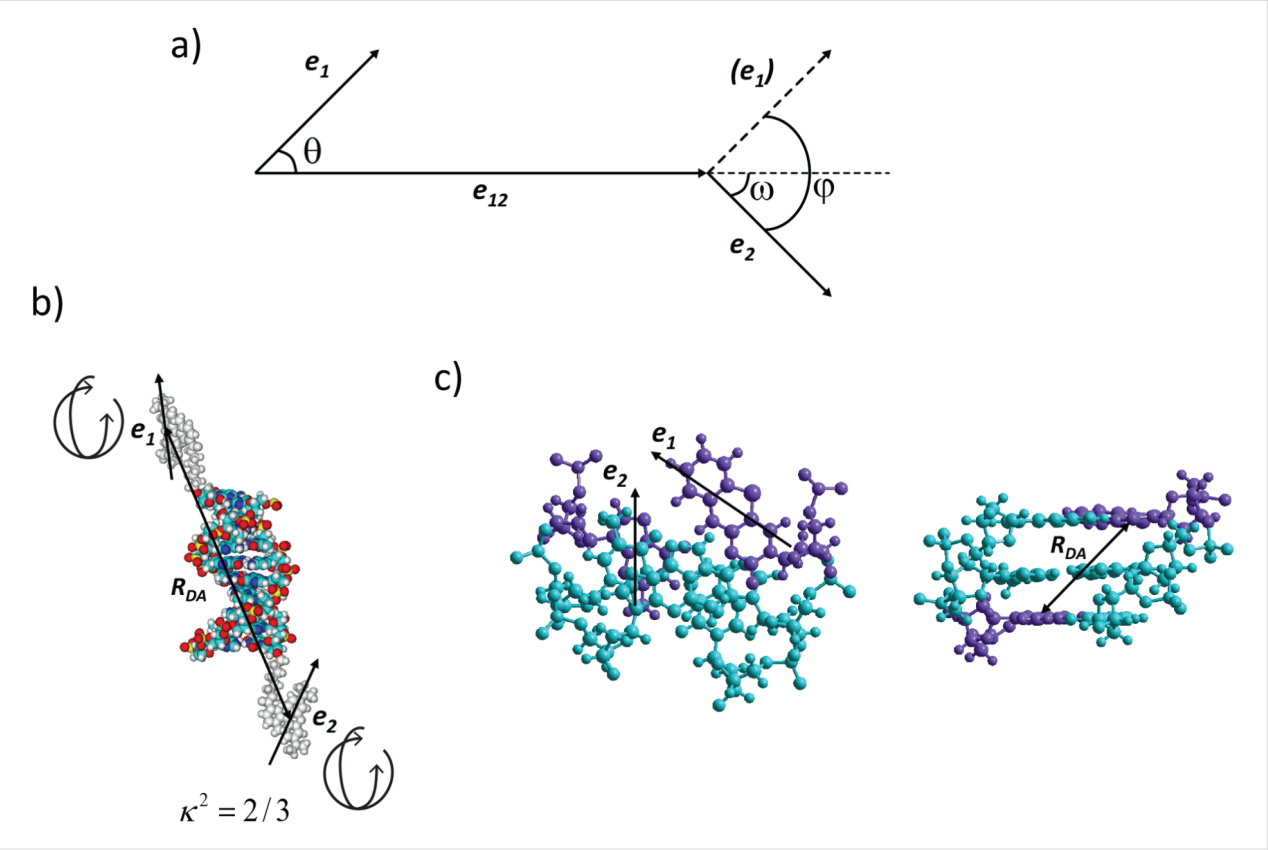


Figure 1: a) Angles and unit vectors used to define the relative orientations of the donor and acceptor transition dipole moments ($\mathbf{e}_1$, $\mathbf{e}_2$) and the separating vector ($\mathbf{e}_{12}$). b) Illustration of the external fluorophores covalently attached to a DNA and their transition dipole moments ($\mathbf{e}_1$, $\mathbf{e}_2$) with free rotation relative to the DNA, i.e., isotropic orientation, $\kappa^2 = 2/3$. R_{DA} is the separation between donor and acceptor. c) DNA top-view (left) and side-view (right) illustrating the typical situation for the virtually static transition dipole moments ($\mathbf{e}_1$, $\mathbf{e}_2$) of fluorescent base analogues in base–base FRET and the distance separating them (R_{DA}).

onal protection groups. A careful consideration of protection groups is paramount as an extensive use adds additional steps as well as complexity to the synthesis. The design and synthesis of fluorescent nucleobase analogues (FBAs) add on additional challenges such as obtaining features that introduce useful photophysical properties, for example, extended conjugation. As an effect of the need for hydrogen bonding properties, size restriction and sterical effects these demands are often conflicting [15,18,19]. However, there is an increasing number of exceptions to this and since the pioneering work of Ward et al. on adenine analogues [20] a whole range of small modifications to nucleobases, such as the 8-vinyldeoxyadenosine [21], has led to the introduction of fluorescence. Considering the differences in the structures of purines and pyrimidines, adenine is unique amongst the natural bases as it offers several sites for modifications: C2, C8, the C6 exocyclic amino functionality and the vastly explored N7 to C7 substitution leading to 7-deaza-adenines. On the contrary, for guanine only the C8 and the C2 exocyclic amine are directly accessible for modifications as well as the previously mentioned 7-deaza substitution. Looking at the monocyclic pyrimidines, only the C5 and C6 positions are available for modifications without directly perturbing the base-pairing properties. The subtle differences between the nucleobases within a class could lead one to believe that the chemistry developed for modifications of adenine would translate easily to guanine. Unfortunately, the variety of functional groups requires different protection group strategies and, moreover, changes the reactivity of the nucleobase. Since the discovery of the gold standard of fluorescent base analogues, 2-AP, a multitude of adenine FBAs has emerged [22]. Notable recent examples of adenine FBAs (see Figure 2 for chemical structures) include C8 to S8 thio-RNA analogue thA [23], the C8-naphtalene substituted adenines ^{cn}A and ^{dn}A [24], as well as our own quadracyclic qAN1 [25]. A handful of fluorescent guanine analogues has been synthesized and characterized and includes the recent turn-on probe BFdG, 3-MI, 2PyG, as well as the emissive RNA analogue thG [23,26-28]. Some notable pyrimidine analogues include our tricyclic analogues tC and tC^O [29-31], pyrrolo-dC [32] and its derivatives [33] as well as thU, thC [23] and ^{DMA}C [34]. Apart from tC, tC^O, qAN1 and thG, FBAs have not yet been utilized in base–base FRET applications. However, the brightest of these FBAs combined with a matching donor or

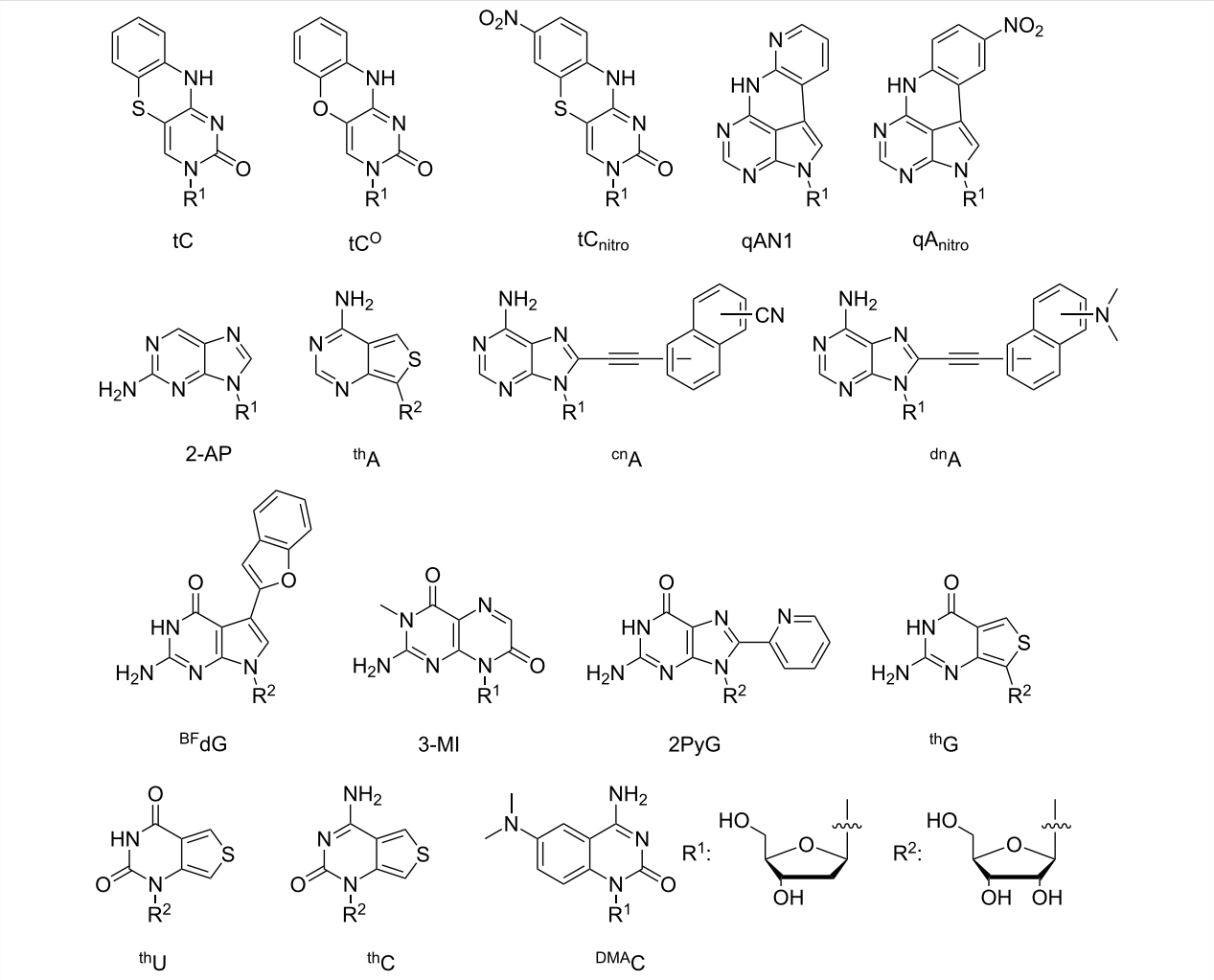
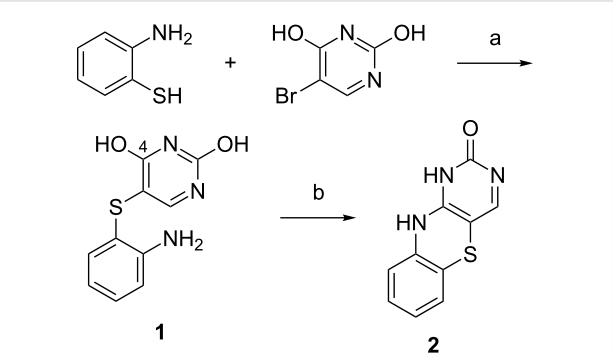


Figure 2: Notable recent examples of fluorescent base analogues. For ^{cn}A and ^{dn}A the attachment point to the substituted naphthalene moiety has been varied.

acceptor molecule could potentially also be used in base–base FRET in the future.

Synthesis of cytosine analogues for base–base FRET in DNA

We have put considerable effort into developing the family of fluorescent base analogues known as the tricyclic cytosines (tC) [14,29-31,35-38]. The aromatic core of tC was first prepared by Roth et al. in 1963 as part of a study to obtain pharmacologically active compounds structurally similar to phenothiazines [39]. Compound **1** (Scheme 1) was readily prepared from condensation of 2,4-dihydroxy-5-bromopyrimidine with 2-aminothiophenol under basic conditions at elevated temperatures and was obtained in 40% isolated yield [40]. Ring-closing of compound **1** to obtain compound **2** was achieved by an acid-catalyzed cyclization which was found to be general for a large set of 4-hydroxy-5-(*o*-aminoarylthio)pyrimidines [39]. The mechanism was thought to proceed via protonation of a pyrimidine ring nitrogen which activates it to nucleophilic attack by an unprotonated anilino nitrogen on the positive C4 of the pyrimidine ring which carries the hydroxy group. The formed com-

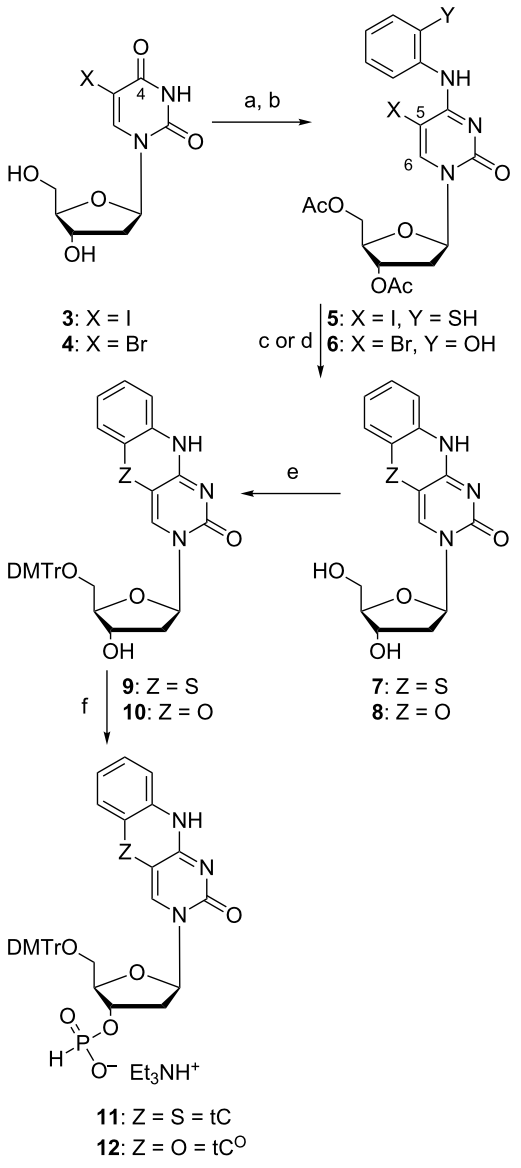


Scheme 1: Synthesis of the tricyclic cytosine aromatic core [39]. (a) Ethylene glycol, K_2CO_3, 120 °C, 1 h, 40%; (b) EtOH, 1 M HCl, reflux, 16 h, 75%.

plex eliminates water and yielded the cyclization product in 75% after isolation.

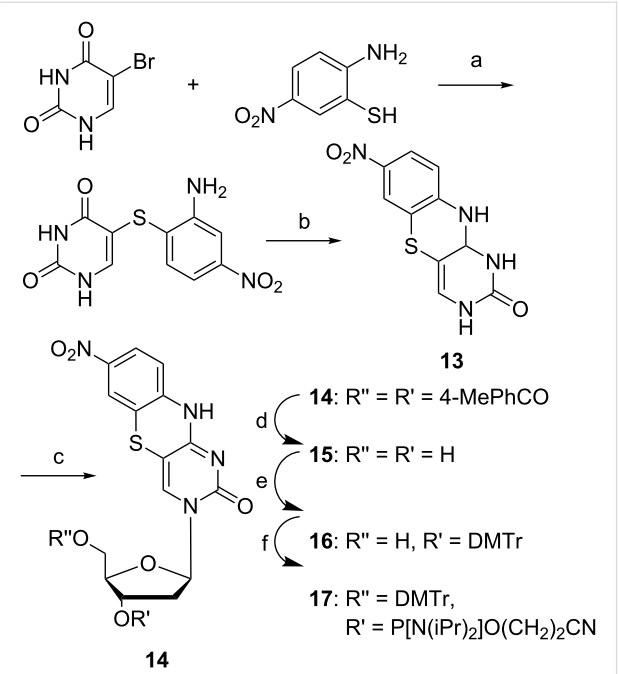
In an effort to increase the thermodynamic stability of oligonucleotide duplexes for antisense purposes, Lin et al. turned to size-expanded nucleobase analogues [41]. To this end they wanted to use the aromatic ring system previously developed by Roth et al. [39] as a nucleobase analogue to furnish greater π–π interactions with the natural bases and possibly also to increased hydrophobic effects. A new strategy for the preparation of tC analogues was used, starting from 5-iodo or 5-bromo-2′-deoxyuridine (Scheme 2) [41]. Compounds **3** and **4** were reacted with acetic anhydride in pyridine to protect the deoxyribose hydroxy groups. The O4 position was then activated by sulfonylation using 2-mesitylenesulfonyl chloride. The subsequent condensation with the appropriate 2-aminothiophenol or 2-aminophenol afforded compounds **5** and **6**, respectively. Refluxing **5** with *t*-BuOK in EtOH generated **7** in 38% isolated yield. When **6** was treated with the same cyclization conditions as **5** only dehalogenation was observed. Compound **8** was obtained by first removing the acetyl protecting groups using ammonia in MeOH, followed by cyclization by refluxing deprotected **8** with KF in EtOH. Presumably, a transient Michael addition of the hydroxy group to the C6 position of compound **6** increases the reactivity of the C5 position towards substitution. Standard dimethoxytritylation of compounds **7** and **8** furnished product **9**, which was used in the next step without isolation and **10** in 50% yield over three steps, respectively. Lastly, phosphitylation yielded the corresponding H-phosphonates **11** in 71% yield over two steps and **12** in 80% yield (Scheme 2).

In 2001, tC was reported as a fluorescent nucleobase analogue [39]. The tricyclic core was synthesized as reported by Roth et al., and subsequently functionalized with a carboxylic acid handle for PNA labeling [39]. In 2003, tC [35] was synthesized bearing a 2′-deoxyribose functionality and thoroughly photophysically characterized (vide infra). tC was later functionalized with a phosphoramidite and incorporated into oligonucleotides [30]. However, the fully detailed synthesis with complete characterization was published in 2007 as a Nature Protocol paper [37]. The aromatic core of tC was prepared according to the procedure of Roth et al. (Scheme 1), followed by a glycosylation using the sodium-salt method as later also performed in the synthesis of tC_{nitro} in 2009 (reaction c, Scheme 3) [14,42]. The synthesis was finished by standard DMTr protection and phosphitylation furnishing tC deoxyribose phosphoramidite in a total of 2.1% yield over 6 steps [43,44]. In 2008, the oxo-analogue tC^O, which Lin et al. initially prepared in 1995 [41], was re-synthesized in order to characterize its photophysical properties, using the same procedure except that *p*-toluoyl protecting groups rather than acetyl were used [31].

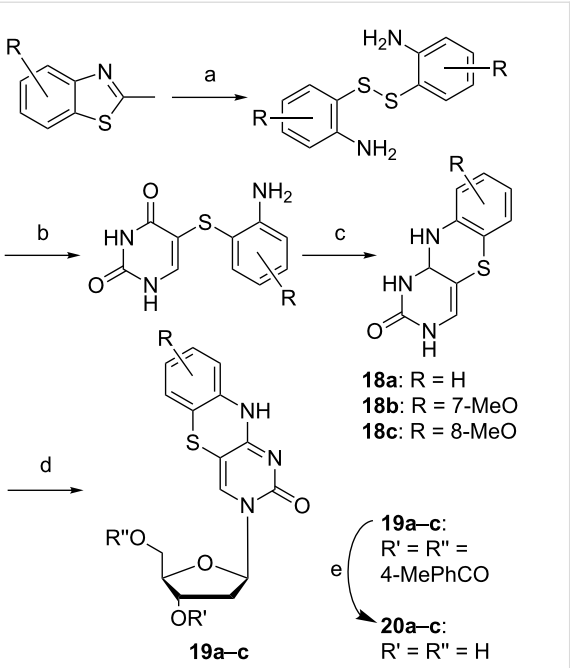


Scheme 2: Synthesis of protected tC and tC^O deoxyribose phosphonates [41]. (a) Ac_2O, pyridine, rt; (b) 2-mesitylenesulfonyl chloride, TEA, then 2-aminothiophenol or 2-aminophenol, DBU, rt, 27% and 54% yield, respectively, over two steps; (c) *t*-BuOK in EtOH, reflux, 38%; (d) NH_3 in MeOH, rt, then 10 equiv of KF, EtOH, reflux; (e) 4,4′-dimethoxytrityl chloride, pyridine, rt, yielding 50% of compound **10** over three steps; (f) 2-chloro-4*H*-1,3,2,-benzodioxaphosphorin-4-one, pyridine, DCM, 0 °C, 71% over two steps and 80%, respectively.

In 2009, we published the first base–base FRET system using tC^O and tC_{nitro} [14]. Nitro groups introduce an increased charge-transfer character to chromophores, which generally results in absorption at lower energies [38,45]. Hence, tC_{nitro} was envisioned to be able to accept the energy transferred from tC or tC^O, and, thus serve as a FRET acceptor. The synthesis of tC_{nitro} followed the procedure of Roth et al. [39] to furnish the aromatic core **13** (Scheme 3). Compound **13** was then glycosy-



Scheme 3: Synthesis of protected tC_{nitro} deoxyribose phosphoramidite [14]. a) aq NaOH, 24 h, reflux; b) EtOH, HCl, 24 h, reflux, 15% over two steps; c) DMF, toluene, 3,5-di-O-*p*-toluoyl-α-D-erythro-pentofuranosyl, NaH, 18 h, rt, 11%; d) MeONa, MeOH, 18 h, rt, 71%; e) pyridine, DMAP, DMTr-Cl, 18 h, rt, 74%; f) DCM, DIPEA, 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite, 1 h, rt, 93%.



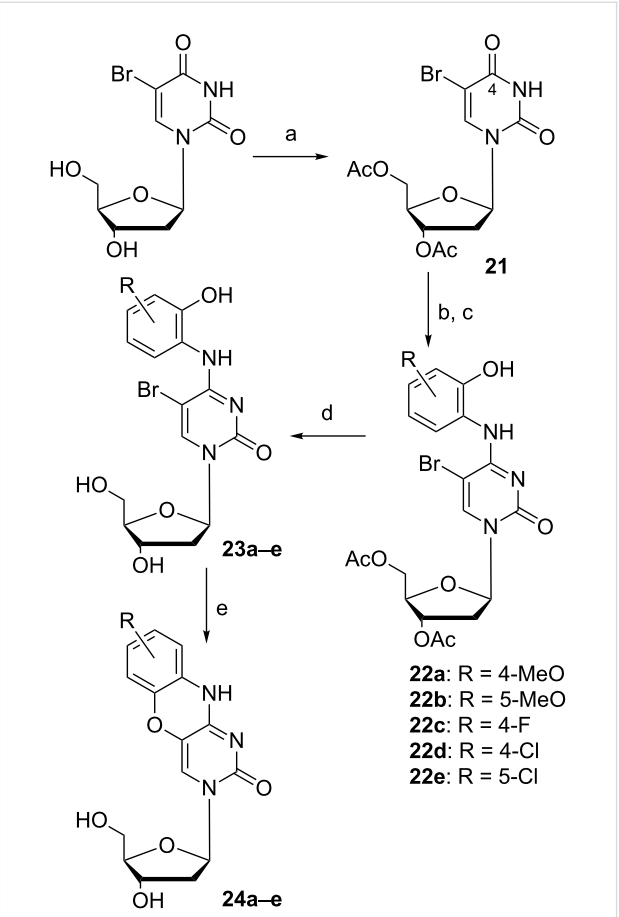
Scheme 4: Improved synthesis of tC and tC derivatives, where R = H, 7-MeO or 8-MeO [47]. a) H_2NNH_2 followed by H_2O_2, 20 h, 100 °C, 60–98%; b) PEt_3, H_2O, diglyme, then Na_2CO_3 and 5-bromouracil, rt to 120 °C, 2 h, 24–86%; c) HCl, butanol, 120 °C, 24–72 h, 27–86%; d) BSA (bis(trimethylsilyl)acetamide), Hoeffer´s α-chloro sugar, $SnCl_4$, 0 °C to rt, 2 h, 12–41%; e) NaOMe, MeOH, 30 min, 69–90%.

lated by making the sodium salt and reacting it with Hoffer´s α-chloro sugar yielding **14** in 11% yield after isolation [46]. The *p*-toluoyl protection groups were cleaved by sodium methoxide in MeOH, which yielded the free nucleoside **15** in 71%. Standard DMTr protection furnished compound **16** which was then activated for oligonucleotide solid-phase synthesis (SPS) by phosphitylation using CEP-Cl. The total yield of tC_{nitro} deoxyribose phosphoramidite was 0.8% over 6 steps where the acid-catalyzed cyclization as well as the glycosylation proved challenging. The latter two steps proceeded with a yield of 15% or less (**17**, Scheme 3).

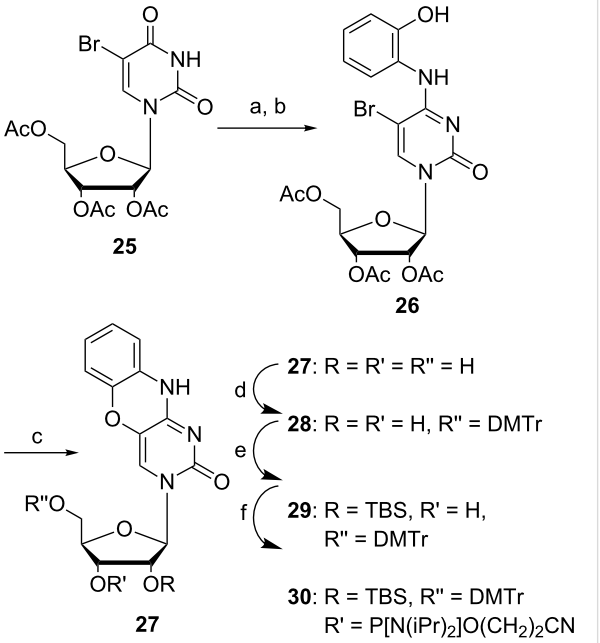
A new synthetic approach to access substituted tricyclic cytosines was envisioned in 2014 by Rodgers et al. (Scheme 4). This protocol increased the yield of the parent compound tC from 10% up to 43% in the glycosylation step of the previously prepared tC nucleobase (Scheme 4) [47]. This was achieved by activation of the aromatic core of compounds **18a**–**c** via trimethylsilylation using BSA (bis(trimethylsilyl)acetamide) [47], instead of the sodium-salt method [42]. The deoxyribosylation was then achieved using the same Hoffer's α-chloro sugar, but in presence of a Lewis acid yielding the protected nucleosides **19a**–**c** [48]. The cleavage of the protection groups was achieved with sodium methoxide to furnish compounds **20a**–**c**.

The improved synthetic route to tC^O derivatives started from the same 3´,5´-di-*O*-acetyl-5-bromo-2´-deoxyuridine (**21**, Scheme 5) as Lin et al. used, but was instead activated for a condensation using Appel chemistry [41,49]. Compound **21** was activated using PPh_3 in CCl_4 which converts the O4 to a 4-Cl and used in situ with various substituted 2-aminophenols in the presence of the strong base DBU which resulted in compounds **22a–e**. A subsequent protection group removal yielded compounds **23a–e** and made the scaffold ready for cyclization. Initially, CsF was used in place of KF, however, the hygroscopic nature of CsF made it impractical to handle. Instead, KF was used in combination with 18-crown-6 in anhydrous diglyme which furnished compounds **24a–e** in modest 3–24% yields after isolation (Scheme 5) [47].

Recently, our group gained interest in RNA chemistry and therefore revisited the synthesis of tC^O containing a ribose unit instead of a deoxyribose [50]. By simply activating the O4 of **25** (Scheme 6) using 2-mesitylenesulfonyl chloride and DIPEA in MeCN, the successful condensation with 2-aminophenol was achieved and afforded compound **26** in 71% yield. The cyclization of **26**, which previously suffered from low yields, was effectively obtained in 86% yield by using an excess of KF in ethanol and microwave heating at 140 °C. Conveniently, at the same time all the three acetyl protecting groups were cleaved



Scheme 5: Improved synthesis of tC^O derivatives [47]. a) Ac_2O, pyridine, 16 h, rt, 85%; b) PPh_3, CCl_4, DCM, 5 h, 44 °C; c) DBU, DCM, R = 4-MeO, 5-MeO, 4-F, 4-Cl, 5-Cl, 15 min, 0 °C; d) NaOMe, MeOH, 3–4 h, rt, 40%, 48%, 61%, 20%, and 41%, respectively, over three steps; e) KF, 18-crown-6, diglyme, 1–2 h, 120 °C, 20% for R = 8-MeO, 11% for R = 7-MeO, 11% for R = 7-F, 24% for R = 8-Cl, and 3% for R = 7-Cl.



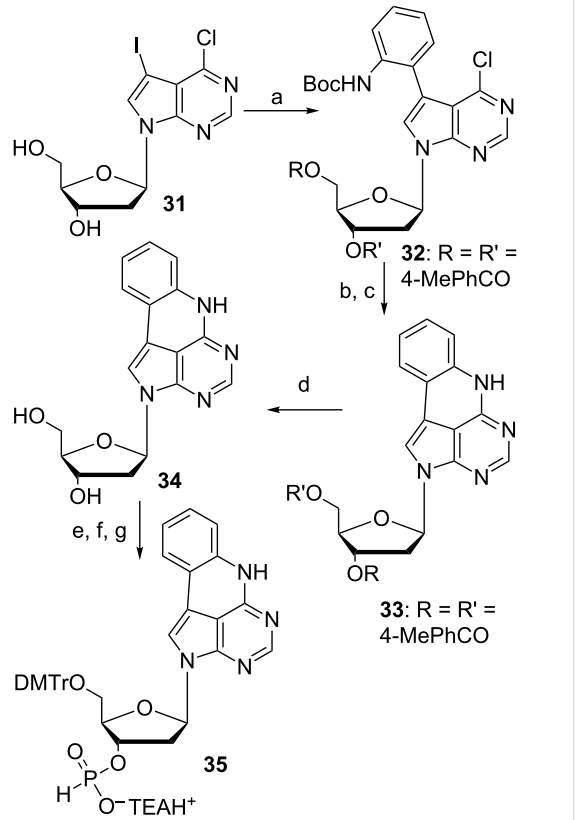
Scheme 6: Synthesis of protected tC^O ribose phosphoramidite [50]. a) $MesSO_2Cl$, DIPEA, MeCN, 4 h, rt; b) 2-aminophenol, 30 min, rt, 71% over two steps; c) KF, EtOH, 2 h, MW 140 °C, 86%; d) DMTr-Cl, pyridine, 1.5 h, rt, 72%; e) $AgNO_3$, TBDMS-Cl, pyridine, THF, 4 h, rt, 76%; f) 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite, DIPEA, THF, 1 h, rt, 93%.

and the free nucleobase was isolated via precipitation. A 5′-DMTr protection followed by 2′-TBDMS protection and phosphitylation using CEP-Cl generated the fully protected monomer ready for solid-phase synthesis [50]. The complete synthesis of the RNA building block of tC^O was in this way achieved over five steps with a total yield of 28%, improved from the four step DNA building block synthesis of tC^O by Lin et al. of 22% [41].

Synthesis of adenine analogues for base–base FRET

Buhr et al. were interested in developing modified adenosines that could thermodynamically stabilize double-stranded nucleic acids [51]. In 1999, a short synthesis article regarding quadracyclic adenine, qA, was published, however, it lacked a full experimental procedure (Scheme 7). The synthesis started from 6-chloro-7-iodo-7-deazapurine functionalized at the N-9 position with Hoffer's α-chloro sugar (**31**, Scheme 7). This material was functionalized further using a Stille coupling to afford a mono-Boc-protected *o*-iodoaniline **32** in 68% yield after isolation [52]. The cyclization was performed via nucleophilic aromatic substitution with DBU and DABCO. Presumably DABCO activates the chlorine and modifies it into a better leaving group allowing the sterically hindered base DBU to abstract a proton from the protected aniline which allows the cyclization. Standard Boc deprotection using TFA gave compound **33** in 96% over two steps. This was followed by *p*-toluoyl deprotection using sodium methoxide in methanol to afford **34** in 64% yield after isolation. Then, the material was protected with DMTr-Cl yielding the protected nucleoside in 65% yield. Subsequent phosphitylation followed by salt-formation finally furnished compound **35** in 52% over two steps.

Since the quadracyclic adenine presented an overall structural similarity with adenine and keeping a very rigid heterocyclic system suggesting few options for the molecule to decay from excited states via non-radiative processes, in 2012 we decided to re-synthesize the quadracyclic adenine according to the procedure of Buhr et al. (Scheme 8) [51,53]. However, in our hands the vital cyclization reaction starting from compound **36** (Scheme 8) never provided more than a 46% yield of **37** after

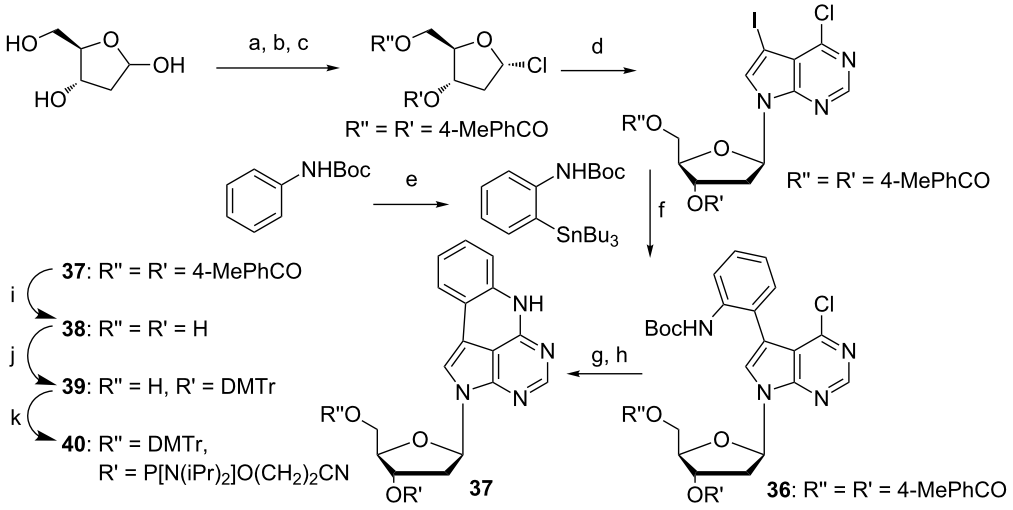


Scheme 7: Synthesis of protected deoxyribose qA [51]. a) *N*-(*tert*-Butoxycarbonyl)-2-(trimethylstannyl)aniline, $(Ph_3P)_2PdCl_2$, DMF, 24 h, 60 °C, 68%; b) DABCO, DBU, DMF, 21 h, 75 °C; c) 25% TFA in CH_2Cl_2, 3 h, rt, 96% over two steps; d) NaOMe, MeOH, rt, 64%; e) DMTr-Cl, pyridine, 65%; f) 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one, pyridine, CH_2Cl_2, 30 min, 0 °C to rt; g) aq triethylammonium bicarbonate, 52% over two steps.

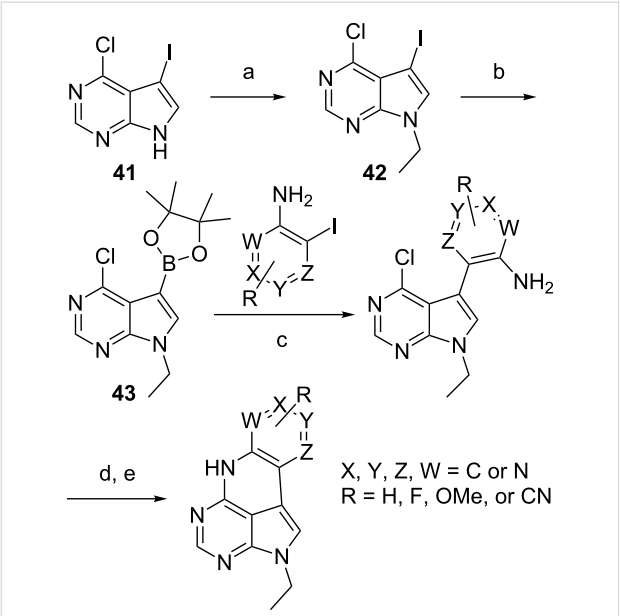
TFA deprotection of the Boc group compared to the previously reported 96% [51].

The base-pairing properties of qA with T and selectivity were found to be excellent. Moreover, the melting temperature of the oligonucleotides remained close to those of unmodified sequences indicating that qA is an excellent adenine analogue [53]. Unfortunately, the photophysical properties of qA were not satisfactory for an internal FRET fluorophore and, thus, we moved on by modifying the quadracyclic aromatic core but leaving the advantageous base-pairing properties. To this end, we needed to develop a more straightforward and versatile synthetic route. The Stille coupling was changed to a Suzuki–Miyaura coupling and the cyclization was performed directly starting from the free aniline nitrogen, as we found that Boc protection was required only for cyclization when using DBU and DABCO. To faster screen a larger set of new compounds for fluorescent properties we envisioned that it was unnecessary to carry the entire sugar moiety through the synthesis. Thus, by alkylation of 6-chloro-7-iodo-7-deazapurine (**41**, Scheme 9) followed by a Miyaura-style borylation of compound **42**, inspired by Thompson et al. we achieved compound **43** in a yield of 77% over two steps [54]. This material was functionalized in two different studies: first by using pyridine-type anilines and later with R-group modifications to the top ring (Scheme 9) [55,56].

Among the quadracyclic adenine analogues in those two studies we found qAN1 to be a promising candidate as a FRET donor due to its high quantum yield of 0.18 (vide infra) [55]. To



Scheme 8: Synthesis of protected deoxyribose qA for DNA SPS [53]. a) AcCl, MeOH, rt, 40 min; b) *p*-toluoyl chloride, pyridine, overnight, 0 °C to rt; c) AcCl, AcOH, H_2O, 0 °C, 36% over three steps; d) NaH, MeCN, 30 min, rt, then 2 h at 60 °C, 73%; e) *t*-BuLi, $SnBuCl_3$, THF, 2h, −78 °C, 65%; f) $Pd(PPh_3)_4$, CuI, CsF, DMF, 1 h, 100 °C, 55%; g) DABCO, DBU, DMF, 16 h, 75 °C; h) 25% TFA in CH_2Cl_2, 1.5 h, 0 °C to rt, 46% over two steps; i) NaOMe, MeOH, overnight, rt, 61%; j) DMTr-Cl, pyridine, 1 h, rt, 68%; k) 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite, DIPEA, CH_2Cl_2, 1.5 h, rt, 79%.



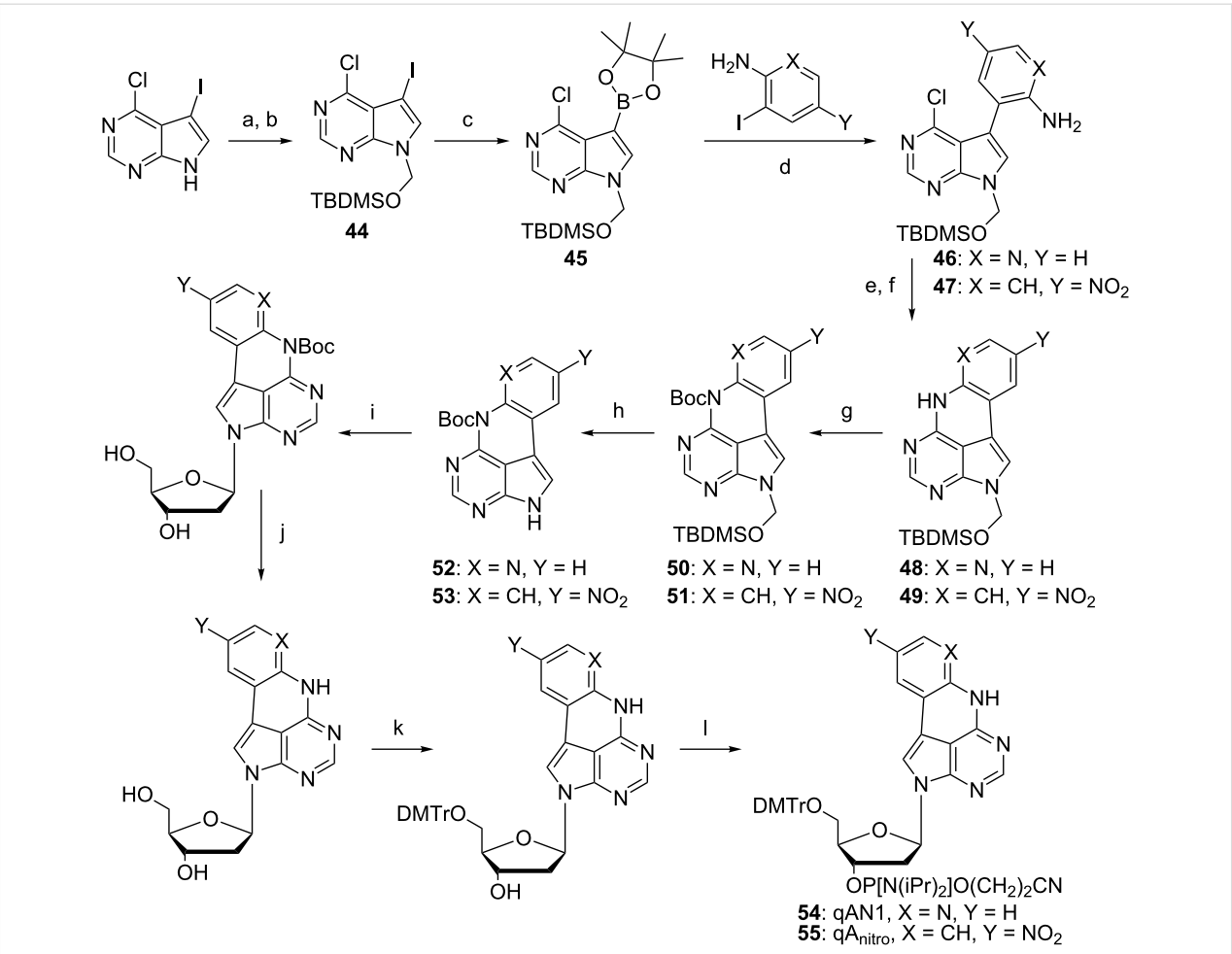
Scheme 9: Synthesis of qA derivatives. a) EtI, Cs_2CO_3, DMF, 4 h, rt, 90%; b) HBPin, $Pd(PPh_3)_4$, Et_3N, 1,4-dioxane, 24 h, 80 °C, 86%; c) $PdCl_2(PPh_3)_2$, K_3PO_4, MeCN, H_2O, 80 °C, 2 h, 56–86%; d) TMS-Cl, THF, 30 min, rt; e) LiHMDS, THF, 100 °C, 3 h, 33–71%.

develop an adenine acceptor for qAN1, a similar approach as for the tricyclic cytosines was performed, i.e., the introduction of a nitro functionality in the outer ring of qA. In an extensive investigation qA_{nitro} was synthesized and characterized and we found it, indeed, to be a suitable FRET acceptor for qAN1 (vide infra). The full synthesis scheme and characterization of this adenine–adenine analogue FRET pair was recently published by our group [25]. The synthetic approach was to first construct a common intermediate that could be used for various Suzuki-coupling partners similar to what we previously reported [55], by first protecting 6-chloro-7-iodo-7-deazapurine with *tert*-butyldimethylsilyloxymethyl (TBDMSOM) in 86% yield over two steps (**44**, Scheme 10). A Miyaura-type borylation afforded the common intermediate **45** in 91% yield and Suzuki coupling was then achieved efficiently for both 2-amino-3-iodopyridine as well as 2-iodo-4-nitroaniline in (**46**) 95% and (**47**) 86% yield, respectively. The activation of the exocyclic amine was achieved by using AcCl which provided a more robust cyclization using LiHMDS than if activating the amine using TMS-Cl. This furnished compounds **48** and **49** in 89% and 87% yield, respectively, over two steps. The subsequent Boc protection gave compound **50** in 89% and compound **51** in 83% yield. The quantitative TBDMSOM deprotection set the stage for a glycosylation using Hoffer's α-chloro sugar and compounds **52** and **53** provided the desired β-anomer after purification in 69% and 55% yield, respectively. Global deprotection using sodium methoxide followed by standard DMTr-protection and phosphitylation provides the activated monomers for solid-phase synthesis [25]. The overall yield of qAN1 and qA_{nitro} phosphoramidite was 19% and 14%, respectively, which is significantly higher than our previous synthesis of qA (6% overall yield) starting from 6-chloro-7-iodo-7-deazapurine (Scheme 10).

Photophysical properties of tricyclic cytosine analogues in nucleic acids

The tricyclic cytosine base analogues 1,3-diaza-2-oxophenothiazine (tC), and its oxo homologue, 1,3-diaza-2-oxophenoxazine (tC^O) (Scheme 1 and Scheme 2) are both excellent fluorescent base analogues as well as donors for base–base FRET in nucleic acids [14,29-31,36]. Extensive evidences that both these base analogues mimic the behavior of natural cytosine have been found using UV–vis [30] and NMR spectroscopy [36], e.g., exchanging cytosine for one of them results in a virtually unperturbed B-form DNA helix. Importantly and as the first fluorescent base analogue with such properties, tC shows high and stable quantum yields (around 20%) both in monomeric form, in single- as well as in double-stranded DNA [29,30]. The quantum yield of tC^O in different environments is even higher than those of tC [31]. While slightly dependent on the neighboring base environment they are still very stable compared to other common fluorescent base analogues [15,16]. The absorption maxima of tC and tC^O in DNA are found at approximately 395 and 365 nm (Figure 3) [30,31], respectively, and, thus, are well separated from the absorption of the natural nucleobases. The emission of tC and tC^O in duplex DNA display large Stokes shifts, cover a broad wavelength region and the maxima are found at 505 and 450 nm (Figure 3), respectively [30,31]. Their spectral envelopes, which are an important factor for the overlap integral in FRET, are robust to changes in the local environment.

The fluorescence decays of tC are all monoexponential in single- as well as in double-stranded DNA resulting in a single lifetime of 5–7 ns depending on the sequence surroundings [30]. For tC^O single-stranded surroundings generally result in biexponential decays, whereas duplex surroundings, as in the case for tC, result in single fluorescence lifetimes (3–5 ns) [31]. High and stable quantum yields and single lifetimes in duplexes along with firm stacking are properties that make tC and tC^O excellent FRET donors. In order to make evaluation of FRET data more exact, through a high precision in the orientation factor (κ^2), we have also determined the direction of the transition dipole moments of tC and tC^O (35° and 33° clockwise from the molecular long-axis as represented in Scheme 1 and Scheme 2, respectively) [31,35]. To complete the first base–base FRET pair there was a need for a FRET acceptor that could match tC and/or tC^O. To this end we developed the nitro-version of tC, 7-nitro-1,3-diaza-2-oxophenothiazine (tC_{nitro}) (Scheme 3) [14,38,45]. From UV–vis spectroscopy we showed



Scheme 10: Synthesis of quadracyclic adenine base–base FRET pair. a) HCHO, NaOH, MeCN, H_2O, 50 °C, 1 h; b) TBDMS-OTf, pyridine, 1 h, 0 °C to rt, 86% over two steps; c) HBPin, $Pd(PPh_3)_4$, Et_3N, 1,4-dioxane, 24 h, 80 °C, 91%; d) $PdCl_2(PPh_3)_2$, K_2CO_3, MeCN, H_2O, 80 °C, 2 h, 86–95%; e) AcCl, pyridine, CH_2Cl_2, 3 h, rt; f) LiHMDS, THF, 100 °C, 2–6 h, 87–89%; g) Boc_2O, DMAP, THF, 10 h, rt, 83–89%; h) ethane-1,2-diamine, TBAF, THF, 2 h, 0 °C to rt, 97–100%; i) NaH, Hoffer´s α-chloro sugar, MeCN, 2 h, 0 °C to rt, 55–69%; j) NaOMe, MeCN or MeOH, 1 h, 50 °C, 81–99%; k) DMTr-Cl, pyridine, 1.5 h, rt, 55–75%; l) 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite, DIPEA, CH_2Cl_2, 2 h, rt, 87–90%.

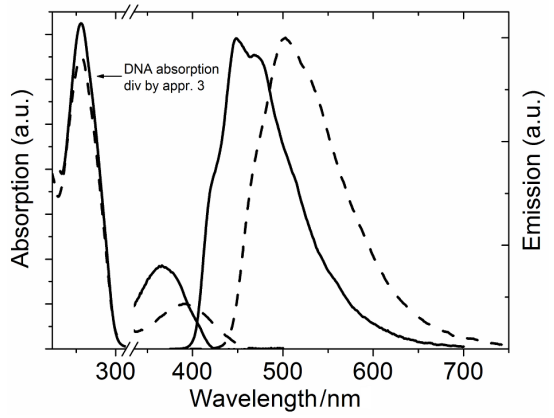


Figure 3: Absorption and emission of tC (dashed line) and tC^O (solid line) in dsDNA. The absorption below 300 nm is divided by three to emphasize the absorption spectral features of the lowest energy absorption bands of tC and tC^O.

that it, as do tC and tC^O, forms stable, B-form duplexes and stacks firmly inside the DNA. It is a non-emissive chromophore that has an absorption maximum around 440 nm which overlaps well with the emission of both tC and tC^O, thus, making it a potential FRET acceptor for both of them [14]. The best spectral overlap is found between the emission of tC^O and the absorption of tC_{nitro} giving a Förster radius (R_0) of 27 Å using an isotropic orientation factor, $\kappa^2 = 2/3$ [14]. Finally, for a high precision in orientation factor, i.e., to enable detailed structure investigations, we determined the direction of the lowest energy transition dipole moment of tC_{nitro} to be 25° in the opposite direction compared to tC and tC^O (i.e., pointing towards the nitro group) [14].

As was mentioned in the synthesis part above, recently we also have developed tC^O as an internal fluorophore for RNA systems [50]. The incorporation into RNA oligonucleotides and hybridi-

zation with a complementary strand results in normal A-form RNA duplexes. Moreover, the useful absorptive and emissive spectral properties of tC^O in DNA are retained in RNA. However, fluorescence decay data for tC^O in RNA suggests a less rigidly stacked conformation in RNA and two lifetimes are normally needed to achieve a good fit of the decays. With virtually stable quantum yields of 20–25% inside duplex RNA, tC^O is the brightest internal RNA fluorophore reported to date and, thus, a promising fluorescence reporter and/or FRET donor also in RNA systems [50].

Photophysical properties of quadracyclic adenine analogues in nucleic acids

Extending the repertoire of base–base FRET pairs to other nucleobases would provide researchers the opportunity of replacing any sequence position in a nucleic acid with a base analogue FRET donor or acceptor. This motivated us to venture into the development of adenine analogues. Quadracyclic adenine (qA) [51], the emission of which was first reported by our group, was our initial adenine analogue candidate [53]. It stabilizes the native B-form DNA and is selective for base pairing with thymine. The emissive properties are decent both for the monomer (Φ_f = 6.8%) and inside DNA even though the quantum yield is quenched in the latter case. However, the average brightness in duplex DNA is still higher than that of 2-aminopurine and together with the excellent base-paring properties it is still a highly useful, environment-sensitive fluorescent-base analogue [53].

Despite its excellent base-analogue properties, the low quantum yield of qA inside DNA disqualifies it for use as a base–base FRET donor. In order to maintain the base-analogue properties and achieve improved photophysical properties, we used quantum chemistry-supported design and developed a series of four, second generation, quadracyclic adenine analogues, qAN1–qAN4 (Scheme 9 and Scheme 10) [55,56]. As monomers, these compounds show significantly improved fluorescence properties. Importantly, one of the derivatives, qAN1, showed a high quantum yield in water (18%) that was not excessively influenced by varying the solvent, indicating that qAN1 is not highly sensitive to the direct surroundings [55]. Once incorporated into DNA strands, qAN1 specifically base-pairs to the complementary base, thymine, and allows formation of stable B-form DNA [25]. Moreover, the quantum yields inside DNA are significantly increased compared to those of qA. However, the quantum yields of qAN1 are slightly sensitive to the directly flanking bases with an average quantum yield of 6% in dsDNA [25]. The wavelength of the emission maximum found around 415 nm (Figure 4) in dsDNA is insensitive to the neighboring bases and the spectrum is more structured compared to the spectrum of monomeric qAN1, implying a firm stacking inside DNA [25]. The fluorescence lifetimes of qAN1 inside dsDNA show biexponential decays (average amplitude-weighted lifetimes ranging from 0.8 to 3.3 ns) for a majority of investigated sequences as compared to triexponential decays in ssDNA [25]. Overall, with a brightness ($\Phi_f \cdot \varepsilon$ = 510 $M^{-1}cm^{-1}$) inside DNA which is 29-times higher than for qA, specificity towards T and a firm stacking inside B-form DNA, qAN1 represents an excellent base–base FRET-donor candidate. To complete the base–base FRET pair the acceptor qA_{nitro} (Scheme 10) was designed and synthesized [25]. Spectroscopy-based investigations of the base analogue properties of qA_{nitro} inside DNA suggest that this derivative of qA is an excellent A-analogue just like qAN1. The lowest absorption maximum for qA_{nitro} in DNA is located at 435 nm (Figure 4) with a molar absorptivity of 5400 M^{-1} cm^{-1}. As in the case of tC^O and tC_{nitro} there is an excellent spectral overlap between the emission of qAN1 and the absorption of qA_{nitro} (Figure 4) resulting in a Förster radius (using an orientation factor κ^2 = 2/3) of 22 Å. This suggested that qAN1 and qA_{nitro} would constitute a good base–base FRET pair [25].

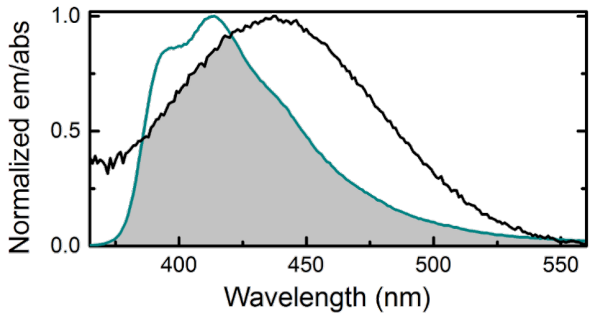


Figure 4: Spectral overlap between the emission of qAN1 (cyan) and the absorption of qA_{nitro} (black) in dsDNA. The shaded region constitutes the overlap integral (*J* integral).

Fluorescent base analogue FRET pairs inside DNA

When using FRET theory on fluorophores/chromophores that replace nucleobases of a normal but static DNA, estimated energy-transfer efficiencies can be simulated using the structure parameters of the B-form duplex together with photophysical parameters of the fluorophores/chromophores. In this way we used the photophysical parameters we already had obtained for our two FRET pairs, tC^O–tC_{nitro} and qAN1–qA_{nitro}, to design the best combination of donor and complementary acceptor-containing DNA oligonucleotides [14,25]. We found that eight DNA strands were sufficient: three donor (tC^O/qAN1) strands, four acceptor (tC_{nitro}/qA_{nitro}) strands and one unmodified strand serving as the complementary strand in donor-only reference samples. Combining these strands in an optimal way we covered distances of 2–13 bases separating the donor and

the acceptor. For each base separation the FRET efficiency was investigated both by steady-state and time-resolved emission measurements. The results of those show an excellent resemblance with our predicted values for a nucleobase FRET pair situated inside a static DNA: an overall sharp (R^6) decrease in the FRET efficiency with increasing numbers of bases separating the donor and the acceptor with an overlaid sinusoidal curvature as a consequence of the effect of the helical nature of B-form DNA on the orientation factor, κ^2 (Figure 5) [14,25]. Both, the measured sets of FRET efficiencies, the one for tC^O–tC_{nitro} as well as the one for qAN1–qA_{nitro}, were fitted to an averaged, static B-form DNA model using an in-house built MATLAB script. The best fits agree excellently with the measured data and suggest that our two FRET pairs are indeed rigidly stacked inside DNA and serve as excellent distance and orientation dependent FRET probes (Figure 5) [14,25]. In the fit we used the associated phase angle (angle between the transition dipole moments of the donor and the acceptor) and the spectral overlap (J_{DA}) as the fitting parameters. The phase angles for the tC^O–tC_{nitro} and the qAN1–qA_{nitro} FRET pairs were 67° and 33°, respectively, that are in good agreement with the experimentally determined one for the cytosine analogue FRET pair (58°) and the TDDFT-estimated one for the adenine analogue FRET pair (41°) [14,25]. Also the spectral overlap integrals show high similarity to the values resulting from the best fit. Taken together these two FRET pairs comprise excellent tools to study detailed structure, dynamics and conformational changes of DNA. An additional advantage with our cytosine and adenine analogue FRET pairs is that they, as a result of their spectral features, can be combined with each other, i.e., qA_{nitro} can replace tC_{nitro} as an acceptor of tC^O, and, similarly, tC_{nitro} can replace qA_{nitro} as an acceptor of qAN1 [25]. A useful advantage of this is that we can now perform base–base FRET between any sequence positions inside duplex DNA.

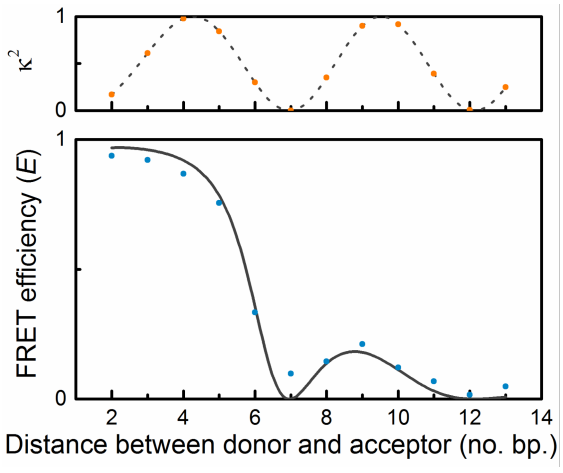


Figure 5: Example of typical FRET efficiency as a function of number of base pairs separating the donor and acceptor (data, blue dots, is an average of steady-state and time-resolved measurements of the FRET pair qAN1 and qA_{nitro}). The line shows a curve fitted to the data based on FRET theory. The top graph shows how the orientation factor, κ^2, varies with number of base pairs separating the donor and acceptor.

Recently Sugiyama and co-workers reported a nucleobase-analogue FRET pair that consists of the 2-aminothieno[3,4-*d*]pyrimidine G-mimic deoxyribonucleoside (thdG) (see Figure 2) [23], developed by the Tor lab, as an energy donor and 1,3-diaza-2-oxophenothiazine (tC), developed by our lab, as an energy acceptor [57]. This G–C analogue FRET pair also displays the general characteristics of an energy efficiency curve of base–base FRET and is able to emit cyan-green light from its acceptor molecule tC. The authors used this FRET pair to study a change from B-to-Z-form DNA using the color and intensity change of the combined donor and acceptor emission [57].

FRETmatrix

To enable detailed studies using our FRET pairs we have developed the freely available software FRETmatrix [58]. It consists of two parts, one that predicts FRET efficiencies from structural input and one that can calculate the most probable structure using measured FRET efficiencies as input.

The first part is useful in the design of a study, as it can predict the change in FRET efficiency between two base analogues upon a structural change of the DNA (for example caused by protein binding). This allows the user to make informed choices of where in a DNA duplex to incorporate the modified bases to get useful FRET-change responses. The second, more powerful part provides structural information based on the FRET efficiencies measured between base analogues positioned on opposite sides of a constraint site. The constraint site can, for example, be a protein-induced kink in the DNA. The software needs the DNA sequence together with photophysical data of the FRET pair and measured FRET efficiencies as input. Then, assuming the rest of the DNA is unchanged, the geometrical parameters of the constraint site can be obtained. This is possible since the base analogues are rigidly positioned inside the DNA (Figure 1c) and the FRET efficiencies depend on the relative distance and angle between them (see Equations 1–3). FRETmatrix, in this way provides a convenient possibility to study structural changes of nucleic acids in solution using only emission measurements [58]. For example, in a small demonstration study we have shown that the method can be used to resolve the structure of a 3A (3 adenine) bulge [58]. The same bulge has been studied by other groups as well, using different techniques and with similar results [59-61]. An elegant and groundbreaking way to study and use detailed FRET has also been reported by Seidel et al. By assuring that the external dyes in use

are truly free to rotate (see Figure 1b; isotropic orientation) an orientation factor, κ^2, of 2/3 can be assumed. In combination with advanced computer modeling, high accuracy structural parameters can thus be resolved from FRET measurements [62,63]. In conclusion, the FRETmatrix (base–base FRET) and Seidel's methodology are in a way two extremes, firmly vs randomly oriented probes, both giving high control of the κ^2 value which in turn facilitates high detail structure information determination.

Studying nucleic acid conformation and conformational changes using base–base FRET

Many biologically important processes such as binding of transcription factors to DNA, polymerase–DNA interactions during replication, gene regulatory systems and structure variation due to changes in conditions (e.g., B-to-Z-form DNA), generally involve conformational changes where base–base FRET can be used with an advantage. The possibility to monitor both, distance and orientation of these conformational changes and inherent dynamics of the systems in real time increases the level of detail accessible in the FRET investigation. Over the less than ten years they have been available, nucleobase analogue FRET pairs have been able to monitor several important processes including transcription and DNA repair. Here we give a short summary of a number of those applications.

Higher detail structure information investigations

DNA exists in a variety of conformations depending on conditions. Z-DNA, a GC-repeat rich, thermodynamically less preferred, left-handed helical conformation that is favored by cytosine methylation is known to form in vivo under negative supercoiling or high salt concentrations [64-68]. Circular dichroism is traditionally the predominant method to investigate Z-DNA and to monitor conformational changes from B-to-Z-form DNA [68-72]. However, the development of nucleobase analogue-based FRET provided an opportunity to sense the significant orientational and distance changes for the B-to-Z-transition in real time using significantly smaller sample amounts. Therefore, we set out to use the tC^O–tC_{nitro} FRET pair to develop new methodologies to investigate Z-form DNA [73]. Two different DNA constructs were selected: one of them containing a $(GC)_7$ and the other a $(GC)_5$. The former is a hairpin which is designed to be able to transform completely into Z-form DNA at high salt concentrations and the latter is able to form a B–Z DNA junction under similar, high salt conditions. The tC^O–tC_{nitro} FRET pair was incorporated at three different base separations (4, 6, and 8 bases between donor and acceptor, respectively). The results show significant changes in the FRET efficiencies upon B-to-Z-DNA transition (e.g., from 35 to 8%) that can, not only, be used to monitor the presence of Z-form DNA but also to determine the rate constants for these transitions [73]. We showed in this investigation that the FRET-based method to study Z-form DNA reduces the amount of sample needed by almost three orders of magnitude compared to the most commonly used CD methodology [73].

Recently we used our adenine analogue FRET pair, qAN1–qA_{nitro}, to study the conformational change of B-form DNA upon interaction with the established minor groove binder netropsin [25]. Netropsin is an archetypal minor groove ligand that binds short (4–5 bp) AT-rich sequences [74-76]. In our investigation we first measured the FRET efficiencies, using both steady-state and time-resolved emission, between qAN1 and qA_{nitro} separated by 2–13 bp in a B-form DNA. Thereafter, we added netropsin until site saturation and again measured the FRET efficiencies (Figure 6), now for base separations of 4–11 bp [25].

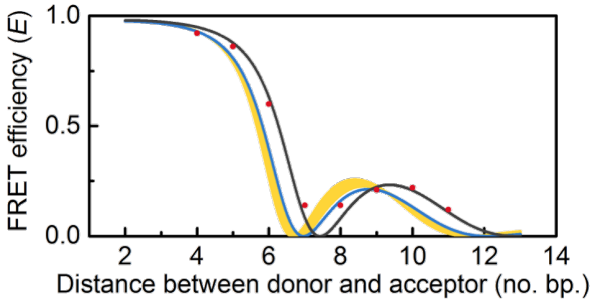


Figure 6: FRET efficiency as a function of number of base pairs separating the donor (qAN1) and acceptor (qA_{nitro}). Red dots mark the measured FRET efficiency with netropsin bound. The black line shows the best fit to the data based on FRET theory. The blue line shows the curve for B DNA. The yellow area depicts the range possible if each netropsin molecule overwinds the DNA as stated in previous literature.

Upon netropsin binding the measured base–base FRET efficiencies change significantly in their orientational component (extreme values are shifted to larger base separations) and also slightly in their distance component (shows up as a higher amplitude for the maximum around 9–10 bp). Fitting these FRET data (Figure 6) points to the best possible DNA helical twist and rise values results in a decrease in twist and rise by 2° and 0.25 Å, respectively [25]. This is in contrast to previously reported values showing slight increases in helical twist and rise measured, for example, by sedimentation [77], gel electrophoresis [78,79], X-ray crystallography [74] and magnetic tweezers [75]. In our investigation we modeled the general appearance of a FRET curve resulting from such, small, helical twist and rise increases (yellow area in Figure 6) and in this way were able to establish that our data unambiguously show decreases in twist and rise [25]. One important difference with our system compared to most previous studies is the fact that we use short DNAs that easily can relax the strain induced by the netropsin

binding. Our netropsin study shows the general strength and potential of base–base FRET to investigate even very small changes in distance and orientation and the finding warrants further studies of this structural change using a larger set of DNA sequences containing netropsin binding sites.

Qualitative base–base FRET to investigate vital cellular processes

As clearly shown above, base–base FRET is a powerful method to obtain structure information with high structure detail. However, one of its obvious and even more straightforward applications is merely for monitoring whether a certain process involves a conformational change or not. A few examples of such applications are described below.

In a collaborative investigation with Falkenberg and Gustafsson, we investigated the role of the transcription factor A (TFAM) in the mitochondrial transcription machinery [80,81]. The investigation, also involving an extensive use of gel electrophoresis studies, shows that TFAM, in contrast to previous reports, indeed is a core component of the machinery. In the study our FRET pair tC^O–tC_{nitro} was site-specifically incorporated in various positions close to the HSP1 transcription initiation site. The results suggest that when TFAM binds to the DNA, it causes significant structural changes [80]. These changes are clearly visible in the tC^O–tC_{nitro} FRET data that also indicate that the conformational changes could be consistent with DNA breathing. Moreover, the data demonstrated that the structural changes upon binding of TFAM near the transcription initiation site are the result of sequence-independent binding to DNA. The investigation establishes the potential of using base–base FRET for studying nucleic acid conformations in vital cellular processes without perturbing the system under study.

In another report using the tC^O–tC_{nitro} FRET pair as a probe of protein interaction, Ansari et al. investigated the DNA damage repair system [82]. Here the FRET pair is used to better understand the conformational dynamics along the DNA-lesion recognition trajectory. The tC^O–tC_{nitro} FRET pair was incorporated on both sides of mismatched regions in a DNA to report on conformational changes upon DNA repair protein Rad4 interaction. The FRET data obtained support a model in which Rad4 binds to the mismatched part causing a “twist-open” mechanism and demonstrates the potential of base–base FRET in short time-scale kinetics investigations [82].

Conclusion

Base–base FRET has a great potential as a detailed structure and dynamics tool in biomolecular sciences. It serves as an interesting complement to FRET pairs based on external fluorophores enabling higher structure resolution and monitoring of a different distance range with high accuracy. With the recent advent of new base–base FRET pairs, the coming years offer great prospects for increased use of such methodology. The combination of base–base FRET and single-molecule-based FRET on nucleic acids with external probes as developed by Seidel et al. [4] comprise a highly interesting opportunity to investigate structure and dynamics of nucleic acid containing systems. Recent progress in the field of fluorescent base analogues also starts to close the gap in brightness to external fluorophores like Cy-, Alexa- and ATTO-dyes and the development of a base analogue with properties that are satisfactory for single molecule use would open up completely new possibilities to study the detailed structure, dynamics and conformational changes of one of the key players in life: nucleic acids.

ORCID® iDs

Sangamesh Sarangamath - https://orcid.org/0000-0001-8430-6426
Moa S. Wranne - https://orcid.org/0000-0003-3315-8288
Morten Grøtli - https://orcid.org/0000-0003-3621-4222
L. Marcus Wilhelmsson - https://orcid.org/0000-0002-2193-6639

References

1. Foster, M. P.; McElroy, C. A.; Amero, C. D. *Biochemistry* **2007,** *46,* 331–340. doi:10.1021/bi0621314
2. Holbrook, S. R. *Annu. Rev. Biophys.* **2008,** *37,* 445–464. doi:10.1146/annurev.biophys.36.040306.132755
3. Preus, S.; Wilhelmsson, L. M. *ChemBioChem* **2012,** *13,* 1990–2001. doi:10.1002/cbic.201200400
4. Kalinin, S.; Peulen, T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. *Nat. Methods* **2012,** *9,* 1218–1225. doi:10.1038/nmeth.2222
5. König, I.; Zarrine-Afsar, A.; Aznauryan, M.; Soranno, A.; Wunderlich, B.; Dingfelder, F.; Stüber, J. C.; Plückthun, A.; Nettels, D.; Schuler, B. *Nat. Methods* **2015,** *12,* 773–779. doi:10.1038/nmeth.3475
6. Sustarsic, M.; Kapanidis, A. N. *Curr. Opin. Struct. Biol.* **2015,** *34,* 52–59. doi:10.1016/j.sbi.2015.07.001
7. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy,* 2nd ed.; Springer: New York, 2006. doi:10.1007/978-0-387-46312-4
8. Dolghih, E.; Roitberg, A. E.; Krause, J. L. *J. Photochem. Photobiol., A* **2007,** *190,* 321–327. doi:10.1016/j.jphotochem.2006.11.009
9. Norman, D. G.; Grainger, R. J.; Uhrin, D.; Lilley, D. M. J. *Biochemistry* **2000,** *39,* 6317–6324. doi:10.1021/bi992944a
10. Iqbal, A.; Wang, L.; Thompson, K. C.; Lilley, D. M. J.; Norman, D. G. *Biochemistry* **2008,** *47,* 7857–7862. doi:10.1021/bi800773f
11. Hurley, D. J.; Tor, Y. *J. Am. Chem. Soc.* **2002,** *124,* 13231–13241. doi:10.1021/ja020172r
12. Lewis, F. D.; Zhang, L.; Zuo, X. *J. Am. Chem. Soc.* **2005,** *127,* 10002–10003. doi:10.1021/ja0524402
13. Iqbal, A.; Arslan, S.; Okumus, B.; Wilson, T. J.; Giraud, G.; Norman, D. G.; Ha, T.; Lilley, D. M. J. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 11176–11181. doi:10.1073/pnas.0801707105
14. Börjesson, K.; Preus, S.; El-Sagheer, A. H.; Brown, T.; Albinsson, B.; Wilhelmsson, L. M. *J. Am. Chem. Soc.* **2009,** *131,* 4288–4293. doi:10.1021/ja806944w

15. Wilhelmsson, L. M. *Q. Rev. Biophys.* **2010,** *43,* 159–183. doi:10.1017/S0033583510000090
16. Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010,** *110,* 2579–2619. doi:10.1021/cr900301e
17. Kato, T.; Kashida, H.; Kishida, H.; Yada, H.; Okamoto, H.; Asanuma, H. *J. Am. Chem. Soc.* **2013,** *135,* 741–750. doi:10.1021/ja309279w
18. Beija, M.; Afonso, C. A. M.; Martinho, J. M. G. *Chem. Soc. Rev.* **2009,** *38,* 2410–2433. doi:10.1039/b901612k
19. Shindy, H. A. *Dyes Pigm.* **2017,** *145,* 505–513. doi:10.1016/j.dyepig.2017.06.029
20. Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem.* **1969,** *244,* 1228–1237.
21. Ben Gaied, N.; Glasser, N.; Ramalanjaona, N.; Beltz, H.; Wolff, P.; Marquet, R.; Burger, A.; Mély, Y. *Nucleic Acids Res.* **2005,** *33,* 1031–1039. doi:10.1093/nar/gki253
22. Matarazzo, A.; Hudson, R. H. E. *Tetrahedron* **2015,** *71,* 1627–1657. doi:10.1016/j.tet.2014.12.066
23. Shin, D.; Sinkeldam, R. W.; Tor, Y. *J. Am. Chem. Soc.* **2011,** *133,* 14912–14915. doi:10.1021/ja206095a
24. Suzuki, A.; Takahashi, N.; Okada, Y.; Saito, I.; Nemoto, N.; Saito, Y. *Bioorg. Med. Chem. Lett.* **2013,** *23,* 886–892. doi:10.1016/j.bmcl.2012.11.029
25. Wranne, M. S.; Füchtbauer, A. F.; Dumat, B.; Bood, M.; El-Sagheer, A. H.; Brown, T.; Gradén, H.; Grøtli, M.; Wilhelmsson, L. M. *J. Am. Chem. Soc.* **2017,** *139,* 9271–9280. doi:10.1021/jacs.7b04517
26. Driscoll, S. L.; Hawkins, M. E.; Balis, F. M.; Pfleiderer, W.; Laws, W. R. *Biophys. J.* **1997,** *73,* 3277–3286. doi:10.1016/S0006-3495(97)78352-8
27. Dumas, A.; Luedtke, N. W. *Nucleic Acids Res.* **2011,** *39,* 6825–6834. doi:10.1093/nar/gkr281
28. Tokugawa, M.; Masaki, Y.; Canggadibrata, J. C.; Kaneko, K.; Shiozawa, T.; Kanamori, T.; Grøtli, M.; Wilhelmsson, L. M.; Sekine, M.; Seio, K. *Chem. Commun.* **2016,** *52,* 3809–3812. doi:10.1039/C5CC09700B
29. Wilhelmsson, L. M.; Holmén, A.; Lincoln, P.; Nielsen, P. E.; Nordén, B. *J. Am. Chem. Soc.* **2001,** *123,* 2434–2435. doi:10.1021/ja0025797
30. Sandin, P.; Wilhelmsson, L. M.; Lincoln, P.; Powers, V. E. C.; Brown, T.; Albinsson, B. *Nucleic Acids Res.* **2005,** *33,* 5019–5025. doi:10.1093/nar/gki790
31. Sandin, P.; Börjesson, K.; Li, H.; Mårtensson, J.; Brown, T.; Wilhelmsson, L. M.; Albinsson, B. *Nucleic Acids Res.* **2008,** *36,* 157–167. doi:10.1093/nar/gkm1006
32. Berry, D. A.; Jung, K.-Y.; Wise, D. S.; Sercel, A. D.; Pearson, W. H.; Mackie, H.; Randolph, J. B.; Somers, R. L. *Tetrahedron Lett.* **2004,** *45,* 2457–2461. doi:10.1016/j.tetlet.2004.01.108
33. Wojciechowski, F.; Hudson, R. H. E. *J. Am. Chem. Soc.* **2008,** *130,* 12574–12575. doi:10.1021/ja804233g
34. Mata, G.; Luedtke, N. W. *J. Am. Chem. Soc.* **2015,** *137,* 699–707. doi:10.1021/ja508741u
35. Wilhelmsson, L. M.; Sandin, P.; Holmén, A.; Albinsson, B.; Lincoln, P.; Nordén, B. *J. Phys. Chem. B* **2003,** *107,* 9094–9101. doi:10.1021/jp034930r
36. Engman, K. C.; Sandin, P.; Osborne, S.; Brown, T.; Billeter, M.; Lincoln, P.; Nordén, B.; Albinsson, B.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2004,** *32,* 5087–5095. doi:10.1093/nar/gkh844
37. Sandin, P.; Lincoln, P.; Brown, T.; Wilhelmsson, L. M. *Nat. Protoc.* **2007,** *2,* 615–623. doi:10.1038/nprot.2007.80
38. Preus, S.; Börjesson, K.; Kilså, K.; Albinsson, B.; Wilhelmsson, L. M. *J. Phys. Chem. B* **2010,** *114,* 1050–1056. doi:10.1021/jp909471b
39. Roth, B.; Schloemer, L. A. *J. Org. Chem.* **1963,** *28,* 2659–2672. doi:10.1021/jo01045a042
40. Roth, B.; Hitchings, G. H. *J. Org. Chem.* **1961,** *26,* 2770–2778. doi:10.1021/jo01066a035
41. Lin, K.-Y.; Jones, R. J.; Matteucci, M. *J. Am. Chem. Soc.* **1995,** *117,* 3873–3874. doi:10.1021/ja00118a026
42. Kazimierczuk, Z.; Cottam, H. B.; Revankar, G. R.; Robins, R. K. *J. Am. Chem. Soc.* **1984,** *106,* 6379–6382. doi:10.1021/ja00333a046
43. Schaller, H.; Weimann, G.; Lerch, B.; Khorana, H. G. *J. Am. Chem. Soc.* **1963,** *85,* 3821–3827. doi:10.1021/ja00906a021
44. Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981,** *22,* 1859–1862. doi:10.1016/S0040-4039(01)90461-7
45. Preus, S.; Kilså, K.; Wilhelmsson, L. M.; Albinsson, B. *Phys. Chem. Chem. Phys.* **2010,** *12,* 8881–8892. doi:10.1039/C000625D
46. Hoffer, M. *Chem. Ber.* **1960,** *93,* 2777–2781. doi:10.1002/cber.19600931204
47. Rodgers, B. J.; Elsharif, N. A.; Vashisht, N.; Mingus, M. M.; Mulvahill, M. A.; Stengel, G.; Kuchta, R. D.; Purse, B. W. *Chem. – Eur. J.* **2014,** *20,* 2010–2015. doi:10.1002/chem.201303410
48. Niedballa, U.; Vorbrueggen, H. *J. Org. Chem.* **1974,** *39,* 3668–3671. doi:10.1021/jo00939a011
49. Appel, R. *Angew. Chem., Int. Ed. Engl.* **1975,** *14,* 801–811. doi:10.1002/anie.197508011
50. Füchtbauer, A. F.; Preus, S.; Börjesson, K.; McPhee, S. A.; Lilley, D. M. J.; Wilhelmsson, L. M. *Sci. Rep.* **2017,** *7,* No. 2393. doi:10.1038/s41598-017-02453-1
51. Buhr, C. A.; Matteucci, M. D.; Froehler, B. C. *Tetrahedron Lett.* **1999,** *40,* 8969–8970. doi:10.1016/S0040-4039(99)01875-4
52. Nair, V.; Turner, G. A.; Chamberlain, S. D. *J. Am. Chem. Soc.* **1987,** *109,* 7223–7224. doi:10.1021/ja00257a071
53. Dierckx, A.; Miannay, F.-A.; Ben Gaied, N.; Preus, S.; Björck, M.; Brown, T.; Wilhelmsson, L. M. *Chem. – Eur. J.* **2012,** *18,* 5987–5997. doi:10.1002/chem.201103419
54. Thompson, A. L. S.; Kabalka, G. W.; Akula, M. R.; Huffman, J. W. *Synthesis* **2005,** 547–550. doi:10.1055/s-2005-861791
55. Dumat, B.; Bood, M.; Wranne, M. S.; Lawson, C. P.; Larsen, A. F.; Preus, S.; Streling, J.; Gradén, H.; Wellner, E.; Grøtli, M.; Wilhelmsson, L. M. *Chem. – Eur. J.* **2015,** *21,* 4039–4048. doi:10.1002/chem.201405759
56. Larsen, A. F.; Dumat, B.; Wranne, M. S.; Lawson, C. P.; Preus, S.; Bood, M.; Gradén, H.; Wilhelmsson, L. M.; Grøtli, M. *Sci. Rep.* **2015,** *5,* No. 12653. doi:10.1038/srep12653
57. Han, J. H.; Yamamoto, S.; Park, S.; Sugiyama, H. *Chem. – Eur. J.* **2017,** *23,* 7607–7613. doi:10.1002/chem.201701118
58. Preus, S.; Kilså, K.; Miannay, F.-A.; Albinsson, B.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2013,** *41,* e18. doi:10.1093/nar/gks856
59. Wozniak, A. K.; Schröder, G. F.; Grubmüller, H.; Seidel, C. A. M.; Oesterhelt, F. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 18337–18342. doi:10.1073/pnas.0800977105
60. Gohlke, C.; Murchie, A. I. H.; Lilley, D. M. J.; Clegg, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91,* 11660–11664. doi:10.1073/pnas.91.24.11660
61. Zacharias, M.; Hagerman, P. J. *J. Mol. Biol.* **1995,** *247,* 486–500. doi:10.1006/jmbi.1995.0155
62. Dimura, M.; Peulen, T. O.; Hanke, C. A.; Prakash, A.; Gohlke, H.; Seidel, C. A. M. *Curr. Opin. Struct. Biol.* **2016,** *40,* 163–185. doi:10.1016/j.sbi.2016.11.012
63. Peulen, T.-O.; Opanasyuk, O.; Seidel, C. A. M. *J. Phys. Chem. B* **2017,** *121,* 8211–8241. doi:10.1021/acs.jpcb.7b03441

64. Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. *Nature* **1979,** *282,* 680–686. doi:10.1038/282680a0
65. Rich, A.; Zhang, S. *Nat. Rev. Genet.* **2003,** *4,* 566–572. doi:10.1038/nrg1115
66. Liu, R.; Liu, H.; Chen, X.; Kirby, M.; Brown, P. O.; Zhao, K. *Cell* **2001,** *106,* 309–318. doi:10.1016/S0092-8674(01)00446-9
67. Pohl, F. M.; Jovin, T. M. *J. Mol. Biol.* **1972,** *67,* 375–396. doi:10.1016/0022-2836(72)90457-3
68. González, V. M.; Fuertes, M. A.; Pérez, J. M.; Alonso, C. *Eur. Biophys. J.* **1998,** *27,* 417–423. doi:10.1007/s002490050150
69. Chen, C.-w..; Knop, R. H.; Cohen, J. S. *Biochemistry* **1983,** *22,* 5468–5471. doi:10.1021/bi00293a002
70. Sheardy, R. D.; Levine, N.; Marotta, S.; Suh, D.; Chaires, J. B. *Biochemistry* **1994,** *33,* 1385–1391. doi:10.1021/bi00172a014
71. Sheardy, R. D.; Suh, D.; Kurzinsky, R.; Doktycz, M. J.; Benight, A. S.; Chaires, J. B. *J. Mol. Biol.* **1993,** *231,* 475–488. doi:10.1006/jmbi.1993.1295
72. Sheardy, R. D.; Winkle, S. A. *Biochemistry* **1989,** *28,* 720–725. doi:10.1021/bi00428a046
73. Dumat, B.; Larsen, A. F.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2016,** *44,* e101. doi:10.1093/nar/gkw114
74. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U. S. A.* **1985,** *82,* 1376–1380. doi:10.1073/pnas.82.5.1376
75. Lipfert, J.; Klijnhout, S.; Dekker, N. H. *Nucleic Acids Res.* **2010,** *38,* 7122–7132. doi:10.1093/nar/gkq598
76. Premvardhan, L.; Maurizot, J.-C. *Eur. Biophys. J.* **2010,** *39,* 781–787. doi:10.1007/s00249-009-0550-x
77. Triebel, H.; Bär, H.; Geuther, R.; Burckhardt, G. Netropsin-induced changes of DANN supercoiling; sedimentation studies. In *Analytical Ultracentrifugation;* Behlke, J., Ed.; *Progress in Colloid and Polymer Science*, Vol. 99; Steinkopff Verlag: Dresden, 1995; pp 45–54. doi:10.1007/BFb0114069
78. Snounou, G.; Malcolm, A. D. B. *J. Mol. Biol.* **1983,** *167,* 211–216. doi:10.1016/S0022-2836(83)80043-6
79. Rettig, M.; Germann, M. W.; Wang, S.; Wilson, W. D. *ChemBioChem* **2013,** *14,* 323–331. doi:10.1002/cbic.201200706
80. Shi, Y.; Dierckx, A.; Wanrooij, P. H.; Wanrooij, S.; Larsson, N.-G.; Wilhelmsson, L. M.; Falkenberg, M.; Gustafsson, C. M. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109,* 16510–16515. doi:10.1073/pnas.1119738109
81. Posse, V.; Hoberg, E.; Dierckx, A.; Shahzad, S.; Koolmeister, C.; Larsson, N.-G.; Wilhelmsson, L. M.; Hällberg, B. M.; Gustafsson, C. M. *Nucleic Acids Res.* **2014,** *42,* 3638–3647. doi:10.1093/nar/gkt1397
82. Velmurugu, Y.; Chen, X.; Sevilla, P. S.; Min, J.-H.; Ansari, A. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* E2296–E2305. doi:10.1073/pnas.1514666113

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Fluorogenic PNA probes

Tirayut Vilaivan

Review

Open Access

Address:
Organic Synthesis Research Unit, Department of Chemistry, Faculty of Science, Chulalongkorn University, Phayathai Road, Patumwan, Bangkok 10330, Thailand

Email:
Tirayut Vilaivan - vtirayut@chula.ac.th

Keywords:
DNA; fluorescence; molecular beacons; molecular probes; oligonucelotides; RNA

Beilstein J. Org. Chem. **2018,** *14,* 253–281.
doi:10.3762/bjoc.14.17

Received: 03 October 2017
Accepted: 03 January 2018
Published: 29 January 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

Fluorogenic oligonucleotide probes that can produce a change in fluorescence signal upon binding to specific biomolecular targets, including nucleic acids as well as non-nucleic acid targets, such as proteins and small molecules, have applications in various important areas. These include diagnostics, drug development and as tools for studying biomolecular interactions in situ and in real time. The probes usually consist of a labeled oligonucleotide strand as a recognition element together with a mechanism for signal transduction that can translate the binding event into a measurable signal. While a number of strategies have been developed for the signal transduction, relatively little attention has been paid to the recognition element. Peptide nucleic acids (PNA) are DNA mimics with several favorable properties making them a potential alternative to natural nucleic acids for the development of fluorogenic probes, including their very strong and specific recognition and excellent chemical and biological stabilities in addition to their ability to bind to structured nucleic acid targets. In addition, the uncharged backbone of PNA allows for other unique designs that cannot be performed with oligonucleotides or analogues with negatively-charged backbones. This review aims to introduce the principle, showcase state-of-the-art technologies and update recent developments in the areas of fluorogenic PNA probes during the past 20 years.

Review

Introduction

The development of molecular probes that can detect and quantify specific biological molecules with a high degree of sensitivity and accuracy, preferably in situ and in real-time, is an important research area since it contributes to the advancement of understanding of complex biomolecular systems and has practical applications in diverse areas, including clinical diagnostics and drug discovery amongst others. Molecular probes generally consist of a recognition element that can bind to the specific target, and a reporter group that, in combination with an appropriate signal transduction mechanism, translates the molecular interaction into a measurable signal. Among several available modes of detection, fluorescence detection offers a number of

distinct advantages, but mainly that it is relatively simple, selective, highly sensitive and can be used for real-time monitoring of the biological targets and even their interactions in vivo. For nucleic acids, the latter valuable information cannot be obtained by sequencing despite the tremendous advances in the field in recent years [1].

Accordingly, fluorescent oligonucleotide probes are still one of the most important tools not only for the detection of nucleic acids, but also proteins and other non-nucleic acid targets by employing aptamer technology [2]. However, ordinary fluorescent oligonucleotide probes generally show indistinguishable signals between the free and target-bound states. This means that additional treatments are required in order to separate the bound and unbound probes, which is most commonly achieved by solid phase hybridization followed by washing to eliminate the unbound probes before performing the fluorescence readout. These assays require multiple steps, and so are time-consuming making them unsuitable for real-time monitoring, such as nucleic acid amplification, monitoring of enzyme activities and localization and quantitation of nucleic acids in living cells. Therefore, it is highly desirable to develop a "smart" fluorogenic oligonucleotide probe that can directly report the presence of the target by a simple fluorescence readout in homogeneous and wash-free format. One of the landmark developments in this area is the DNA molecular beacon – a double-labeled hairpin DNA that can change its fluorescence in response to the conformational change induced by binding with the complementary nucleic acid target (Figure 1) [3,4].

While a number of signal transduction and readout strategies have been developed, even for fluorescence-based detection alone, the recognition element of the probe is almost always an oligodeoxynucleotide. During the past few decades, there have been continuous developments of unnatural oligonucleotides with superior properties that make them worthwhile alternatives to the traditionally used oligodeoxynucleotide probes. In this respect, peptide nucleic acids (PNA) [5] have been one of the most widely used alternative oligonucleotide probes in addition to locked and morpholino nucleic acids [6,7]. Although many review articles on fluorogenic oligonucleotide probes exist [8-17], very few have specifically dealt with fluorogenic unnatural oligonucleotide probes [18,19]. This review aims to fill this gap by introducing the principle, showcasing state-of-the-art technologies and updating recent developments in the areas of fluorogenic PNA probes. Some examples of PNA probes for nucleic acid hybridization and detection that may not be strictly fluorogenic by definition have been included where deemed suitable.

Overview of PNA and fluorescent PNA probes

Peptide nucleic acids (PNA) are a unique class of oligonucleotide mimics that consist of a peptide-like backbone. Although the idea of replacing the whole sugar-phosphate backbone of DNA with a completely unrelated scaffold such as peptide had been around since the 1970s [20], it was not until the 1990s that the first PNA system with an *N*-2-aminoethylglycine (aeg) backbone that can recognize its target DNA and RNA was reported [21]. Considering the enormous difference between the two backbones, it is quite surprising that PNA can still retain the ability to recognize natural oligonucleotides having a complementary sequence with high affinity and specificity according to the Watson–Crick base pairing rules [22]. In fact, PNA exhibits an even higher affinity and better discrimination between complementary and mismatched nucleic acid targets than natural oligonucleotides. In addition, the uncharged peptide-like backbone of PNA contributes to several unique properties not observed in other classes of oligonucleotide analogues with negatively charged phosphate groups. These include the relative insensitivity of the PNA–DNA or PNA–RNA hybrids to the ionic strength of the solvent [23], and the complete stability towards nucleases as well as proteases [24].

Overall, these properties enable the use of PNA for several applications, including therapeutics and diagnostics [25], and as a unique tool for DNA manipulation, such as PCR clamping [26] and PNA openers [27]. Ironically, the very same uncharged

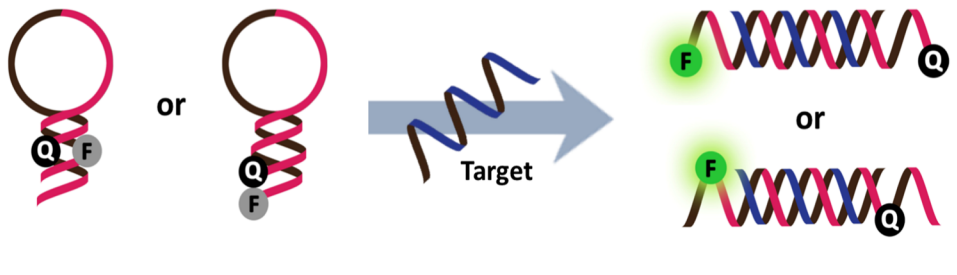


Figure 1: The design of classical DNA molecular beacons.

nature of PNA that is the basis of several of the aforementioned advantages poses new challenges, including a poor aqueous solubility and the tendency to bind non-specifically to hydrophobic surfaces, including other PNA molecules to form aggregates. These can be largely overcome by conjugation of the PNA to charged or hydrophilic groups or by backbone modification [28]. The simplicity of the PNA structure offers unlimited possibilities to design new PNA systems with improved solubilities and other properties and to incorporate new functions into the PNA molecule [29]. Some notable performance improvements of the original PNA system based on the aeg backbone have been realized by constraining the conformational flexibility by incorporating a suitable substituent at a suitable position (such as in γPNA) [30], or by incorporating cyclic structures into the PNA backbones (such as in acpcPNA, Figure 2) [31,32].

The high binding affinity and excellent specificity of PNA towards their respective nucleic acid targets immediately suggested their potential use as diagnostic probes. In addition, a broad range of salt tolerance and stability against most enzymes are added benefits of PNA probes. Moreover, the ability of PNA to bind to double stranded (ds)DNA (Figure 3) [33], dsRNA [34], and other unusual structures, such as G-quadruplexes [35], makes PNA an ideal tool for targeting structured nucleic acid targets. Simple fluorescent-labeled PNA probes have found extensive applications in nucleic acid detection and quantitation. Examples of such assays that have successfully employed PNA probes include array hybridization [36,37] and staining of intracellular nucleic acids by fluorescence in situ hybridization (FISH) [38,39]. For the same reasons explained above in the case of oligonucleotide probes, it is more desirable to develop a fluorogenic PNA probe that can change its fluorescence signal in response to hybridization to its specific target. The most obvious strategy would be to employ the same principle as DNA beacons, whereby two interacting dyes are placed on a hairpin-forming PNA probe that can switch to an open conformation upon target hybridization [3,4].

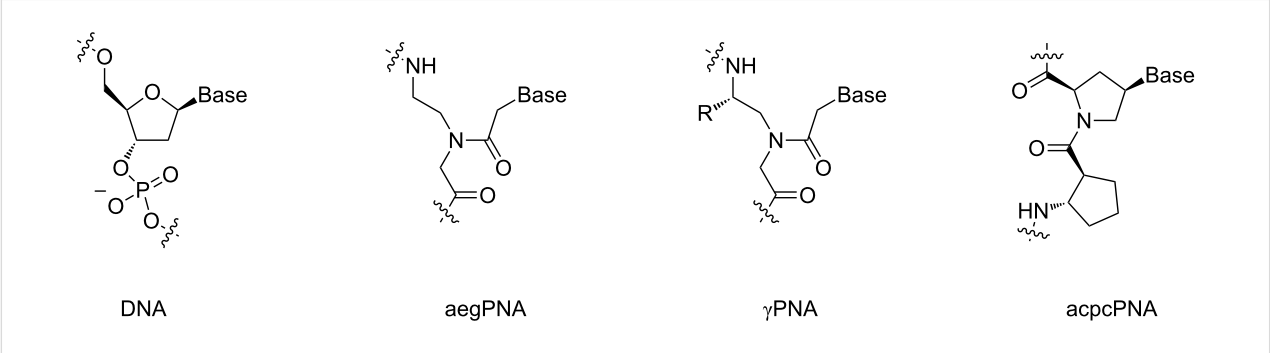


Figure 2: Structures of DNA and selected PNA systems.

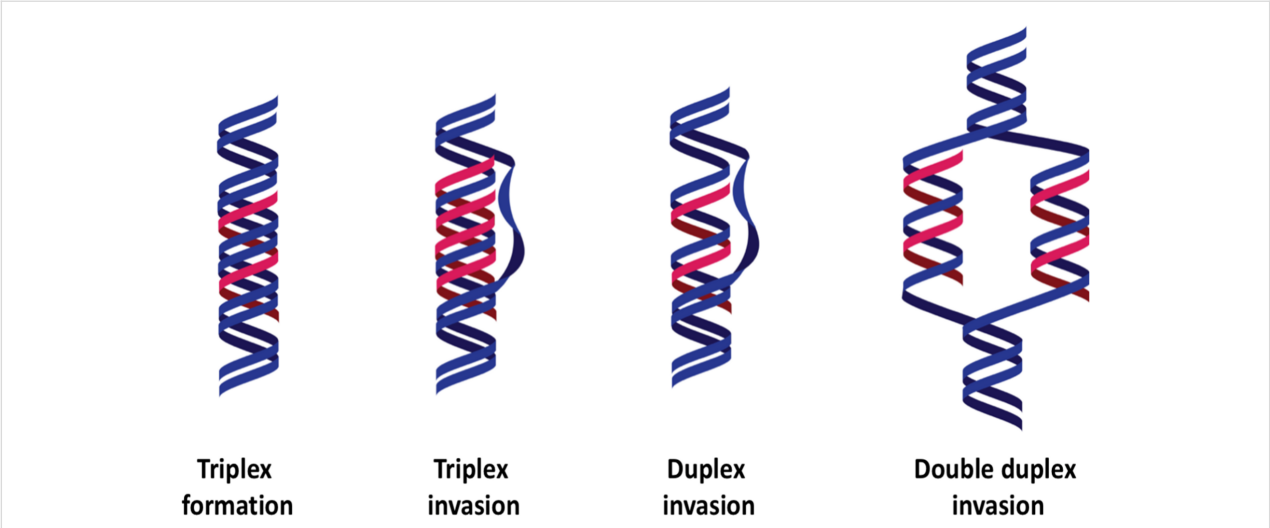


Figure 3: Various binding modes of PNA to double stranded DNA including triplex formation, triplex invasion, duplex invasion and double duplex invasion.

It has been reported that short, linear PNA beacons can perform surprisingly well when compared to linear DNA beacons [40]. This could be attributed to both the ability of PNA to form a compact structure in aqueous media, thereby forcing the two dyes in close contact, and the excellent mismatch discrimination ability of PNA. Several other strategies originally developed for fluorogenic oligonucleotide probes have also been successfully applied to PNA probes. These include binary probes and nucleic acid-templated reactions, strand displacement probes and the combination of ordinary PNA probes with nanomaterials as an external quencher. In addition, the uncharged backbone of PNA offers other unique designs, including the combination of PNA probes with cationic conjugated polymers that simultaneously act as a light harvesting antenna and fluorescent resonance energy transfer (FRET) partner, and the combination of short linear PNA probes with environment-sensitive dyes, such as in light-up and forced intercalation (FIT) PNA probes.

PNA probes carrying two or more interacting dyes in the same strand

The very first PNA probe of this type were independently reported in 1998 by Ortiz et al. [41] and Armitage et al. [42]. These first-generation PNA beacons employed a similar design principle to that of classical DNA beacons, having a hairpin structure carrying a fluorophore and a quencher attached to the stem part. In the normal state, the adopted hairpin structure forced the fluorescence dye and quencher to be in close proximity, resulting in quenching of the dye. Binding to the complementary DNA target opened the hairpin structure and restored the fluorescence. In Armitage's design [42], the entire beacon was purely PNA, while in the Ortiz design [41], PNA was employed as the DNA binding domain and the stem part consisted of a chimeric PNA–DNA hybrid linked together by a disulfide bond (Figure 4). The chimeric beacon was immobilized to a microtiter plate via a biotin tag and was employed for the detection of PCR amplicons of rDNA from *Entamoeba histolytica* [41]. The use of PNA allowed direct detection of double stranded (ds)DNA targets after simple denaturation due to the ability of PNA to bind more strongly to DNA and so effectively compete with the DNA re-association.

Another research group from Japan also reported a similar stem-loop non-chimeric PNA beacon carrying two labels (TAMRA/Dabcyl) at both ends of the PNA molecule and demonstrated that the fluorescence changed in response to the presence of the DNA target [43]. The high stability of PNA–PNA self-hybrids allows the design of a beacon with a relatively short stem (four bases). In this particular case, an interesting pH dependency of the performance of the PNA beacon was observed, but no explanation was offered.

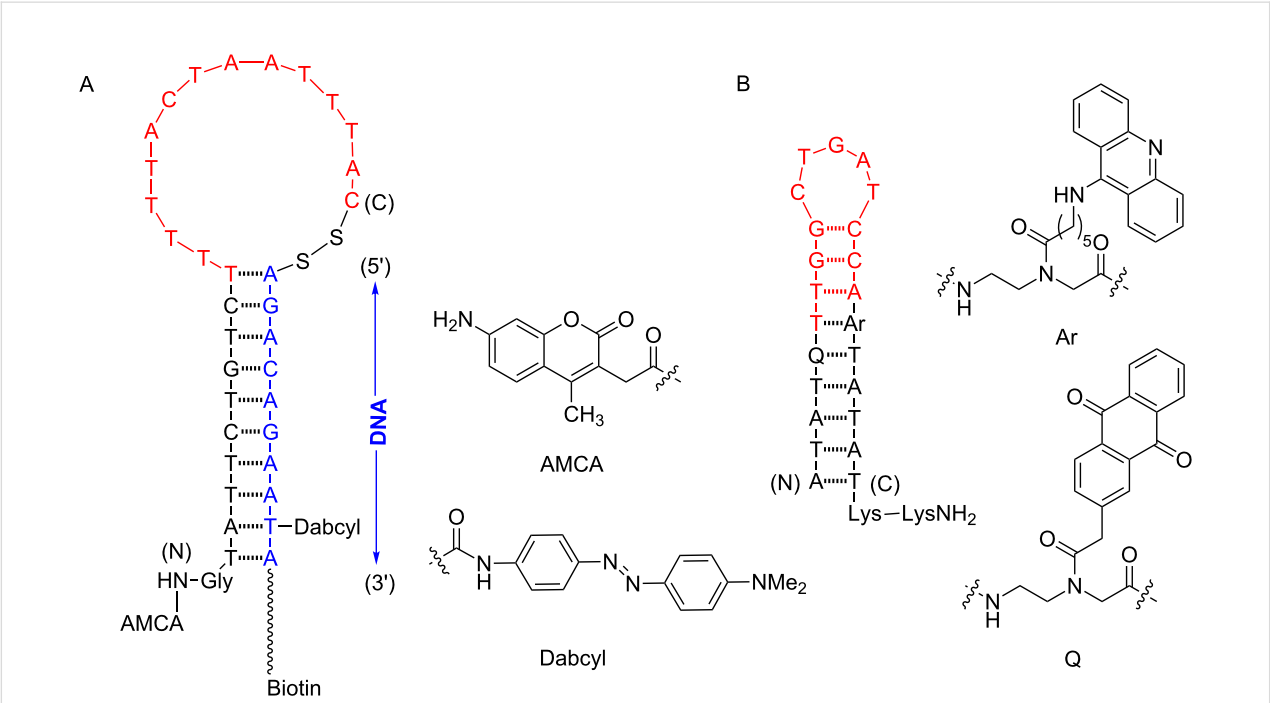


Figure 4: The design and working principle of the PNA beacons according to (A) Ortiz et al. [41] and (B) Armitage et al. [42]. The DNA binding domains are shown in red.

A new generation of PNA beacons was simultaneously reported by Seitz [44,45] and a group at Boston Probes [40,46]. According to this design, a short linear PNA strand was double-labeled at both ends with a quencher and a fluorophore (Figure 5). Seitz proposed that the stem-loop structure was not required because the hydrophobic backbone of PNA tends to adopt a compact structure to minimize hydrophobic surfaces, allowing the fluorophore and quencher labeled at opposite ends of the PNA molecule to make a close contact. However, based on their detailed comparative studies between stemless PNA and DNA beacons, the Boston Probes group argued that the quenching was more likely due to hydrophobic and electrostatic interactions between the fluorophore and the quencher. Nevertheless, different fluorophore-quencher pairs were used in the two cases, and so no single explanation may be applicable to all circumstances. What is clear is that the stemless, unstructured linear PNA beacons showed an impressive performance in DNA detection in terms of an excellent specificity, very fast hybridization kinetics and high signal-to-background ratio regardless of the salt concentration. Moreover, it can be applied for the direct detection of dsDNA targets under non-denaturing conditions either directly for short dsDNA sequences [47] or with the assistance of a pair of PNA openers for longer dsDNA targets (Figure 6) [40].

To ensure a close contact between the two labels, amino acids with opposite charges (typically glutamic acid and lysine) are usually placed in the vicinity of the labels. This was the practice originally used by the Boston Probe design and has been followed since by many others [40]. In another strategy, the PNA may be terminally functionalized with chelating ligands, such as iminodiacetic acid, iminotriacetic acid or terminal dihistidine, to form the so-called Snap-to-it probes (Figure 7) [48]. The addition of divalent metal ions, such as Ni^{2+}, results in an intramolecular chelation that contributes to the reduction in the background fluorescence of the single stranded (ss) probe by forcing the labels in close contact. In addition, the chelation also introduces an additional thermodynamic barrier into the binding of the DNA target to the unstructured PNA beacons that is akin to the difference observed between linear and hairpin DNA beacons [49]. These result in a significant (up to 40-fold) improvement in the signal-to-background ratio and target binding specificity [48].

A different strategy to eliminate the background signal problem from free PNA beacons is to use ion exchange HPLC with fluorescence detection which can separate the signals from the free and hybridized PNA probes into different channels. In this HPLC-based assay, the selectivity for single mismatch detec-

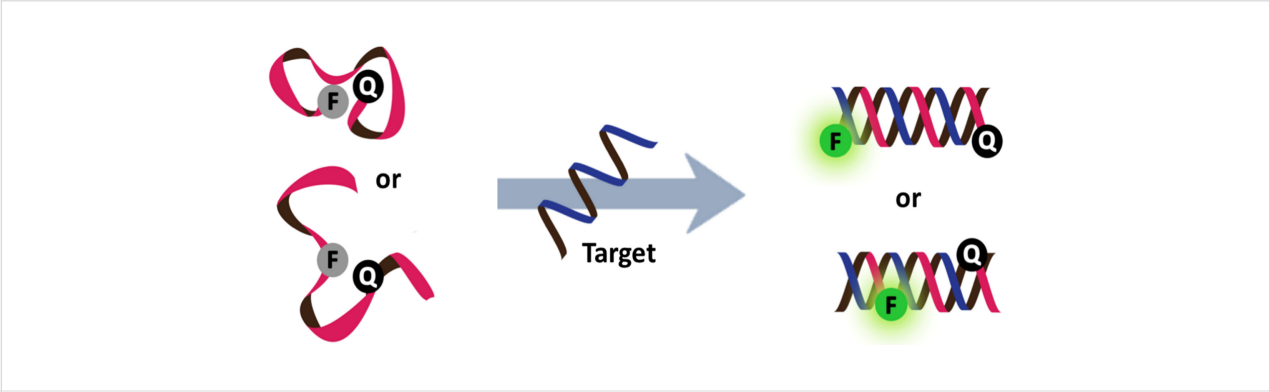


Figure 5: The design of "stemless" PNA beacons.

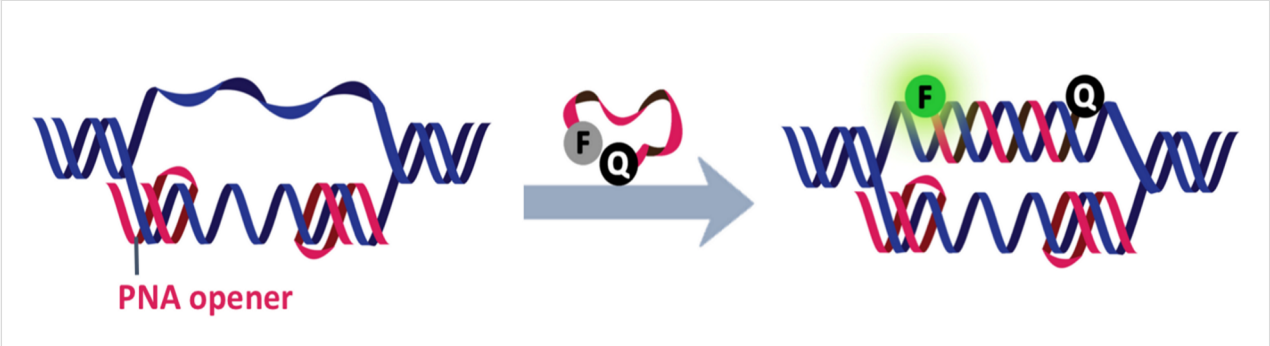


Figure 6: The applications of PNA openers to facilitate the binding of PNA beacons to double stranded DNA [40,47].

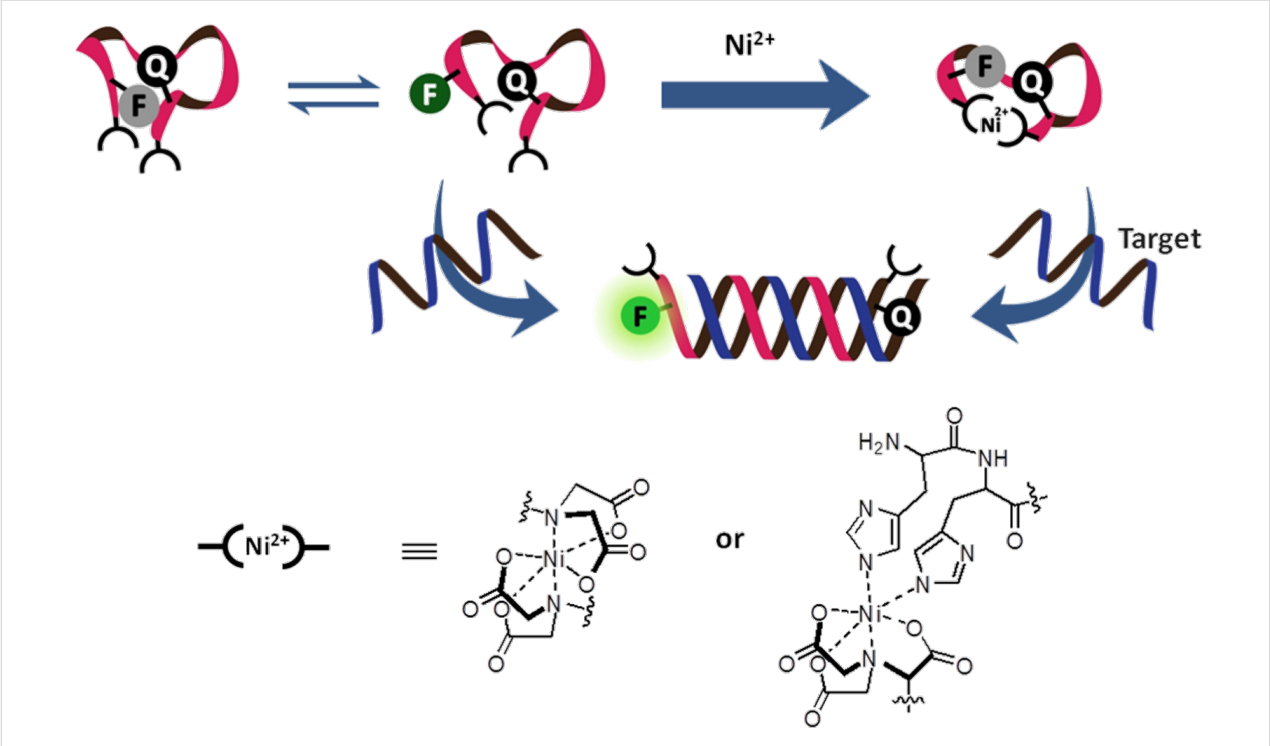


Figure 7: The working principle of snap-to-it probes that employed metal chelation to bring the dyes in close contact and to improve the mismatch discrimination by increasing the thermodynamic barrier of the probe–target binding [48].

tion was greatly improved compared to the homogeneous assay using the same probe and targets, which is probably due to the denaturing effects of the chromatographic conditions employed [50,51].

Most dual-labeled PNA beacons discussed so far carry a fluorophore and a quencher attached at opposite ends of the PNA strand via amide bonds. The advantage of this strategy is that the standard as-synthesized PNA can be used directly without having to modify the synthetic protocols, but it is difficult to vary the positions of the label attachment. The use of preformed dye labeled PNA monomers [42,52,53] or post-synthetically functionalizable PNA monomers [54-59] (Figure 8) permits facile incorporation of the labels anywhere in the molecular beacon, which may allow better control of the interactions between the two dyes and is, therefore, an important step towards a further improvement in the performance of the PNA probes.

In an alternative design, the fluorophore and either the quencher or another fluorophore may be placed in close proximity in the probe molecule, thereby maximizing the interactions between the two dyes and leading to a more effective quenching. Duplex formation will alter the interaction if one of the dyes can intercalate into the duplex or form a more stable end-stacking complex with the terminal base pair of the duplex (Figure 9). Several examples of DNA probes with this design are known [60-64]. This concept has been applied in the design of a double-end-labeled conformationally constrained acpcPNA probe for DNA sequence detection that showed a low background and good response with DNA, as well as offering an excellent specificity [65].

In addition to fluorophore-quencher interactions, FRET [66] and pyrene monomer–excimer switching [67,68] have been employed as alternative mechanisms for inducing fluorescence changes in the DNA probes. The FRET and monomer–excimer switching approaches have some advantages over fluorescence quenching because of the large Stokes shifts and the ability to measure the signals at two different wavelengths, thereby providing a means for self-referencing. Furthermore, unlike the fluorophore-quencher beacons, the FRET and monomer–excimer switching beacons are also fluorescent in the unbound state, and therefore it is possible to monitor the success of cellular delivery. In the case of pyrene monomer–excimer switching, the long fluorescence lifetime of the pyrene excimers allows facile elimination of background signals from autofluorescence by time-resolved fluorescence measurements. A few monomer–excimer switching PNA probes have been reported. Two or more pyrene labels may be conveniently placed anywhere in the PNA molecule by attaching them to a C5-functionalized thymine via amide or

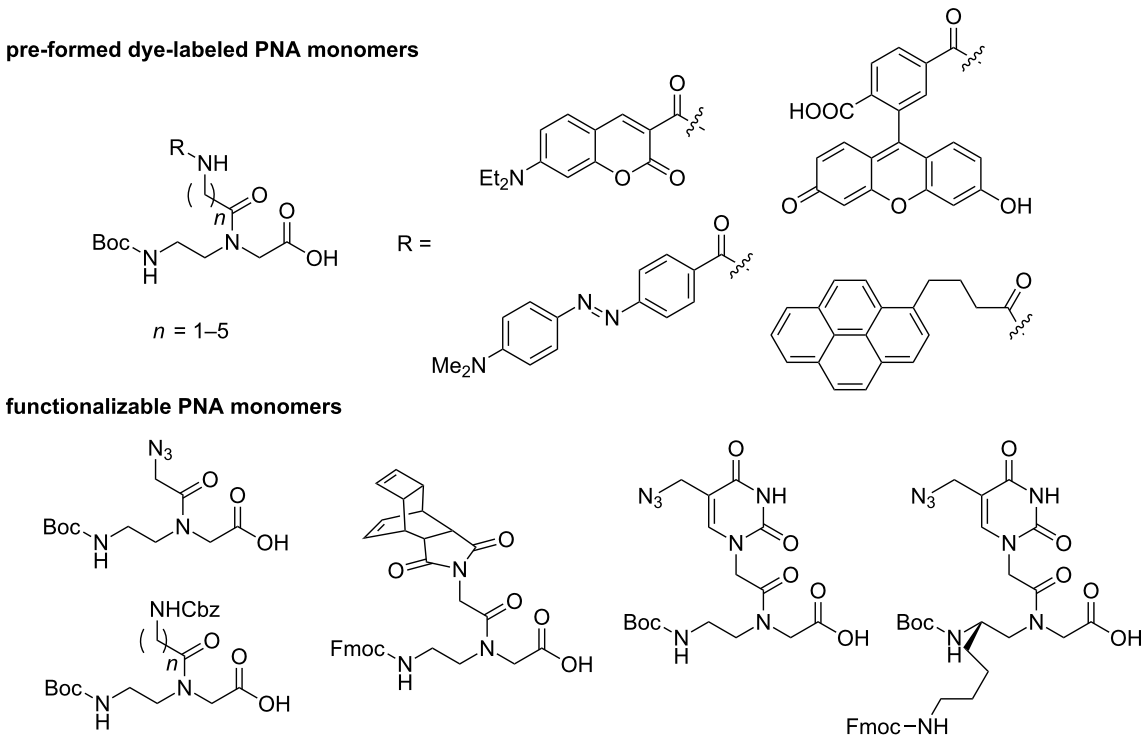


Figure 8: Examples of pre-formed dye-labeled PNA monomers and functionalizable PNA monomers.

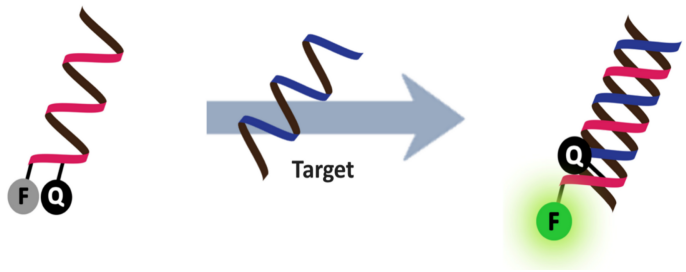


Figure 9: Dual-labeled PNA beacons with end-stacking or intercalating quencher.

click chemistries, and applications for the detection of DNA by duplex [54] or triplex [69] formation have been demonstrated. Alternatively, the pyrene label can be placed onto the PNA backbone as a base replacement (in aegPNA) [55] or as a tethered label through a flexible linker (in acpcPNA) [70].

In most cases, excimer emission is predominant in the single stranded probe, and hybridization with the DNA target resulted in an increased monomer emission with a simultaneous decrease in the excimer emission. A switching ratio of >30 upon hybridization with the correct DNA target was obtained with double-pyrene-labeled acpcPNA probes [70]. Interestingly, switching in the opposite direction, when the excimer signal is enhanced upon hybrid formation, was observed in triplex forming double-pyrene-labeled aegPNA probes [69]. For FRET-based PNA beacons, these have so far been investigated in the context of FIT PNA probes (vide infra), which will be discussed under the topic of PNA probes carrying fluorescent nucleobases (vide infra).

Dual-labeled PNA molecular beacons have been extensively used on their own or in combination with fluorescence melting curve analyses for genotyping [71-73] and analyses of genetic mutations, including single nucleotide polymorphisms (SNP), insertions and deletions [74-76]. They have also been used for the detection and quantification of ribosomal RNA in a wash-free FISH [77]. Side-by-side comparison showed that the PNA beacons gave faster hybridization kinetics, a higher signal-to-

noise ratio and a much better specificity than DNA beacons. A PNA beacon has been used as a sensing probe for PCR amplicons in a droplet-based microfluidic device [78], while an integrated microfluidic device that combined sample preparation, hybridization and confocal fluorescence spectroscopy allowed amplification-free detection of 16S RNA from a single cell of *Escherichia coli* [79]. In addition to the detection of DNA and RNA targets, hybrid PNA-peptide beacons were designed for the detection of non-DNA targets, such as proteins (Figure 10A). According to this design, which was simultaneously proposed in 2007 by Seitz [80] and Plaxco [81], the peptide part acted as the recognition element while the labeled stem-forming PNA part acted as the switch. Binding of the target molecule resulted in a conformational change, leading to a change in the fluorescence signal. The same principle can be applied for the detection of protease activities (Figure 10B) [82]. The advantages of using PNA as the switch over DNA would be the excellent biological stability and the compatibility of PNA with peptide chemistry. Obviously, the ease of the opening of the beacon could be fine-tuned to accommodate the different strengths of the biomolecular interactions to be investigated by adjusting the length and sequence of the PNA stem.

PNA-based binary probes and DNA/RNA-templated reaction of PNA probes

The working principle of binary probes involves the co-hybridization of two dye-labeled probes at adjacent positions on the same nucleic acid scaffold or template, which allows the two dyes to optically interact, such as by FRET or excimer formation (Figure 11) [83]. In principle, the hybridization of two very short PNA probes should offer a better selectivity than a single longer PNA probe. The relative positions of the two labels, as well as the types of linkers, need to be fine-tuned in order to

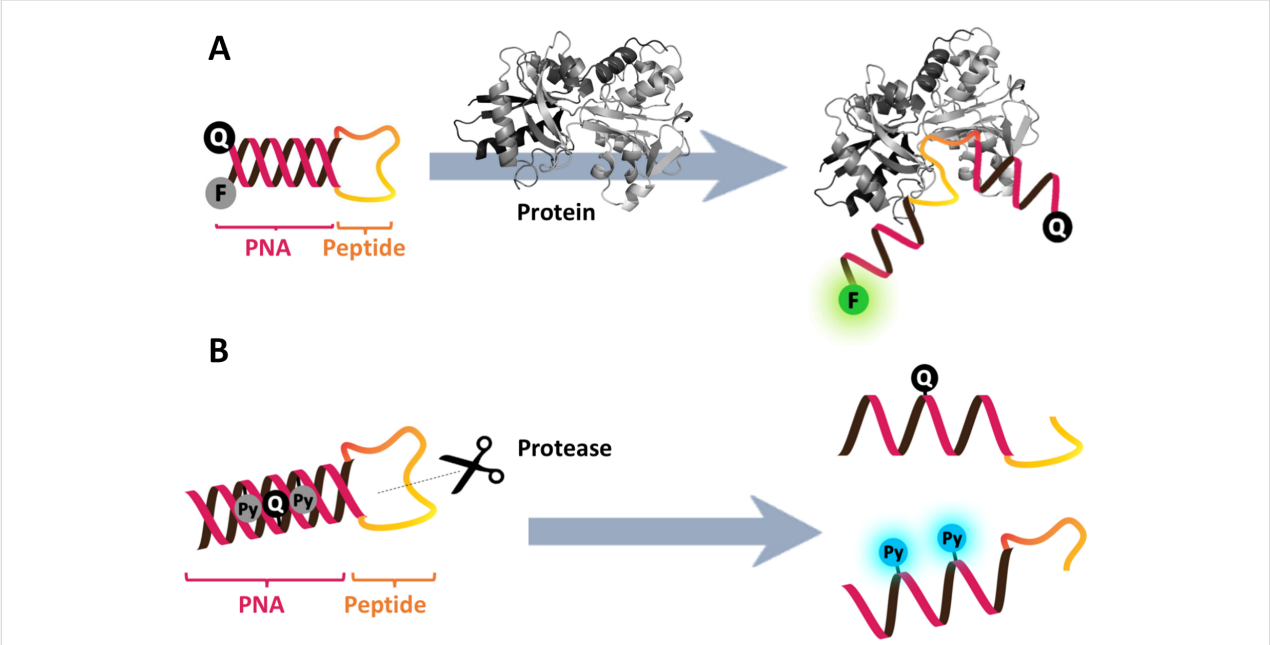


Figure 10: The working principle of hybrid PNA-peptide beacons for detection of (A) proteins [80] and (B) protease activities [82].

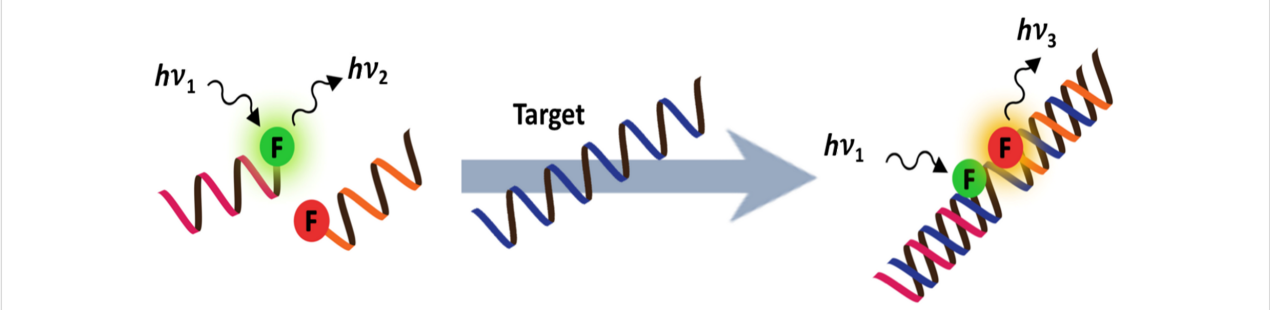


Figure 11: The working principle of binary probes.

obtain optimal results. In this respect, the ability to site-specifically label anywhere in the PNA molecule using a pre-formed dye-labeled monomer or a functionalized monomer that allows post-synthetic labeling is important [84]. Accordingly, FRET-based binary PNA probes have been used to monitor RNA splicing whereby the decrease in distance between the two probes after splicing resulted in an increased FRET efficiency [85]. A fluorescent PNA probe was used for the selective amplification of a rare DNA mutation in K-ras by PCR clamping and, at the same time, as a sensor probe in combination with another dye-labeled DNA probe in a single-tube operation by real-time fluorescence monitoring [86]. More challenging applications of binary PNA probes for imaging mRNA expression in living cells is possible at a proof-of-concept stage, but there is still room for improvement [87].

While numerous nucleic acid templated reactions involving oligonucleotide probes are known [88,89], only a few are PNA-based and fewer are fluorogenic. Each PNA probe carries one or more reactive groups that cannot react with each other at the low probe concentrations typically required for detection of the target molecules (nanomolar range or below). However, binding of the probes at adjacent positions on the same DNA or RNA template increases the local concentration of the probes so that they can readily react, resulting in creating and/or breaking one or more covalent bonds. They can be sub-divided into reactions that provide ligated products (Figure 12A) or non-ligated products (Figure 12B). In the latter case, the reacted probes readily dissociate and are replaced by unreacted probes, creating a catalytic cycle that leads to signal amplification (Figure 13) [90]. Signal amplification is also possible in the former case provided that the reaction was designed to give a ligated product that

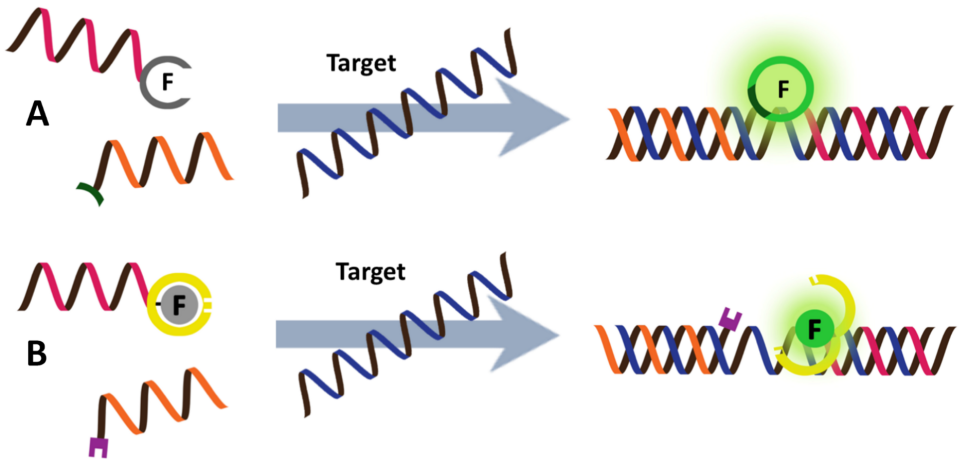


Figure 12: The working principle of nucleic acid templated fluorogenic reactions leading to a (A) ligated product and (B) non-ligated product.

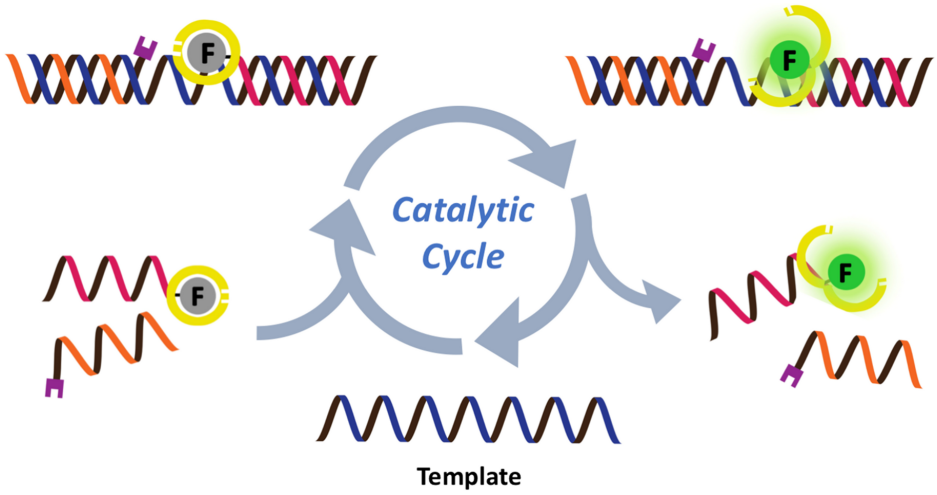


Figure 13: Catalytic cycles in fluorogenic nucleic acid templated reactions [90].

does not bind too tightly to the template, for example by using isocysteine-mediated native chemical ligation to provide a ligated PNA with a sub-optimal extended backbone at the ligation site [91].

Examples of the fluorogenic templated ligation of PNA are given in Table 1 which include the simple native chemical ligation of two fluorophore-labeled probes [92], formation of cyanine dyes [93,94], fluorogenic Michael addition [95,96] and cross-linking of dicysteine PNA probes that can bind with profluorescent bisarsenical dyes [97]. Interestingly, in the native ligation reaction, the FRET efficiency increased substantially following the ligation even though the positions of the two PNA probes hardly changed [92]. It should also be noted that the selectivity for matched over mismatched templates for these short PNA probes (>100:1) is higher than in the case of molecular beacons or other DNA-based probes that are typically much longer [92].

Examples of fluorogenic non-ligated templated reactions are summarized in Table 2. Many of these reactions feature the unmasking of quenched or caged fluorophores by chemical reactions, such as the Staudinger reaction [98-101], hydrolysis [102] and group transfer mediated by nucleophilic substitution [103,104]. Optimizing the conditions requires a good balance between the affinity of the probe to the template and the effective strand exchange. This can be achieved by performing the reaction at low probe concentrations and at a temperature close

Table 1: Examples of nucleic-acid-templated fluorogenic reactions of PNA probes leading to ligated products.

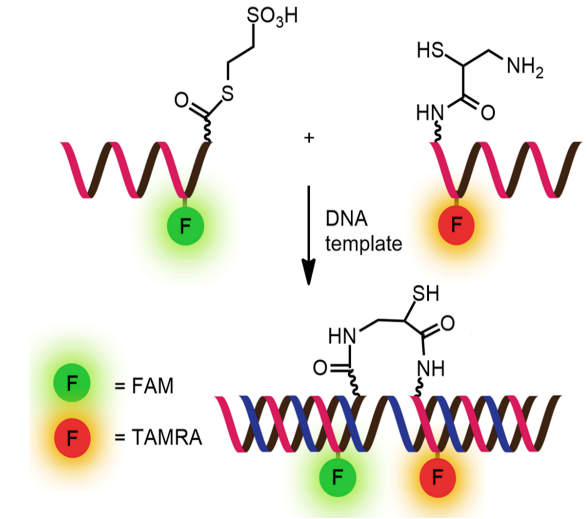
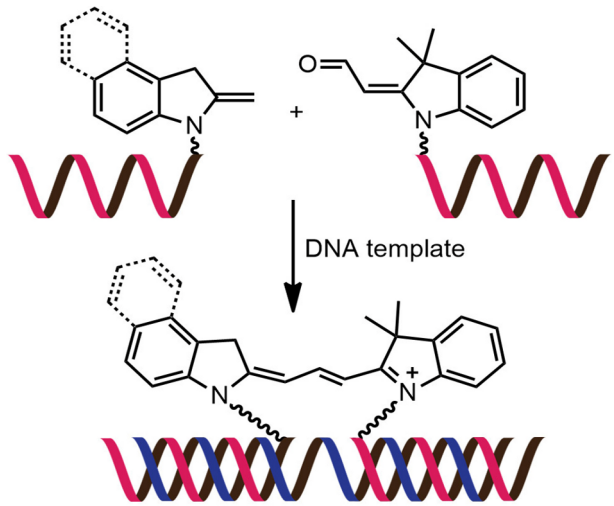
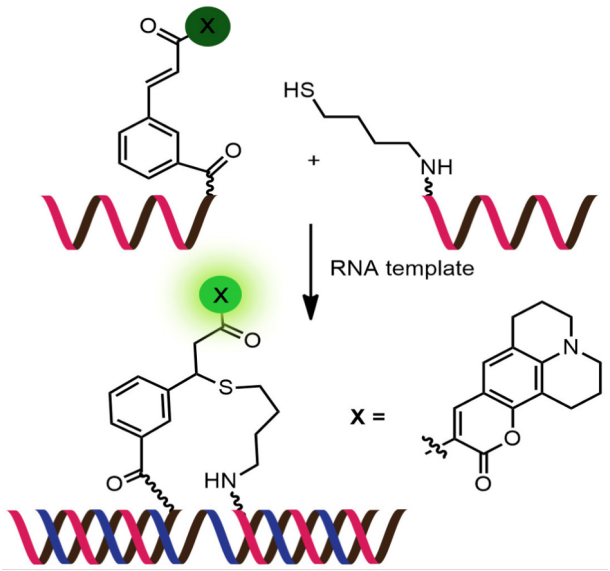
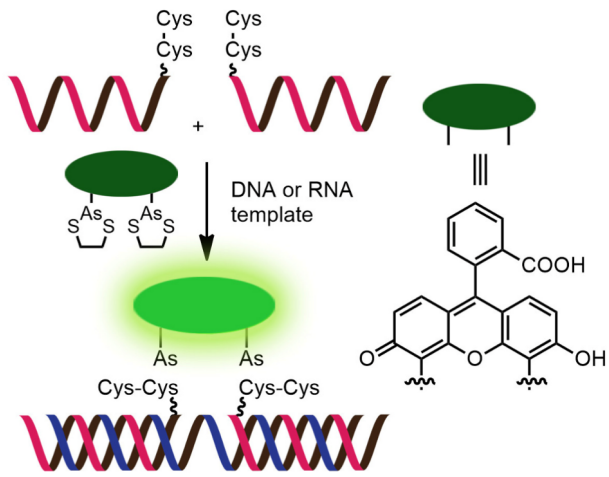
System	Remarks
	Reaction: native chemical ligation of two fluorescent labeled probes Probe: aegPNA Template: oligodeoxynucleotides Selectivity: 4.3 × 10^4 (relative to non-templated reaction); 10^2 to 10^3 (relative to single-mismatched template) LOD: N/A (all experiments were conducted at 1 μM or higher concentrations) Applications: in vitro detection of oligodeoxynucleotides Ref: [92]
	Reaction: formation of cyanine dyes Probe: aegPNA Template: oligodeoxynucleotides Selectivity: >50 (relative to non-templated reaction) LOD: <500 nM Applications: detection of conformational change of G-quadruplex- and hairpin-forming oligodeoxynucleotides Ref: [93,94]

Table 1: Examples of nucleic-acid-templated fluorogenic reactions of PNA probes leading to ligated products. (continued)

	Reaction: Michael addition Probe: aegPNA Template: oligoribonucleotide Selectivity: ≈2–3 (relative to single-mismatched template); up to 80 (relative to non-templated reaction) LOD: ≈60 nM; improved to 0.1 nM by confinement in hydrogels Applications: amplification-free detection and quantitation of extracted miRNA biomarkers for prostate cancers (miRNA-141, miRNA-132, miRNA-375) Ref: [95,96]
	Reaction: cross-linking with bisarsenical dyes Probe: aegPNA Template: oligodeoxynucleotide and oligoribonucleotide Selectivity: ≈16 (relative to single-mismatched template); >50 (relative to non-templated reaction) LOD: 2.5 nM (without amplification); subnanomolar sensitivity for miRNA detection (with rolling circle amplification, RCA) Applications: in vitro detection of miRNA let-7a Ref: [97]

to the melting temperature of the probe-template complex [104]. Impressive performances of several of these systems have been demonstrated, mostly with synthetic oligonucleotide templates, and applications for intracellular nucleic acid detection are also emerging [90,100]. The unmasking by template-catalyzed hydrolysis or thiolysis is susceptible to other non-selective pathways, such as enzymatic hydrolysis, giving rise to false positive signals or a high background. The masking of fluorophores by a more stable group, such as azides, is more attractive. Unfortunately, the phosphine probe generally required for the unmasking of the azide probe is susceptible to rapid aerobic oxidation. Accordingly, new developments in this area are still required. One promising example is the release of fluorophores by a phosphine-free photocatalyzed reduction of a self-immolative azide-based linker [105].

Single-labeled PNA probes with additional interacting partners

PNA-based strand displacement probes

Strand displacement probes consist of a fluorescence oligonucleotide probe strand annealed to another oligonucleotide strand that is labeled with a second dye and which can interact with the first dye by quenching or FRET. In the presence of the complementary nucleic acid target, strand displacement takes place and results in separation of the two dyes, and so gives rise to a fluorescence change (Figure 14). The strand displacement probe combines the advantages of the design simplicity of linear probes with the high specificity of hairpin molecular beacons [49]. Labeled PNA has occasionally been used in combination with another DNA as a strand displacement probe. One of the earliest examples makes use of PNA probes immobilized

on fluorescence-polymer-coated polystyrene microspheres and a quenching DNA strand [106]. The fluorescence of the microsphere was restored upon displacement of the quencher DNA strand by the DNA target in a sequence-dependent manner. Better mismatch discrimination was observed with PNA probes over DNA probes. In another example, a TAMRA-labeled PNA probe was used in combination with a short Cy5-labeled DNA oligonucleotide to form a FRET pair for the homogeneous assay of a SNP. Strand displacement with the unlabeled DNA target resulted in an increased TAMRA and decreased Cy5 fluorescence. Room temperature discrimination of a single base mismatch from the complementary DNA target was possible with a detection limit of 10 nM [107].

The use of pyrrolidinyl PNA probes gave an even better mismatch discrimination than conventional PNA probes

Table 2: Examples of nucleic-acid-templated fluorogenic reactions of PNA probes leading to non-ligated products.

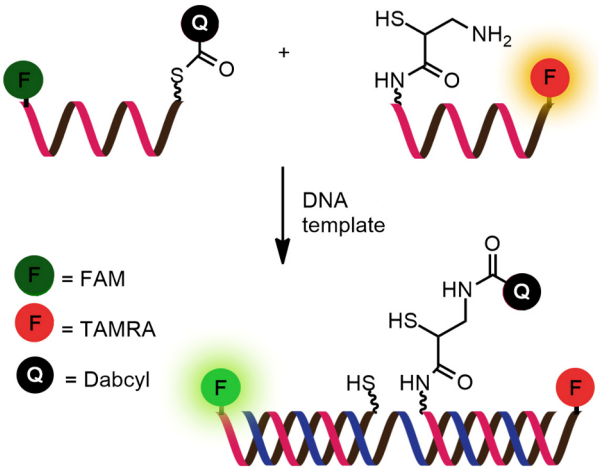
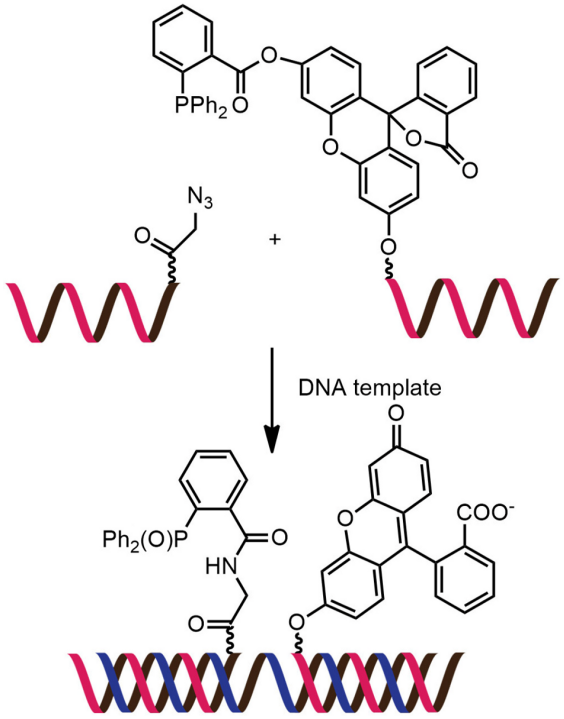
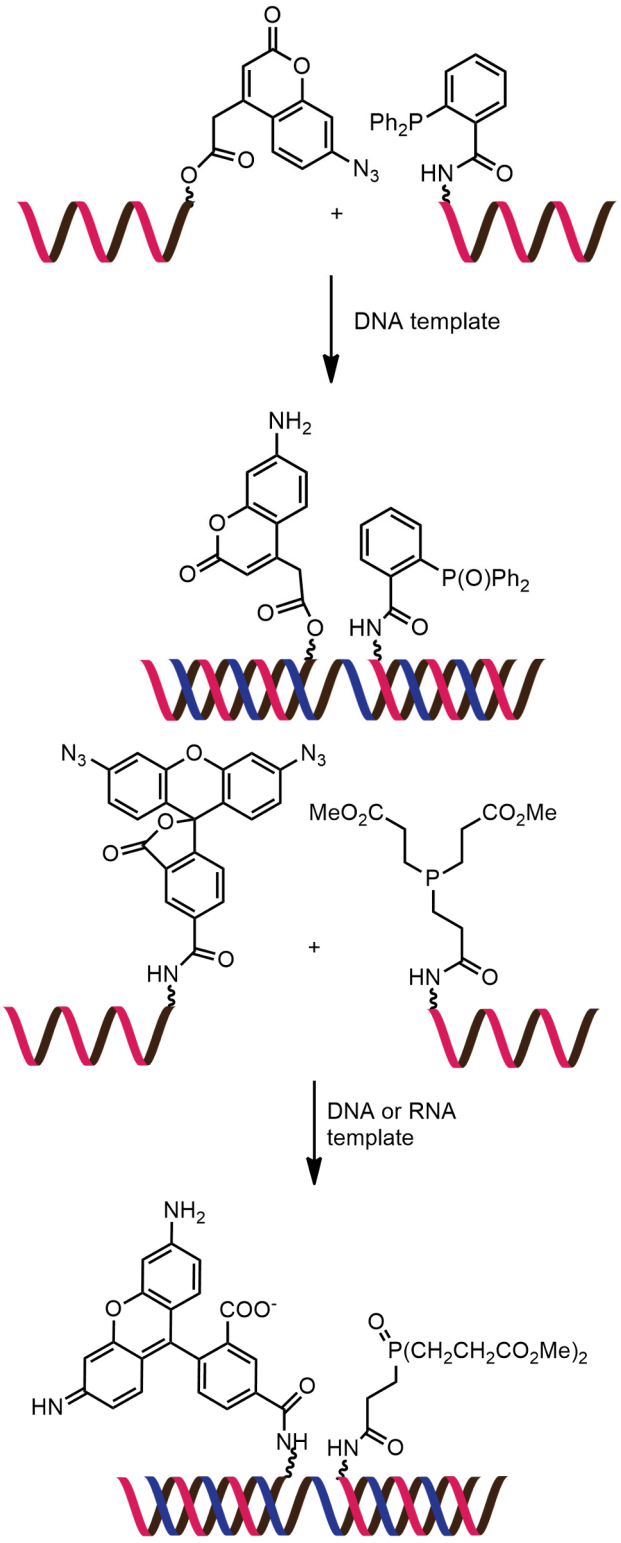
System	Remarks
	Reaction: thiol-mediated transfer of quencher Probe: aegPNA Template: oligodeoxynucleotide Rate acceleration: up to 1000 (relative to non-templated reaction) Selectivity: up to 138 (relative to single-mismatched template) TON: 8–38, depending on the type of linker LOD: 0.02 nM or less Applications: in vitro detection of oligodeoxynucleotides Ref: [103,104]
	Reaction: uncaging of fluorophore by Staudinger reaction Probe: aegPNA Template: oligodeoxynucleotide Rate acceleration: ca. 188 (relative to non-templated reaction) Selectivity: 31–37 (relative to single-mismatched template) TON: N/A LOD: N/A (all experiments were conducted in low μM range) Applications: in vitro detection of oligodeoxynucleotides Ref: [98]

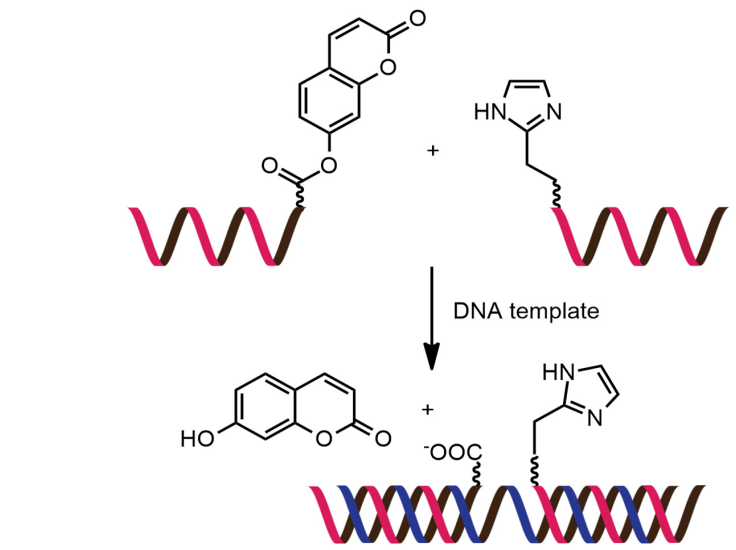
Table 2: Examples of nucleic-acid-templated fluorogenic reactions of PNA probes leading to non-ligated products. (continued)



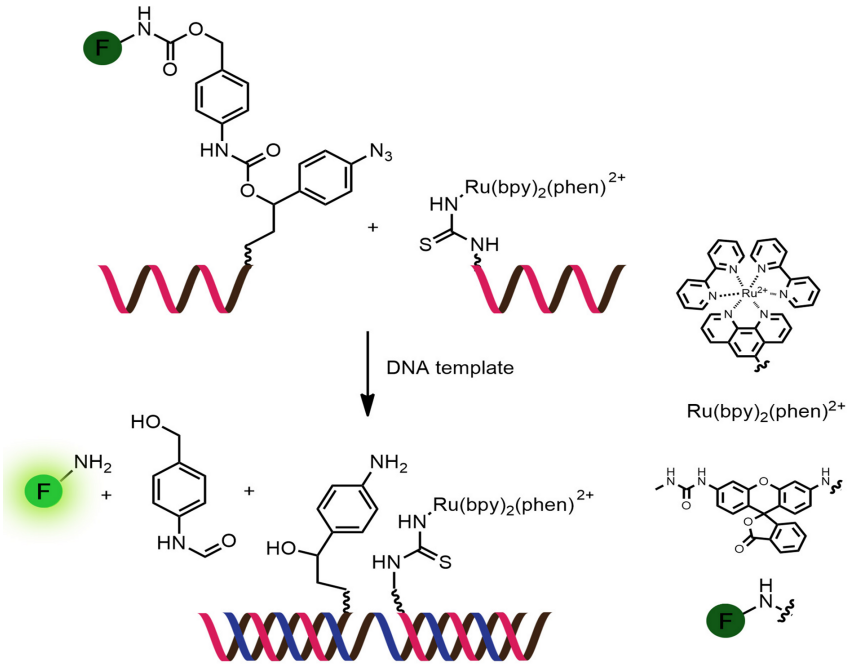
Reaction: uncaging of fluorophore by Staudinger reaction
Probe: aegPNA
Template: oligodeoxynucleotide
Rate acceleration: N/A
Selectivity: >10 (relative to single-mismatched template)
TON: >100
LOD: low nM
Applications: in vitro detection of oligodeoxynucleotides
Ref: [99]

Reaction: uncaging of fluorophore by Staudinger reaction
Probe: γGPNA
Template: oligodeoxynucleotide and oligoribonucleotide
Rate acceleration: N/A
Selectivity: >10 (relative to single-mismatched template)
TON: >10
LOD: low nM
Applications: imaging of 23S RNA in fixed *E. coli* cells
Ref: [101]

Table 2: Examples of nucleic-acid-templated fluorogenic reactions of PNA probes leading to non-ligated products. (continued)



Reaction: unmasking of fluorophores by imidazole-catalyzed hydrolysis
Probe: aegPNA
Template: oligodeoxynucleotide
Rate acceleration: 5.5×10^5 (relative to non-templated reaction); 11 (relative to DNA-based reaction)
Selectivity: 23–30 (relative to single-mismatched template)
TON: N/A
LOD: N/A (all experiments were conducted in low μM range)
Applications: in vitro detection of oligodeoxynucleotides
Ref: [102]



Reaction: photocatalytic release of fluorophore by reduction of self-immolative azide linker
Probe: γ-Ser PNA
Template: oligodeoxynucleotide, miRNA
Rate acceleration: first order rate constant (k) 0.64×10^{-3} s^{-1}; could not be measured for non-templated reaction
Selectivity: 8–10 (relative to single-mismatched template)
TON: >4000
LOD: 5 pM
Applications: in vitro and intracellular visualization of miRNA
Ref: [105]

[108,109]. The slow kinetics of the strand displacement involving highly stable PNA–DNA hybrids can be a major limitation of all the PNA-based strand displacement probes. Strand exchange can be facilitated by heating or increasing the salt concentration, where Mg^{2+} is more effective than Na^+ [107]. On the other hand, a polycationic comb-type dextran-poly(lysine) copolymer was shown to strongly promote the strand exchange of PNA–DNA duplexes by another DNA strand [110]. The equilibrium of the strand displacement reaction lies on the side of the more stable duplex, and so the original strand displacement probes are designed to have lower stabilities than the final duplexes. This can generally be achieved by employing a short complementary strand or by introducing mismatch pairs into the probe [109], but a C·I pseudocomplementary base pair, which is somewhat less stable than the C·G pair, has also been successfully employed in one case [111]. A PNA-based strand displacement probe was delivered into mammalian cells by employing cationic shell

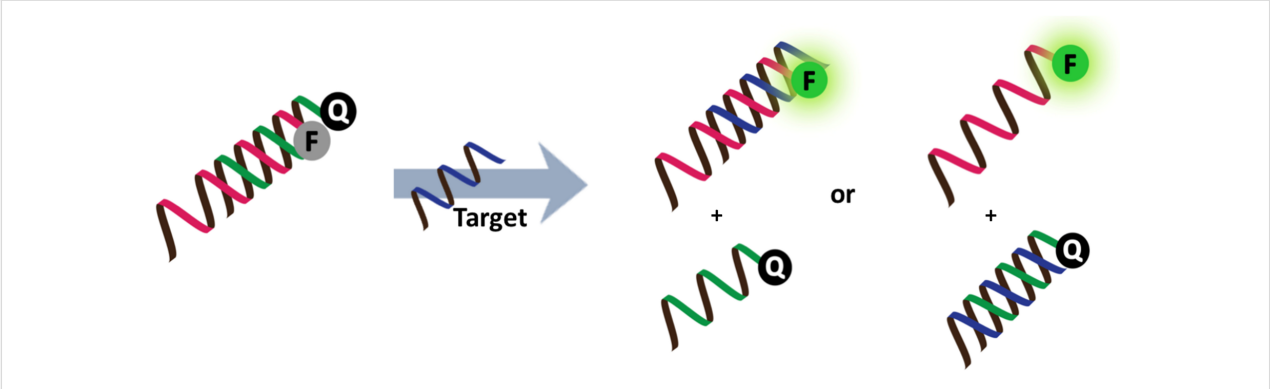


Figure 14: The working principle of strand displacement probes.

cross-linked nanoparticles, whereupon its application in the imaging of cellular mRNA expression was demonstrated [112]. Related to the concept of strand displacement probes, a self-reporting PNA–DNA primer was designed, whereby a fluorophore-labeled DNA probe formed a duplex with a short quencher-labeled PNA strand. The sticky end of the DNA part in the chimeric probe acted as a primer in a PCR reaction, which, upon chain extension displaced the quencher PNA and resulted in an increased fluorescence after several rounds of PCR [113].

Combination of single-labeled PNA probes with cationic conjugated polymers

The absence of negative charges on the PNA backbone offers a unique advantage for the development of novel DNA assays that cannot be made with oligonucleotides or analogues. One notable example is the use of a labeled PNA probe in combination with water-soluble cationic conjugated polyelectrolytes (CCP) for the FRET-based detection of DNA [114]. The CCP is typically an extended π-conjugated system, such as oligophenylene/fluorene with appending quaternary ammonium side chains (Figure 15A). Being electrostatically neutral, PNA cannot bind to these positively charged CCP unless it is first hybridized with DNA to form a negatively charged PNA–DNA complex. The CCP acts as a light harvesting antenna that transfers the energy to another label on the PNA probe and, therefore, the FRET is observed only when the labeled PNA probe, its complementary DNA and the CCP are present together (Figure 15B) [115,116]. It was demonstrated that the fluorescence signal obtained via FRET from CCP to fluorescein-labeled PNA was of an order of magnitude higher than the direct excitation of the fluorescein-labeled PNA [117]. Accordingly, the light harvesting properties of the oligomeric CCP improved the detection efficiency by signal amplification.

Although the high specificity of PNA probes can already differentiate between complementary and mismatched DNA targets on its own, the specificity for SNP detection could be further improved by the use of S1 nuclease, a single stranded specific nuclease that can digest single stranded DNA or mismatched duplexes more rapidly than perfect complementary duplexes [115]. The S1 digestion also improved the signal-to-background ratio by preventing the binding of the CCP on the DNA strand in the region remote from the PNA binding site, which is quite problematic in the detection of long DNA targets derived from PCR. On the other hand, addition of an organic solvent, such as *N*-methylpyrrolidinone [118], or a surfactant [119,120] improved the performance of the system by reducing the non-specific aggregation of the hydrophobic CPP and by reducing its non-specific binding with ssDNA. Up to a 10-fold improvement in the detection sensitivity in the presence of sodium dodecylsulfate (SDS) was reported [121]. Detection of dsDNA targets via the formation of high order complexes between PNA–dsDNA was also possible using a combination of PNA probes and CPP [122]. The use of a labeled PNA probe can be avoided by employing a fluorescence dye that can bind to PNA–DNA duplexes or triplexes, such as thiazole orange (TO) [123]. A polycationic dendritic fluorophore has also been used as a FRET donor to fluorescein-labeled pyrrolidinyl PNA probes giving a highly sequence-specific DNA detection at room temperature without requiring S1 digestion at a subnanomolar detection limit [124].

A new CCP called PFBT that can be excited at 488 nm (the wavelength commonly found in commercial microarray readers) has been used in combination with Cy5-labeled PNA probes immobilized on glass slides for the FRET-based detection of DNA in microarray formats [125]. Approximately 10^{10} copies (ca. 20 fmol) of unlabeled DNA can be readily detected following standard surface-hybridization protocols. The use of the solid support assay format also allows detection of DNA directly from the fluorescence signal of the CPP bound to the solid support without requiring labeling on the PNA probe [126]. Polystyrene microbeads self-assembled on

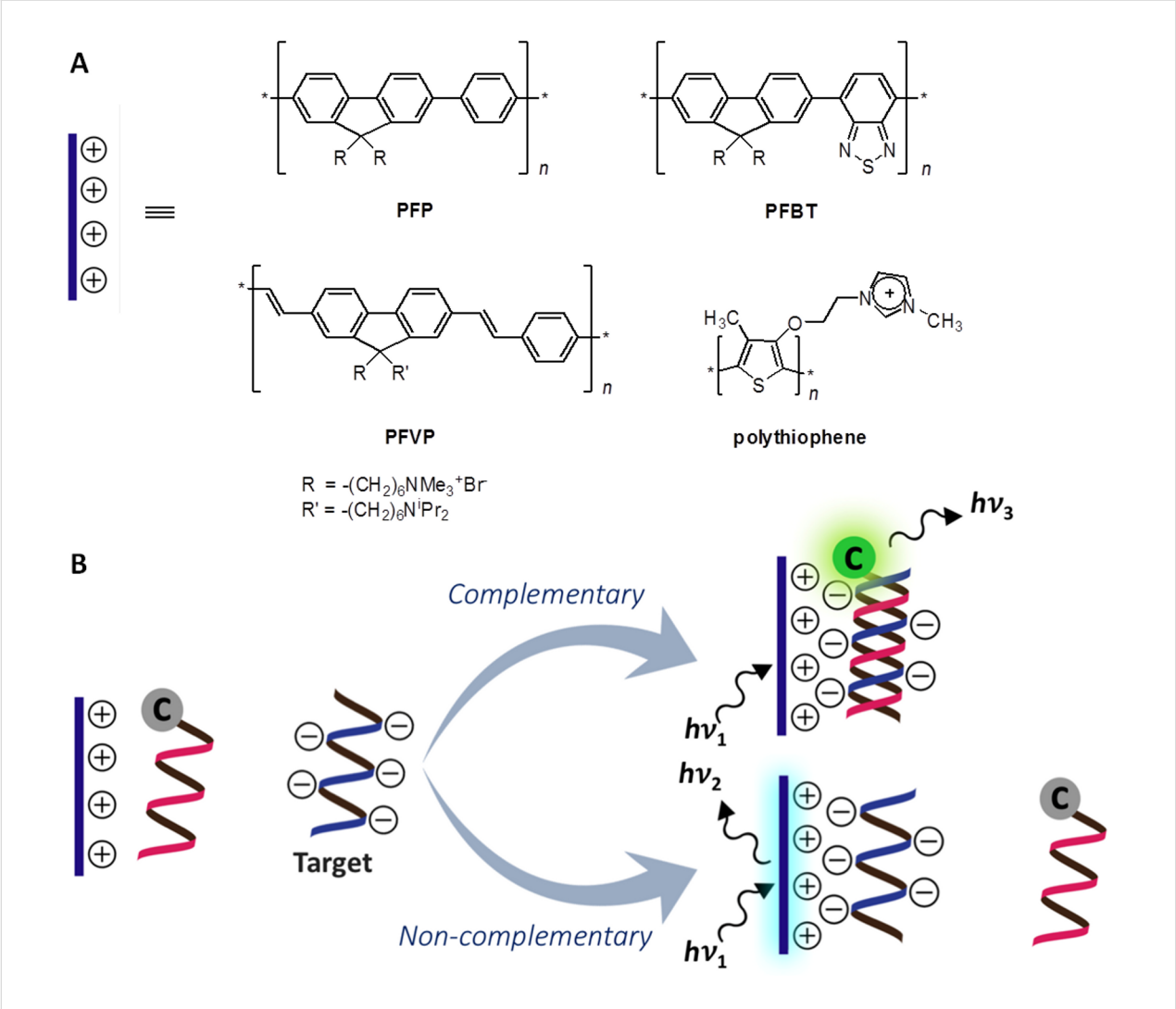


Figure 15: (A) Examples of CPP successfully used with labeled PNA probes. (B) The use of single-labeled PNA probes in combination with cationic conjugated polymers for FRET-based DNA detection.

patterned silicon chips have been used as the solid support for the PNA–CPP based DNA assay. Using this assay format, a combination of long-wavelength emissive CPP and PNA allowed detection of as low as 150 attomol of unlabeled DNA, or 300 copies (0.5 zeptomol) of Cy5-labeled DNA targets by confocal microscopy [127]. A fluorescence cationic polythiophene that fluoresces upon binding to DNA [128] has been used as a transducer for the highly specific detection of DNA, captured by an unlabeled PNA probe immobilized on glass slides, giving a subpicomole sensitivity that could conceivably be improved by the use of labeled PNA probes [129].

Combination of PNA probes with nanomaterials as external quenchers/FRET partners

Graphene oxide (GO) has been used extensively in combination with dye-labeled DNA for fluorescence-based DNA/RNA sensing [130,131]. Although ssDNA interacts strongly with GO via a combination of multiple interactions, it is mainly governed by hydrophobic interactions and hydrogen bonding between nucleobases and the GO surface [132]. Duplex DNA interacts less strongly with GO since the pairing nucleobases are buried inside the duplex in dsDNA. This phenomenon, together with the ability of GO to act as an effective fluorescence quencher, leads to the development of a novel fluorescence DNA sensor platform based on fluorescence labeled oligonucleotide probes and GO (Figure 16) [133,134]. Although the applications of DNA-based fluorescent probes have been well documented, the non-specific nature of the interaction between ssDNA and GO can result in competitive displacement of the adsorbed DNA probes by non-target ssDNA [135-137]. The degree of such non-specific displacement would depend strongly on the base sequence and, therefore, false positive results can be expected.

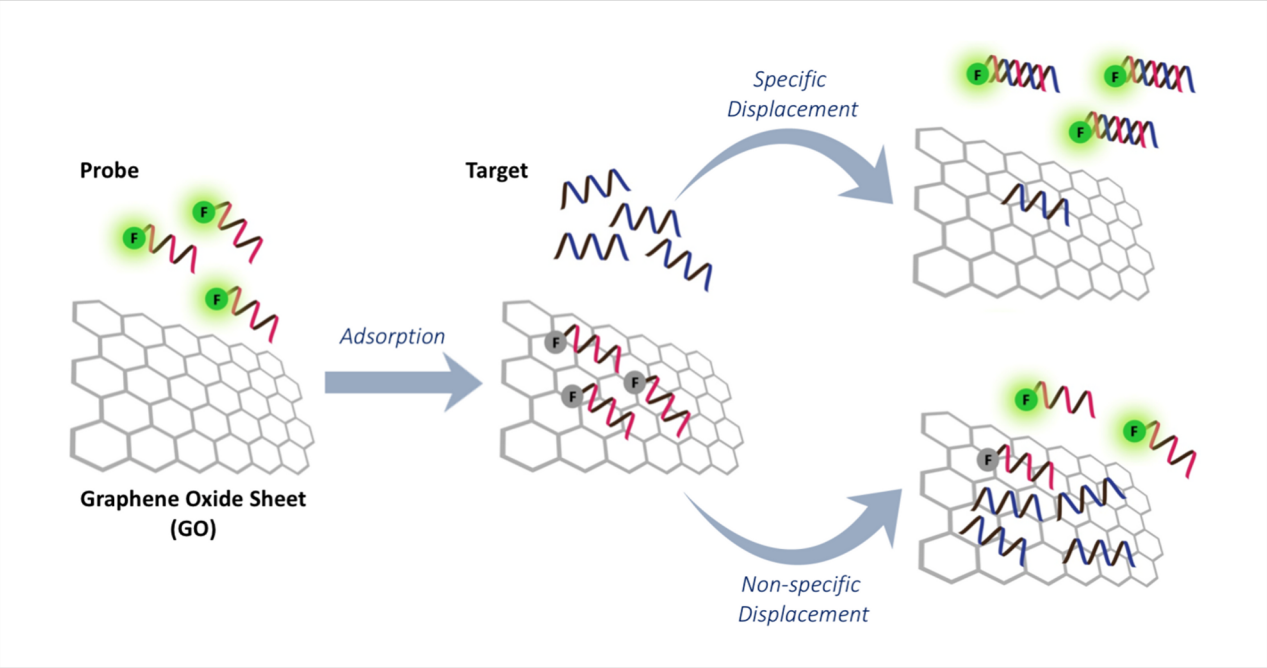


Figure 16: The concept of PNA–GO platform for DNA/RNA sensing.

The use of PNA as a probe offers a solution to overcome this issue since the hydrophobic backbone of the PNA probe appears to enhance the interaction with GO further, making the non-specific displacement of PNA by non-target DNA unfavorable [138]. The stronger interaction also means that fluorescent labeled PNA probes are more strongly quenched by GO than DNA probes with an identical sequence, leading to a lower background fluorescence. Importantly, the adsorbed PNA probes are still available for the formation of PNA–DNA duplexes, which are less strongly adsorbed by GO and so this leads to a release of the hybridized PNA probe from the GO and an enhanced fluorescence. The drawback of the use of the strongly adsorbed PNA probes is the rather slow kinetics and therefore an elevated temperature is often required to promote the desorption and hybridization of the PNA probes. Detection limits in the nano- to picomolar range have been achieved, while the high specificity of PNA allowed differentiation between complementary and non-complementary DNA, including single mismatched DNA as well as RNA targets, although temperature shifting is often necessary to obtain perfect discrimination [139,140].

The ability of PNA to invade into the DNA duplex allows the use of a PNA–GO sensing platform for the direct detection of dsDNA targets without requiring denaturation. Using three different labeled PNA probes, multiplex detection of short synthetic dsDNA that corresponded to the DNA sequences of HVA (hepatitis A Val17 polyprotein), HIV and HVB (hepatitis B virus surface antigen) in a single tube was demonstrated [141]. Addition of a very low concentration of bovine serum albumin (BSA, 0.01%) was shown to enhance the performance of the PNA–GO-based DNA sensing platform by reducing the non-specific adsorption of PNA–DNA duplexes to GO, which resulted in a lower limit of detection (LOD) by almost two orders of magnitude [142]. The detection limit could be improved further for the detection of microRNA by a sequence specific RNA-templated DNA ligation to form a circular DNA. This circular DNA was then used as a template for rolling circle amplification reaction (RCA) initiated by phi29 DNA polymerase to give a long tandem repeat DNA. As low as 0.4 pM of miRNA could be readily detected with single mismatch specificity [143].

In addition to the direct detection of DNA and RNA targets, the PNA–GO platform has also been used for an in vitro assay of RNA polymerase activities and its inhibition by detection of the RNA formed [144]. The cell-penetrating ability of nanosized GO (NGO) allowed cellular internalization of the PNA–NGO complexes, enabling direct, quantitative and multiplex monitoring of various micro (mi)RNAs in living cells [145]. Side-by-side comparison showed clear advantages of PNA over DNA probes, where the DNA–NGO complexes lighted up non-specifically in the presence of the cell lysate, presumably by the interference of cellular matrices with the DNA–NGO interactions or by cleavage of the DNA probes by cellular nucleases. Related to this, hyarulonic acid-coated GO was also employed as a carrier for introducing PNA into cancer cells for the simultaneous detection and inhibition of endogeneous miRNA-21 in

CD44-positive MBA-MB231 cells, leading to a decreased proliferation and inducing apoptosis [146].

In addition to GO, several other nanomaterials have been used as external quenchers in combination with labeled PNA probes for the fluorescent detection of DNA or RNA. WS_2 nanosheets can selectively adsorb ssPNA and quench its fluorescence [147]. The addition of complementary DNA targets resulted in the restoration of the fluorescence in pretty much the same way as GO. Likewise, carbon nitride nanosheets [148] and a zirconium-based nano metal-organic framework (UiO-66) [149] behaved similarly. In addition, they have been successfully used in combination with PNA for monitoring of miRNA inside living cells [148,149].

Quantum dots (QDs) are another class of nanomaterial that offer great promise as a FRET partner for fluorescence detection of nucleic acids [150-153]. Despite the fact that several examples of QDs and oligonucleotide probes for DNA/RNA detection are known [154,155], the use of QDs in combination with labeled PNA probes was only recently reported for a sandwich-type fluorescence detection of DNA at nanomolar concentrations [156].

Single-labeled fluorogenic PNA probes

Fluorogenic linear PNA probes carrying a single label (Figure 17) are highly attractive in terms of their ease of design and synthesis. To allow the wash-free, homogeneous detection of target nucleic acids, it is necessary to incorporate an environment sensitive label onto the PNA probe. Ideally, the label should interact differently and yield different responses to ssDNA, ssPNA and PNA–DNA or PNA–RNA duplexes by various mechanisms, such as groove binding or intercalation. This fluorescent label may be a stand-alone entity (referred to as a tethered label) or combined with a nucleobase that integrates the base pairing and the fluorescence sensing in a single event. The latter strategy should allow a more precise control of the position and orientation of the label than the former, and therefore should be more sensitive to local environment changes than the former, which indirectly and non-specifically senses the global formation of duplex or base stack.

PNA probes with tethered labels

One of the earliest PNA probes in this category are the so-called "light-up probes", which consist of a short single stranded PNA probe (ssPNA) linked to the DNA binding dye TO (Figure 18) at one end of the molecule, usually at the N-terminus, via a flexible linker [157]. The TO binds non-specifically with DNA duplexes and shows a large fluorescence enhancement upon binding [158]. PNA–DNA hybridization facilitates the binding of TO to the PNA–DNA duplexes, resulting in a fluorescence enhancement due to the increased co-planarity and restricted motion of the two conjugated aromatic rings in the TO (Figure 19). The electrostatically neutral backbone of PNA was originally thought to offer a unique advantage over DNA and other negatively charged oligonucleotide analogues, since it would minimize self-binding between TO and the probe itself. However, subsequent studies have revealed that there are still significant fractions of TO that are back-bound to the PNA probe (molar ratio between 0.7 to almost 1.0 at 30 °C) [159]. The smallest fraction was observed in probes with homopyrimidine sequences. In addition, the fluorescence quantum yields of the probe in the back-bound conformation were lowest in pyrimidine rich sequences. Therefore, the performance of light-up probes is rather sequence-dependent, although it could be marginally improved by increasing the temperature.

Rapid and specific detection of PCR products by light-up PNA probes in a simple mix-and-read fluorescence assay [160] or by real-time PCR [161,162] have been demonstrated in the context

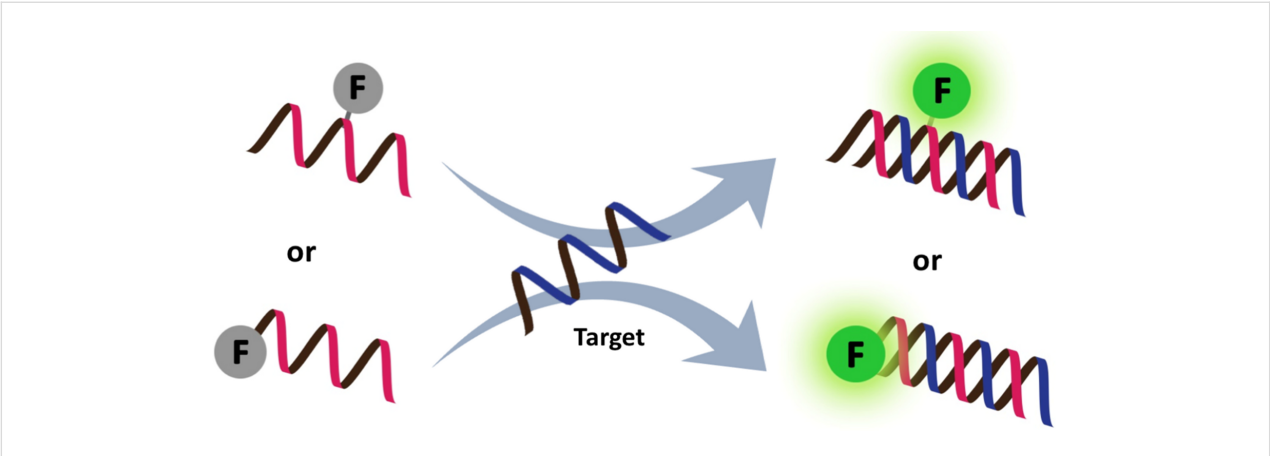


Figure 17: Single-labeled fluorogenic PNA probes.

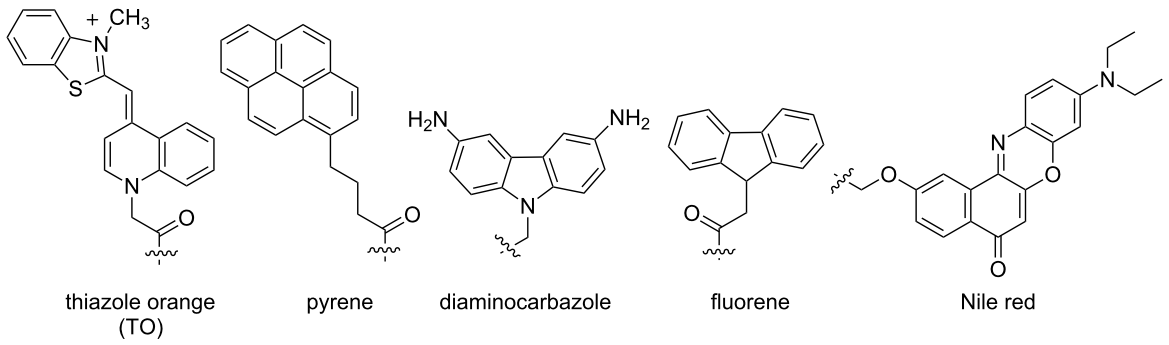


Figure 18: Examples of environment sensitive fluorescent labels that have been incorporated into PNA probes as a tethered label.

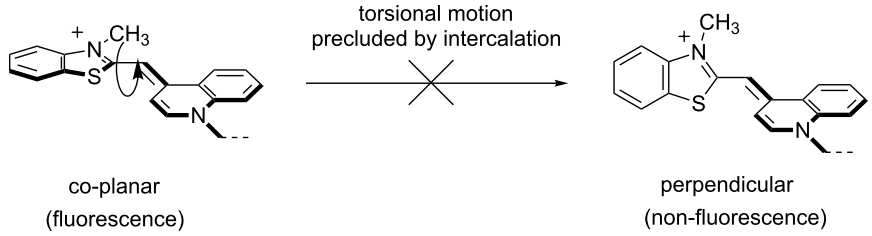


Figure 19: The mechanism of fluorescence change in TO dye.

of pyrimidine rich sequences. A pair of pseudo-complementary PNA probes terminally labeled with TO allowed the detection of specific DNA sequences in long dsDNA and plasmids without requiring prior denaturation [163]. Although non-selective binding of the TO label to DNA or RNA duplexes were occasionally observed, this could be suppressed by using a shorter linker (as in FIT PNA probes, vide infra) or by combination with a second dye, such as pyrene, that can interact with TO by π-stacking. Such a combination in a short PNA probe has been used as a non-covalent affinity label for monitoring small interfering (si)RNA delivery [164,165]. In addition, TO has also been linked as a tethered label at internal positions of γPNA [166] and acpcPNA [167], where both showed dramatic light-up behaviors (50–100 fold) when hybridized to complementary DNA.

In addition to TO, other environment-sensitive labels, such as fluorene [168], pyrene [167,169,170], diaminocarbazole [171] and Nile red [172], have been successfully employed in the development of single-labeled fluorogenic aegPNA, γPNA and acpcPNA probes (Figure 18). Up to a 73-fold fluorescence enhancement was observed in the case of a single pyrene-labeled acpcPNA probe [170]. This is in sharp contrast to pyrene-tethered DNA probes, which are usually quenched after hybridization due to intercalation of the pyrene in the DNA–DNA duplex [173,174], unless the pyrene is attached to the base via a rigid linker that disfavors such intercalation [175,176].

PNA probes carrying fluorescent nucleobases

There are three major strategies to introduce fluorescent nucleobases into oligonucleotides or PNA probes. These are (i) the extension of conjugation in the natural nucleobase, (ii) appending a fluorophore onto a natural nucleobase and (iii) the use of unnatural, intrinsically fluorescent nucleobases. They can be divided into fluorescent nucleobases capable of hydrogen bond formation, which can form specific base pairs with canonical nucleobases, and those that cannot (i.e., universal bases). In this section, only those that have been used in combination with PNA will be discussed. More general reviews of fluorescent base analogues in the DNA/RNA context can be found elsewhere [177,178].

Fluorescent nucleobases capable of hydrogen bonding ("base discriminating fluorophores")

The structures of fluorescent nucleobases with hydrogen-bonding abilities that have been studied in the PNA context are shown in Figure 20. The earliest examples of PNA carrying an intrinsically fluorescent nucleobase 2-aminopurine (2-AP) were

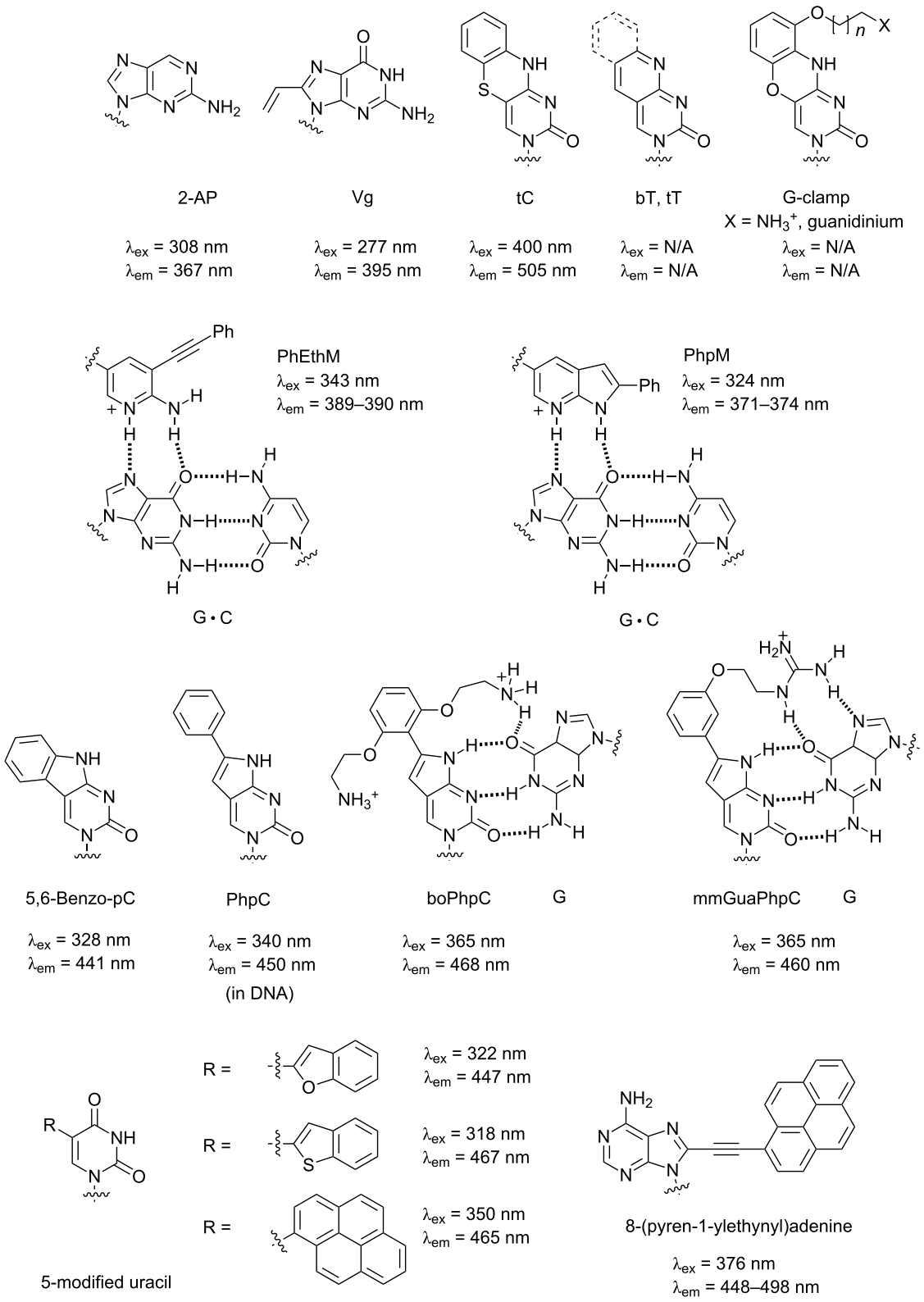


Figure 20: Fluorescent nucleobases capable of hydrogen bonding that have been incorporated into PNA probes.

reported since 1997 [179]. The 2-AP in aegPNA selectively recognizes dT in the DNA strand with comparable affinity to that for A, and the base pairing resulted in fluorescence quenching. While the quenching of 2-AP can be useful for probing the structure and dynamics of the PNA and its DNA/RNA hybrids, the quenching effect is small and not highly specific as quenching may also be partly observed in the case of mis-pairing [180]. In another example, the fluorescence 8-vinylguanine (Vg) in aegPNA forms specific base pairs with C in DNA and RNA. In addition, Vg can still participate in G-quadruplex formation similar to dG. The formation of base pairs or G-quadruplexes was accompanied by fluorescence quenching, and Vg-modified aegPNA has been used to probe the interaction between RNA quadruplexes and PNA [181].

Another example of a nucleobase with extended conjugation that has been incorporated into PNA is 1,3-diaza-2-oxophenothiazine (tC), which is a tricyclic analogue of cytosine [182]. The tC exhibits decent quantum yields of ≈0.2 both in free form and when incorporated into DNA or aegPNA [183], and forms specific base pairs with dG, but does not show appreciable fluorescence change upon the base pairing. While this could be advantageous for several applications, it means that tC–PNA alone is not useful as a fluorogenic probe. The related tri- or bicyclic thymine analogues tT and bT (without the additional aromatic ring) pair specifically with dA when incorporated into aegPNA, but their fluorescence properties have not been reported [184]. PNA carrying phenoxazine analogues of tC (tC^O) with a positively-charged pendant (so-called "G-clamp") were prepared in order to improve the affinity of PNA–DNA duplex by additional hydrogen bonding [185-187]. No fluorescence properties of these tC^O-based G-clamps PNA have been reported, but it is quite likely that they will be non-responsive to the base pairing similar to PNA carrying tC or tC^O [177].

Pyrrolocytosine is another intrinsically fluorescent hydrogen-bond-forming nucleobase that has been extensively studied in a DNA context [188,189]. When incorporated into PNA, the simple phenyl-substituted pyrrolocytosine (PhpC) recognizes dG in DNA and G in RNA with a slightly increased and decreased affinity, respectively, relative to C [190]. Addition of a positively charged pendant group at the *ortho*-position of the phenyl substituent, such as in boPhpC, substantially increased the binding affinity as well as the specificity due to the additional hydrogen-bonding interactions with the pairing G residue analogous to G-clamp phenoxazine. Moving the substituent to the *meta*-position, as in mmGuaPhpC, increased the binding affinity towards RNA while still maintaining good DNA binding [191]. When incorporated into PNA, these PhpC derivatives exhibited a bright blue fluorescence with large fluorescence quantum yield (0.5–0.6), which can be useful for monitoring the cellular uptake and distribution of PNA [192]. In addition, the fluorescence was responsive to the base pairing, where up to 60% quenching was observed upon duplex formation with DNA or RNA. The fused ring fluorescence cytosine analogue 5,6-benzo-pC gave a large Stokes shift (113 nm) and good quantum yield (0.79) as a monomer. Unfortunately, severe fluorescence quenching was observed upon incorporation of this monomer into PNA sequences and no discrimination was observed among complementary and mismatched DNA targets [193]. PNA probes carrying PhpM, a deoxy analogue of PhpC, and its open-chain analogue PhEthM were designed as a fluorogenic probe for the detection of RNA duplexes. Isothermal calorimetric titration suggested that PhpM was less selective than PhEthM in recognizing dsRNA, and that the binding was pH dependent. A lower pH was required for protonation of the nitrogen atom to allow binding with G·C base pairs in a Hoogsteen fashion. Moderate fluorescence quenching (50–60%) was observed upon triplex formation [194].

Fluorophore-modified uracils have also been extensively studied as fluorescent nucleobases in the context of aegPNA. Modifications have invariably been made at the 5-position of uracil, which could be functionalized by various aromatics or alkynes via palladium-catalyzed cross couplings of the corresponding iodouridine derivative. 5-Benzothiophene- and 5-benzofuran-modified uracil in aegPNA exhibited a fluorescence that was marginally sensitive to the environment [195]. When incorporated into PNA, benzothiophene-uracil exhibited an increased fluorescence compared to the free nucleoside. The opposite effect was observed with benzofuran-uracil, and the fluorescence was almost completely quenched when G was the flanking nucleobase. Significant fluorescence enhancement with a small blue-shift of the emission maxima was observed upon duplex formation with DNA for both modified uracil derivatives. Unfortunately, the discrimination between complementary and mismatched duplex observed with benzothiophene-uracil was limited to sequences with flanking Cs. No discrimination was observed with benzofuran-uracil PNA unless it was used in combination with GO as a quencher for ssPNA [196]. In this respect, the thiophene-modified uracil acted rather like a general fluorescence label.

On the other hand, the fluorescent nucleobase 5-(pyren-1-yl)uracil in acpcPNA formed a specific Watson–Crick type base pairing with dA in the DNA strand, and the duplex formation was accompanied by a strong (up to 42-fold) fluorescence emission increase at 465 nm [197]. This is in sharp contrast with the behavior of the same pyrene-modified uracil in DNA, where no discrimination was observed among the four canonical nucleobases in terms of both thermal stabilities and fluorescence responses [198,199].

A similar selective recognition of dT with light-up behavior was also observed with the fluorescent nucleobase 9-(pyrenylethynyl)adenine when incorporated in acpcPNA, but not in DNA [200]. In the case of DNA–DNA duplexes, the nucleobase most likely adopts a *syn* conformation, thereby placing the hydrophobic pyrene moiety in the base stack at the expense of hydrogen bonding. The stronger base-pairing in PNA–DNA duplexes probably make this process less favorable. This emphasizes the subtle different behavior of PNA and DNA that may make PNA useful in certain circumstances.

Fluorescent nucleobases incapable of hydrogen bonding

The FIT probe first reported in 1999 [201] is perhaps the most well-known representative PNA probe in this class [202,203]. It contains the DNA-staining dye TO attached to the backbone of aegPNA as a nucleobase surrogate (Figure 21). Being incapable of hydrogen binding, TO behaves as a universal nucleobase as it can pair equally well with all canonical nucleobases in the DNA strand [204]. However, unlike most universal bases that usually destabilize the duplex [205], the base pairing strength involving TO was comparable to that of A·T pairs. The fluorescence of the TO-based FIT probe was increased substantially in the presence of the complementary DNA target as a result of the restricted rotation of the TO chromophore upon its intercalation into the base stack (Figure 19). Considering the bulkiness of the TO dye, it is quite likely that the opposite nucleobase was forced away from the duplex and so did not directly participate in the base pairing. The responsiveness of TO is sequence-dependent, and the largest responses (>20-fold) were observed when there is at least one A residue adjacent to the TO base. In most cases the fluorescence increase is quite general for various sequence context [206], Nevertheless, a smaller fluorescence changes (less than 10-fold) were observed with other flanking nucleobases. Importantly, when a mismatched base pair is present adjacent to the TO, the fluorescence increase was much lower than the complementary duplex, and the discrimination could be improved further, albeit at the expense of sensitivity, by increasing the temperature.

Unlike light-up probes, whereby the TO was linked at the end of the PNA probe as a tethered label via a long and flexible linker, FIT probes have the TO linked to the PNA backbone as a base replacement via a short and rigid linker (Figure 21). The working principle of the FIT probe is, therefore, distinctly different from light-up probes and should in principle allow for better control since the TO is forced to intercalate into the duplex at a well-defined position and so can directly sense the mismatched base pairing adjacent to the intercalation site. In contrast, the tethered TO label in a light-up probe globally senses the duplex formation by intercalation of the dye into the PNA–DNA duplexes making it difficult to differentiate between the complementary and mismatched duplexes. The residual fluorescence of the single stranded FIT probe was also more easily predicted than that of light-up probes since it is the result of direct interaction between the TO and the nearest neighbor, and an adjacent G is to be avoided since it contributes to a large background signal [206].

Detailed studies of the fluorescence lifetime indicated several distinct fluorescence decay processes are associated with different hybridization states of FIT probes [207]. Systematic variation of the linker length and position of attachment of the TO label indicated that the best response and mismatch discrimination were observed with a short linker linked to the quinoline ring of the TO [208]. Interestingly, a larger fluorescence increase was observed in the mismatched than complementary duplex when the TO dye was linked to PNA via the benzothiazole ring [209]. Subsequent work revealed that the responsiveness of the probe can be further improved by the use of a D-ornithine-derived TO-modified PNA monomer, which increased the number of rotatable bonds [210].

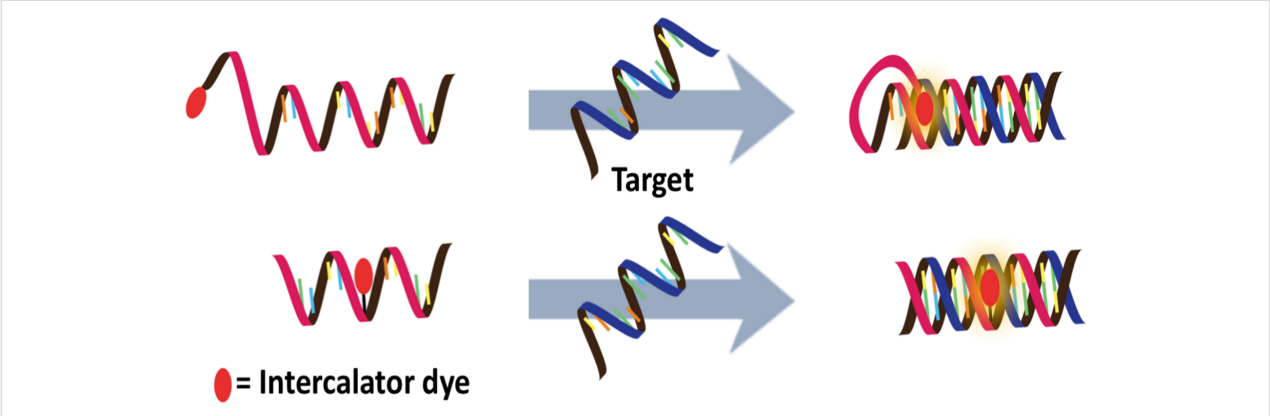


Figure 21: Comparison of the designs of the (A) light-up PNA probe and (B) FIT PNA probe.

Applications of FIT PNA probes for the real-time detection of DNA [210] and RNA [211] by qPCR have been demonstrated. Moreover, FIT PNA probes also recognize dsRNA by triplex formation, yielding an extremely large fluorescence enhancement (>200-fold) that is highly sensitive to base mismatches adjacent to the position of TO, similar to that with ssDNA and ssRNA targets [212,213]. With appropriate cellular delivery techniques, it has been possible to image specific RNA in living cells using FIT PNA probes [214-218]. In addition, FIT PNA probes were shown to offer fast hybridization kinetics, a higher signal-to-background ratio and a better specificity than classical DNA-based molecular beacons. Other related TO derivatives have been explored as alternative fluorophores for FIT PNA probes, but only few close analogs of TO show the selective fluorescence enhancement upon duplex formation that may allow their use in combination with TO-based FIT PNA probes for multiplex detection of DNA or RNA. These include the oxazole yellow (YO) [219], BO [220] and BisQ [221] dyes (Figure 22). Multiplex imaging of two RNAs in living cells was enabled by the combined use of BO- and TO-labeled FIT PNA probes [220]. The wash-free protocol and fast hybridization response of FIT PNA probes allows for both spatial and temporal monitoring of the mRNA expression in the cells [220].

The performance of FIT PNA probes can be further improved by combination of the TO with an additional dye to form a donor–acceptor pair (Figure 23). This dual-labeled FIT probe design offers advantages over classical stemless PNA beacons with other fluorophore/quencher pairs because the TO dye will also be quenched by the additional dye rather than relying on the rapid internal rotation alone. This results in a much lower background signal of such dual-labeled FIT PNA probes (more than 99.9% lower than single-labeled FIT probes) [222,223]. Upon hybridization, the TO will intercalate and act as a FRET

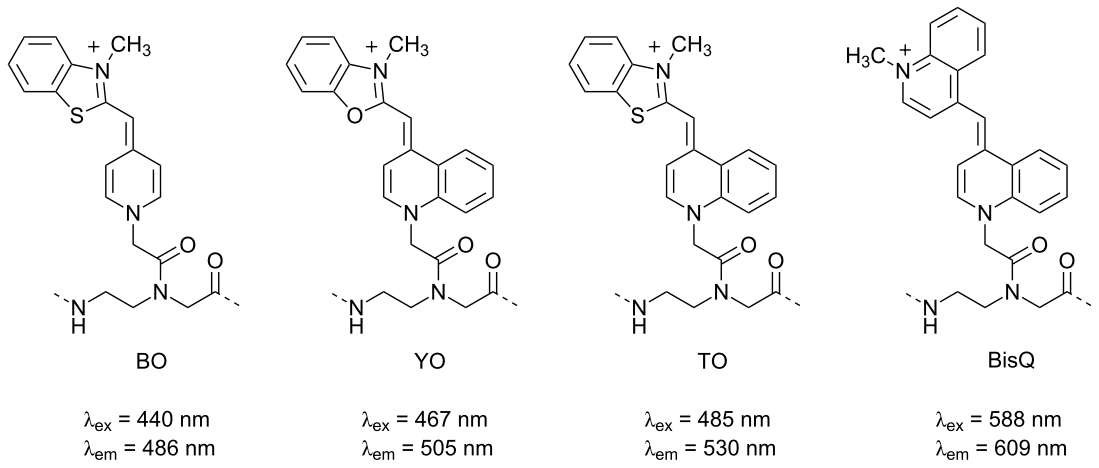


Figure 22: The structures of TO and its analogues that have successfully been used in FIT PNA probes.

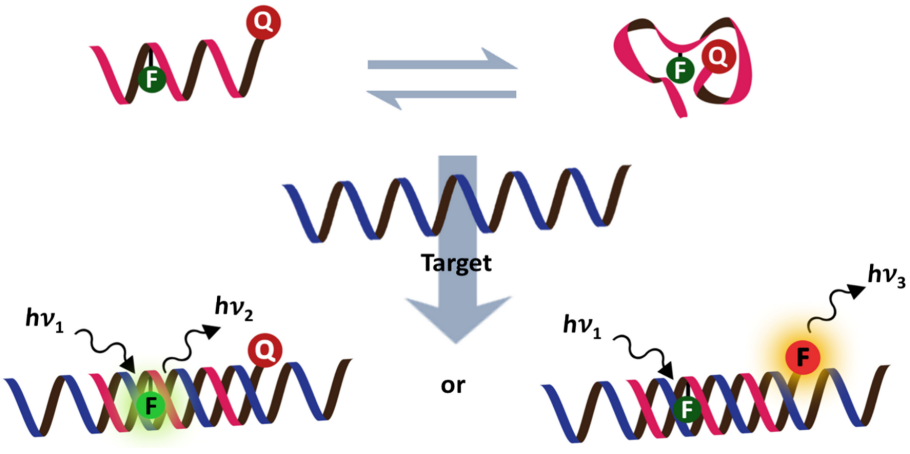


Figure 23: The working principle of dual-labeled FIT PNA probes [222,223].

donor provided that the second dye possesses a suitable spectral overlap. Up to a 450-fold enhanced fluorescence emission of the acceptor dyes was observed with TO/indotricarbocyanine (ITCC) FRET pairs (excited at TO and detected at ITCC). On the other hand, if the second dye did not possess a suitable spectral overlap, the dye would simply act as a quencher and the responsiveness of the TO dye was nevertheless still improved as a result of the greatly reduced background signal. Direct excitation of the acceptor dye also showed a markedly improved light-up signal compared to the single-labeled probe due to the mutual quenching of the dye by TO.

In addition to TO, simple polycyclic aromatic hydrocarbons, such as pyrene, have been explored as a non-hydrogen-bonding fluorescence universal base in aegPNA [224,225] and acpcPNA [226]. In the case of aegPNA with triazole-linked pyrene, the duplex stabilities were decreased by 10–12 °C regardless of the nature of the opposite base in the DNA or RNA strand. Hybrid formation resulted in quenching of the pyrene, and stronger quenching was observed with RNA over DNA [225]. The same triazole-linked pyrene in acpcPNA also behaved as a universal base with only slight destabilization of the duplexes. However, the fluorescence change of pyrene-labeled acpcPNA was in the opposite direction to that of aegPNA (light-up instead of quenching) [226]. This was explained by molecular dynamics simulations, which suggested that the triazole formed hydrogen bonds with the opposite nucleobase and pushed the pyrene towards the major groove instead of stacking within the PNA–DNA duplexes [226]. Other chromophores that can potentially behave as non-hydrogen-bonding fluorescent nucleobases in aegPNA include phenylazonaphthalene [227], flavin [228], naphthalimide [229] and psoralen [230]. Fluorescence studies have only been performed with psoralen, and revealed that the fluorescence was only marginally decreased when the PNA probe was hybridized with the complementary DNA target [230].

Conclusion

This review summarizes various strategies that have been successfully used for the design of fluorogenic PNA probes as well as their performances and applications when applicable. It can be seen that several designs originally developed for oligonucleotide probes can be applied to PNA. The tendency of the uncharged backbone of PNA to fold into a compact structure and the inability to interact with positively charged species that normally bind to nucleic acids offer unique opportunities to design new sensing platforms that have no equivalent in the case of DNA and related analogues. The superior properties of PNA over DNA probes in terms of their improved sensitivity, specificity and/or biological stability have been clearly demonstrated in several cases. Nevertheless, the availability and cost of PNA, together with other unfamiliar characteristics, such as its hydrophobicity, poor water solubility and cellular uptake, make the research community reluctant to adopt PNA as a new tool despite these potential advantages over more conventional oligodeoxynucleotide probes. While it is unlikely that PNA-based fluorogenic probes will generally replace DNA probes for routine applications, there are niche areas that PNA probes can potentially offer real advantages. These include applications that require very high specificity under non-stringent conditions, such as detection of SNPs or closely related nucleic acid targets, especially in multiplex fashion. In addition, PNA probes are especially suitable for targeting nucleic acid targets with secondary structures that may be difficult to access by other probes. The excellent biological stability of PNA-based fluorogenic probes make them particularly attractive as an alternative to oligonucleotide probes for intracellular nucleic acids detection. Although some impressive examples are emerging, most of these are proof-of-principle studies with highly expressed RNA targets. It remains to see if it is possible to increase the sensitivity to detect low abundant targets in situ and in real time. By combination of PNA as a recognition element together with brighter, more stable fluorophores, new signal transduction mechanisms and more sophisticated detection techniques, such as fluorescence lifetime measurement and single molecule fluorescence, there are still a lot of further opportunities for fluorogenic PNA probes to advance the field further.

Acknowledgements

The author acknowledges financial support from The Thailand Research Fund and Chulalongkorn University (DPG5780002), Mr. Kriangsak Faikhruea and Ms. Kotchakorn Supaboworn-sathit for the preparation of illustrations, and Dr. Robert Butcher (Chulalongkorn University) for the proof-reading.

ORCID® iDs

Tirayut Vilaivan - https://orcid.org/0000-0002-9184-9506

References

1. Heather, J. M.; Chain, B. *Genomics* **2016,** *107,* 1–8. doi:10.1016/j.ygeno.2015.11.003
2. Mascini, M.; Palchetti, I.; Tombelli, S. *Angew. Chem., Int. Ed.* **2012,** *51,* 1316–1332. doi:10.1002/anie.201006630
3. Tyagi, S.; Kramer, F. R. *Nat. Biotechnol.* **1996,** *14,* 303–308. doi:10.1038/nbt0396-303
4. Tyagi, S.; Bratu, D. P.; Kramer, F. R. *Nat. Biotechnol.* **1998,** *16,* 49–53. doi:10.1038/nbt0198-49
5. Nielsen, P. E. *Acc. Chem. Res.* **1999,** *32,* 624–630. doi:10.1021/ar980010t
6. Briones, C.; Moreno, M. *Anal. Bioanal. Chem.* **2012,** *402,* 3071–3089. doi:10.1007/s00216-012-5742-z
7. Karkare, S.; Bhatnagar, D. *Appl. Microbiol. Biotechnol.* **2006,** *71,* 575–586. doi:10.1007/s00253-006-0434-2

8. Tan, W.; Wang, K.; Drake, T. J. *Curr. Opin. Chem. Biol.* **2004,** *8,* 547–553. doi:10.1016/j.cbpa.2004.08.010
9. Tan, L.; Li, Y.; Drake, T. J.; Moroz, L.; Wang, K.; Li, J.; Munteanu, A.; Yang, C. J.; Martinez, K.; Tan, W. *Analyst* **2005,** *130,* 1002–1005. doi:10.1039/b500308n
10. Ranasinghe, R. T.; Brown, T. *Chem. Commun.* **2005,** 5487–5502. doi:10.1039/b509522k
11. Marras, S. A. E. *Mol. Biotechnol.* **2008,** *38,* 247–255. doi:10.1007/s12033-007-9012-9
12. Wang, K.; Tang, Z.; Yang, C. J.; Kim, Y.; Fang, X.; Li, W.; Wu, Y.; Medley, C. D.; Cao, Z.; Li, J.; Colon, P.; Lin, H.; Tan, W. *Angew. Chem., Int. Ed.* **2009,** *48,* 856–870. doi:10.1002/anie.200800370
13. Juskowiak, B. *Anal. Bioanal. Chem.* **2011,** *399,* 3157–3176. doi:10.1007/s00216-010-4304-5
14. Huang, K.; Marti, A. A. *Anal. Bioanal. Chem.* **2012,** *402,* 3091–3102. doi:10.1007/s00216-011-5570-6
15. Guo, J.; Ju, J.; Turro, N. J. *Anal. Bioanal. Chem.* **2012,** *402,* 3115–3125. doi:10.1007/s00216-011-5526-x
16. Giannetti, A.; Tombelli, S.; Baldini, F. *Anal. Bioanal. Chem.* **2013,** *405,* 6181–6196. doi:10.1007/s00216-013-7086-8
17. Zheng, J.; Yang, R.; Shi, M.; Wu, C.; Fang, X.; Li, Y.; Li, J.; Tan, W. *Chem. Soc. Rev.* **2015,** *44,* 3036–3055. doi:10.1039/C5CS00020C
18. Wang, Q.; Chen, L.; Long, Y.; Tian, H.; Wu, J. *Theranostics* **2013,** *3,* 395–408. doi:10.7150/thno.5935
19. Demidov, V. V. *Trends Biotechnol.* **2003,** *21,* 4–7. doi:10.1016/S0167-7799(02)00008-2
20. De Koning, H.; Pandit, U. K. *Recl. Trav. Chim. Pays-Bas* **1971,** *90,* 1069–1080. doi:10.1002/recl.19710901003
21. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991,** *254,* 1497–1500. doi:10.1126/science.1962210
22. Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. *Nature* **1993,** *365,* 566–568. doi:10.1038/365566a0
23. Tomac, S.; Sarkar, M.; Ratilainen, T.; Wittung, P.; Nielsen, P. E.; Nordén, B.; Gräslund, A. *J. Am. Chem. Soc.* **1996,** *118,* 5544–5552. doi:10.1021/ja960495l
24. Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskil, M. D.; Egholm, M.; Buchard, O.; Sönnichsen, S. H.; Nielsen, P. E. *Biochem. Pharmacol.* **1994,** *48,* 1310–1313. doi:10.1016/0006-2952(94)90171-6
25. Pellestor, F.; Paulasova, P. *Eur. J. Hum. Genet.* **2004,** *12,* 694–700. doi:10.1038/sj.ejhg.5201226
26. Ørum, H. *Curr. Issues Mol. Biol.* **2000,** *2,* 27–30.
27. Demidov, V. V.; Frank-Kamenetskii, M. D. PNA Openers and Their Applications. In *Peptide Nucleic Acids. Methods in Molecular Biology;* Nielsen, P. E., Ed.; Springer: Totowa, NJ, 2002; Vol. 208. doi:10.1385/1-59259-290-2:119
28. Sahu, B.; Sacui, I.; Rapireddy, S.; Zanotti, K. J.; Bahal, R.; Armitage, B. A.; Ly, D. H. *J. Org. Chem.* **2011,** *76,* 5614–5627. doi:10.1021/jo200482d
29. Sukiyama, T.; Kittaka, A. *Molecules* **2013,** *18,* 287–310. doi:10.3390/molecules18010287
30. Dragulescu-Andrasi, A.; Rapireddy, S.; Frezza, B. M.; Gayathri, C.; Gil, R. R.; Ly, D. H. *J. Am. Chem. Soc.* **2006,** *128,* 10258–10267. doi:10.1021/ja0625576
31. Kumar, V. A.; Ganesh, K. N. *Acc. Chem. Res.* **2005,** *38,* 404–412. doi:10.1021/ar030277e
32. Vilaivan, T. *Acc. Chem. Res.* **2015,** *48,* 1645–1656. doi:10.1021/acs.accounts.5b00080
33. Nielsen, P. E. *Methods Enzymol.* **2001,** *340,* 329–340. doi:10.1016/S0076-6879(01)40429-0
34. Li, M.; Zengeya, T.; Rozners, E. *J. Am. Chem. Soc.* **2010,** *132,* 8676–8681. doi:10.1021/ja101384k
35. Lusvarghi, S.; Murphy, C. T.; Roy, S.; Tanious, F. A.; Sacui, I.; Wilson, W. D.; Ly, D. H.; Armitage, B. A. *J. Am. Chem. Soc.* **2009,** *131,* 18415–18424. doi:10.1021/ja907250j
36. Metaferia, B.; Wei, J. S.; Song, Y. K.; Evangelista, J.; Aschenbach, K.; Johansson, P.; Wen, X.; Chen, Q.; Lee, A.; Hempel, H.; Gheeya, J. S.; Getty, S.; Gomez, R.; Khan, J. *PLoS One* **2013,** *8,* e58870. doi:10.1371/journal.pone.0058870
37. Shi, H.; Yang, F.; Li, W.; Zhao, W.; Nie, K.; Dong, B.; Liu, Z. *Biosens. Bioelectron.* **2015,** *66,* 481–489. doi:10.1016/j.bios.2014.12.010
38. Stender, H. *Expert Rev. Mol. Diagn.* **2003,** *3,* 649–655. doi:10.1586/14737159.3.5.649
39. Forrest, G. N. *Expert Rev. Mol. Diagn.* **2007,** *7,* 231–236. doi:10.1586/14737159.7.3.231
40. Kuhn, H.; Demidov, V. V.; Coull, J. M.; Fiandaca, M. J.; Gildea, B. D.; Frank-Kamenetskii, M. D. *J. Am. Chem. Soc.* **2002,** *124,* 1097–1103. doi:10.1021/ja0041324
41. Ortiz, E.; Estrada, G.; Lizardi, P. M. *Mol. Cell. Probes* **1998,** *12,* 219–226. doi:10.1006/mcpr.1998.0175
42. Armitage, B.; Ly, D.; Koch, T.; Frydenlund, H.; Ørum, H.; Schuster, G. B. *Biochemistry* **1998,** *37,* 9417–9425. doi:10.1021/bi9729458
43. Kitagawa, F.; Ohori, Y.; Ikeda, H.; Fujimori, H.; Murakami, Y.; Nakamura, Y. *Nucleic Acids Symp. Ser.* **2002,** *2,* 143–144. doi:10.1093/nass/2.1.143
44. Seitz, O. *Angew. Chem., Int. Ed.* **2000,** *39,* 3249–3252. doi:10.1002/1521-3773(20000915)39:18<3249::AID-ANIE3249>3.0.CO;2-M
45. Seitz, O.; Köhler, O. *Chem. – Eur. J.* **2001,** *7,* 3911–3925. doi:10.1002/1521-3765(20010917)7:18<3911::AID-CHEM3911>3.0.CO;2-1
46. Kuhn, H.; Demidov, V. V.; Gildea, B. D.; Fiandaca, M. J.; Coull, J. C.; Frank-Kamenetskii, M. D. *Antisense Nucleic Acid Drug Dev.* **2001,** *11,* 265–270. doi:10.1089/108729001317022269
47. Smolina, I. V.; Demidov, V. V.; Cantor, C. R.; Broude, N. E. *Anal. Biochem.* **2004,** *335,* 326–329. doi:10.1016/j.ab.2004.07.022
48. Morgan, J. R.; Lyon, R. P.; Maeda, D. Y.; Zebala, J. A. *Nucleic Acids Res.* **2008,** *36,* 3522–3530. doi:10.1093/nar/gkn219
49. Bonnet, G.; Tyagi, S.; Libchaber, A.; Kramer, F. R. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96,* 6171–6176. doi:10.1073/pnas.96.11.6171
50. Totsingan, F.; Rossi, S.; Corradini, R.; Tedeschi, T.; Sforza, S.; Juris, A.; Scaravelli, E.; Marchelli, R. *Org. Biomol. Chem.* **2008,** *6,* 1232–1237. doi:10.1039/b718772f
51. Totsingan, F.; Tedeschi, T.; Sforza, S.; Corradini, R.; Marchelli, A. *Chirality* **2009,** *21,* 245–253. doi:10.1002/chir.20659
52. Ikeda, H.; Kitagawa, F.; Nakamura, Y. *Tetrahedron* **2007,** *63,* 5677–5689. doi:10.1016/j.tet.2007.03.109
53. Moustafa, M. E.; Hudson, R. H. E. *Nucleosides, Nucleotides Nucleic Acids* **2011,** *30,* 740–751. doi:10.1080/15257770.2011.604661
54. Manicardi, A.; Bertucci, A.; Rozzi, A.; Corradini, R. *Org. Lett.* **2016,** *18,* 5452–5455. doi:10.1021/acs.orglett.6b02363
55. Wang, X.; Hudson, R. H. E. *ChemBioChem* **2015,** *16,* 2156–2161. doi:10.1002/cbic.201500248

56. Manicardi, A.; Accetta, A.; Tedeschi, T.; Sforza, S.; Marchelli, R.; Corradini, R. *Artif. DNA PNA XNA* **2012,** *3,* 53–62. doi:10.4161/adna.20158
57. Mitra, R.; Ganesh, K. N. *J. Org. Chem.* **2012,** *77,* 5696–5704. doi:10.1021/jo300860f
58. Jain, D. R.; Ganesh, K. N. *J. Org. Chem.* **2014,** *79,* 6708–6714. doi:10.1021/jo500834u
59. Wiessler, M.; Waldeck, W.; Pipkorn, R.; Kliem, C.; Lorenz, P.; Fleischhacker, H.; Hafner, M.; Braun, K. *Int. J. Med. Sci.* **2010,** *7,* 213–223. doi:10.7150/ijms.7.213
60. Shinozuka, K.; Sato, Y.; Sawai, H. *J. Chem. Soc., Chem. Commun.* **1994,** 1377–1378. doi:10.1039/c39940001377
61. Ranasinghe, R. T.; Brown, L. J.; Brown, T. *Chem. Commun.* **2001,** 1480–1481. doi:10.1039/b104421b
62. Kodama, S.; Asano, S.; Moriguchi, T.; Sawai, H.; Shinozuka, K. *Bioorg. Med. Chem. Lett.* **2006,** *16,* 2685–2688. doi:10.1016/j.bmcl.2006.02.016
63. Yamane, A. *Nucleic Acids Res.* **2002,** *30,* e97. doi:10.1093/nar/gnf096
64. Qiu, J.; Wilson, A.; El-Sagheer, A. H.; Brown, T. *Nucleic Acids Res.* **2016,** *44,* e138. doi:10.1093/nar/gkw579
65. Yotapan, N.; Nim-anussornkul, D.; Vilaivan, T. *Tetrahedron* **2016,** *72,* 7992–7999. doi:10.1016/j.tet.2016.10.040
66. Jockusch, S.; Martí, A. A.; Turro, N. J.; Li, Z.; Li, X.; Ju, J.; Stevens, N.; Akins, D. L. *Photochem. Photobiol. Sci.* **2006,** *5,* 493–498. doi:10.1039/b600213g
67. Wu, C.; Wang, C.; Yan, L.; Yang, C. J. *J. Biomed. Nanotechnol.* **2009,** *5,* 495–504. doi:10.1166/jbn.2009.1074
68. Fujimoto, K.; Shimizu, H.; Inouye, M. *J. Org. Chem.* **2004,** *69,* 3271–3275. doi:10.1021/jo049824f
69. Manicardi, A.; Guido, L.; Ghidini, A.; Corradini, R. *Beilstein J. Org. Chem.* **2014,** *10,* 1495–1503. doi:10.3762/bjoc.10.154
70. Maneelun, N.; Vilaivan, T. *Tetrahedron* **2013,** *69,* 10805–10810. doi:10.1016/j.tet.2013.10.096
71. Lee, K.; Park, H.-C.; An, S.; Ahn, E.-R.; Lee, Y.-H.; Kim, M.-J.; Lee, E.-j.; Park, J. S.; Jung, J. W.; Lim, S. *Leg. Med.* **2015,** *17,* 334–339. doi:10.1016/j.legalmed.2015.04.001
72. Ahn, J. J.; Kim, Y.; Lee, S. Y.; Hong, J. Y.; Kim, G. W.; Hwang, S. Y. *Anal. Biochem.* **2015,** *484,* 143–147. doi:10.1016/j.ab.2015.05.022
73. Kim, E.-M.; Song, M. S.; Hur, D. H.; An, C. M.; Kang, J.-H.; Park, J. Y. *BioChip J.* **2015,** *9,* 247–258. doi:10.1007/s13206-015-9402-1
74. Ahn, J. J.; Lee, S. Y.; Hong, J. Y.; Kim, Y.; Kim, G. W.; Hwang, S. Y. *Biotechnol. Prog.* **2015,** *31,* 730–735. doi:10.1002/btpr.2054
75. Hur, D.; Kim, M. S.; Song, M.; Jung, J.; Park, H. *Biol. Proced. Online* **2015,** *17,* 14. doi:10.1186/s12575-015-0027-5
76. Petersen, K.; Vogel, U.; Rockenbauer, E.; Nielsen, K. V.; Kølvraa, S.; Bolund, L.; Nexø, B. *Mol. Cell. Probes* **2004,** *18,* 117–122. doi:10.1016/j.mcp.2003.10.003
77. Xi, C.; Balberg, M.; Boppart, S. A.; Raskin, L. *Appl. Environ. Microbiol.* **2003,** *69,* 5673–5678. doi:10.1128/AEM.69.9.5673-5678.2003
78. Zanoli, L. M.; Licciardello, M.; D' Agata, R.; Lantano, C.; Calabretta, A.; Corradini, R.; Marchelli, R.; Spoto, G. *Anal. Bioanal. Chem.* **2013,** *405,* 615–624. doi:10.1007/s00216-011-5638-3
79. Rane, T. D.; Zec, H. C.; Puleo, C.; Lee, A. P.; Wang, T.-H. *Lab Chip* **2012,** *12,* 3341–3347. doi:10.1039/c2lc40537g
80. Thurley, S.; Röglin, L.; Seitz, O. *J. Am. Chem. Soc.* **2007,** *129,* 12693–12695. doi:10.1021/ja075487r
81. Oh, K. J.; Cash, K. J.; Lubin, A. A.; Plaxco, K. W. *Chem. Commun.* **2007,** 4869–4871. doi:10.1039/b709776j
82. Fischbach, M.; Resch-Genger, U.; Seitz, O. *Angew. Chem., Int. Ed.* **2014,** *53,* 11955–11959. doi:10.1002/anie.201406909
83. Kolpashchikov, D. M. *Chem. Rev.* **2010,** *110,* 4709–4723. doi:10.1021/cr900323b
84. Oquare, B. Y.; Taylor, J.-S. *Bioconjugate Chem.* **2008,** *19,* 2196–2204. doi:10.1021/bc800284x
85. Robertson, K. L.; Yu, L.; Armitage, B. A.; Lopez, A. J.; Peteanu, L. A. *Biochemistry* **2006,** *45,* 6066–6074. doi:10.1021/bi052050s
86. Chiou, C.-C.; Luo, J.-D.; Chen, T.-L. *Nat. Protoc.* **2006,** *1,* 2604–2612. doi:10.1038/nprot.2006.428
87. Wang, Z.; Zhang, K.; Shen, Y.; Smith, J.; Bloch, S.; Achilefu, S.; Wooley, K. L.; Taylor, J.-S. *Org. Biomol. Chem.* **2013,** *11,* 3159–3167. doi:10.1039/c3ob26923j
88. Silverman, A. P.; Kool, E. T. *Chem. Rev.* **2006,** *106,* 3775–3789. doi:10.1021/cr050057+
89. Percivalle, C.; Bartolo, J.-F.; Ladame, S. *Org. Biomol. Chem.* **2013,** *11,* 16–26. doi:10.1039/C2OB26163D
90. Pianowski, Z.; Gorska, K.; Oswald, L.; Merten, C. A.; Winssinger, N. *J. Am. Chem. Soc.* **2009,** *131,* 6492–6497. doi:10.1021/ja809656k
91. Dose, C.; Ficht, S.; Seitz, O. *Angew. Chem., Int. Ed.* **2006,** *45,* 5369–5373. doi:10.1002/anie.200600464
92. Dose, C.; Seitz, O. *Bioorg. Med. Chem.* **2008,** *16,* 65–77. doi:10.1016/j.bmc.2007.04.059
93. Meguellati, K.; Koripelly, G.; Ladame, S. *Angew. Chem., Int. Ed.* **2010,** *49,* 2738–2742. doi:10.1002/anie.201000291
94. Koripelly, G.; Meguellati, K.; Ladame, S. *Bioconjugate Chem.* **2010,** *21,* 2103–2109. doi:10.1021/bc100335f
95. Metcalf, G. A. D.; Shibakawa, A.; Patel, H.; Sita-Lumsden, A.; Zivi, A.; Rama, N.; Bevan, C. L.; Ladame, S. *Anal. Chem.* **2016,** *88,* 8091–8098. doi:10.1021/acs.analchem.6b01594
96. Al Sulaiman, D.; Chang, J. Y. H.; Ladame, S. *Angew. Chem., Int. Ed.* **2017,** *56,* 5247–5251. doi:10.1002/anie.201701356
97. Fang, G.-m.; Seitz, O. *ChemBioChem* **2017,** *18,* 189–194. doi:10.1002/cbic.201600623
98. Cai, J.; Li, X.; Yue, X.; Taylor, J. S. *J. Am. Chem. Soc.* **2004,** *126,* 16324–16325. doi:10.1021/ja0452626
99. Pianowski, Z. L.; Winssinger, N. *Chem. Commun.* **2007,** 3820–3822. doi:10.1039/b709611a
100. Gorska, K.; Manicardi, A.; Barluenga, S.; Winssinger, N. *Chem. Commun.* **2011,** *47,* 4364–4366. doi:10.1039/C1CC10222B
101. Gorska, K.; Keklikoglou, I.; Tschulena, U.; Winssinger, N. *Chem. Sci.* **2011,** *2,* 1969–1975. doi:10.1039/c1sc00216c
102. Cai, J.; Li, X.; Taylor, J. S. *Org. Lett.* **2005,** *7,* 751–754. doi:10.1021/ol0478382
103. Grossmann, T. N.; Seitz, O. *J. Am. Chem. Soc.* **2006,** *128,* 15596–15597. doi:10.1021/ja0670097
104. Grossmann, T. N.; Seitz, O. *Chem. – Eur. J.* **2009,** *15,* 6723–6730. doi:10.1002/chem.200900025
105. Sadhu, K. K.; Winssinger, N. *Chem. – Eur. J.* **2013,** *19,* 8182–8189. doi:10.1002/chem.201300060
106. Kushon, S. A.; Bradford, K.; Marin, V.; Suhrada, C.; Armitage, B. A.; McBranch, D.; Whitten, D. *Langmuir* **2003,** *19,* 6456–6464. doi:10.1021/la034323v
107. Habl, G.; Böhm, H.; Marmé, N.; Knemeyer, J.-P. *Int. J. Environ. Anal. Chem.* **2005,** *85,* 613–623. doi:10.1080/03067310500154940
108. Bohländer, P. R.; Vilaivan, T.; Wagenknecht, H.-A. *Org. Biomol. Chem.* **2015,** *13,* 9223–9230. doi:10.1039/C5OB01273B
109. Boonlua, C.; Charoenpakdee, C.; Vilaivan, T.; Praneenararat, T. *ChemistrySelect* **2016,** *1,* 5691–5697. doi:10.1002/slct.201601075

110. Grossmann, T. N.; Sasaki, S.; Ritzefeld, M.; Choi, S. W.; Maruyama, A.; Seitz, O. *Bioorg. Med. Chem.* **2008,** *16,* 34–39. doi:10.1016/j.bmc.2007.04.066
111. Matsumoto, K.; Nakata, E.; Tamura, T.; Saito, I.; Aizawa, Y.; Morii, T. *Chem. – Eur. J.* **2013,** *19,* 5034–5040. doi:10.1002/chem.201204183
112. Wang, Z.; Zhang, K.; Wooley, K. L.; Taylor, J.-S. *J. Nucleic Acids* **2012,** No. 962652. doi:10.1155/2012/962652
113. Fiandaca, M. J.; Hyldig-Nielsen, J. J.; Gildea, B. D.; Coull, J. M. *Genome Res.* **2001,** *11,* 609–613. doi:10.1101/gr.170401
114. For a previous review that covers this topic see: Liu, X.; Fan, Q.; Huang, W. *Biosens. Bioelectron.* **2011,** *26,* 2154–2164. doi:10.1016/j.bios.2010.09.025
115. Gaylord, B. S.; Heeger, A. J.; Bazan, G. C. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 10954–10957. doi:10.1073/pnas.162375999
116. Gaylord, B. S.; Heeger, A. J.; Bazan, G. C. *J. Am. Chem. Soc.* **2003,** *125,* 896–900. doi:10.1021/ja027152+
117. Gaylord, B. S.; Massie, M. R.; Feinstein, S. C.; Bazan, G. C. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102,* 34–39. doi:10.1073/pnas.0407578101
118. Xu, Q.-H.; Gaylord, B. S.; Wang, S.; Bazan, G. C.; Moses, D.; Heeger, A. J. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101,* 11634–11639. doi:10.1073/pnas.0404574101
119. Al Attar, H. A.; Norden, J.; O'Brien, S.; Monkman, A. P. *Biosens. Bioelectron.* **2008,** *23,* 1466–1472. doi:10.1016/j.bios.2008.01.005
120. Al Attar, H. A.; Monkman, A. P. *J. Phys. Chem. B* **2007,** *111,* 12418–12426. doi:10.1021/jp070827g
121. Mathur, N.; Aneja, A.; Bhatnagar, P. K.; Mathur, P. C. *J. Biomater. Sci., Polym. Ed.* **2011,** *22,* 379–387. doi:10.1163/092050610X486973
122. Baker, E. S.; Hong, J. W.; Gaylord, B. S.; Bazan, G. C.; Bowers, M. T. *J. Am. Chem. Soc.* **2006,** *128,* 8484–8492. doi:10.1021/ja060069s
123. Li, K.; Liu, B. *Anal. Chem.* **2009,** *81,* 4099–4105. doi:10.1021/ac9003985
124. Rashatasakhon, P.; Vongnam, K.; Siripornnoppakhun, W.; Vilaivan, T.; Sukwattanasinitt, M. *Talanta* **2012,** *88,* 593–598. doi:10.1016/j.talanta.2011.11.041
125. Liu, B.; Bazan, G. C. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102,* 589–593. doi:10.1073/pnas.0408561102
126. Sun, C.; Gaylord, B. S.; Hong, J. W.; Liu, B.; Bazan, G. C. *Nat. Protoc.* **2007,** *2,* 2148–2151. doi:10.1038/nprot.2007.307
127. Wang, C.; Zhan, R.; Pu, K.-Y.; Liu, B. *Adv. Funct. Mater.* **2010,** *20,* 2597–2604. doi:10.1002/adfm.201000459
128. Ho, H.-A.; Boissinot, M.; Bergeron, M. G.; Corbeil, G.; Doré, K.; Boudreau, D.; Leclerc, M. *Angew. Chem., Int. Ed.* **2002,** *41,* 1548–1551. doi:10.1002/1521-3773(20020503)41:9<1548::AID-ANIE1548>3.0.CO;2-I
129. Raymond, F. R.; Ho, H.-A.; Peytavi, R.; Bissonnette, L.; Boissinot, M.; Picard, F. J.; Leclerc, M.; Begeron, M. G. *BMC Biotechnol.* **2005,** *5,* 10. doi:10.1186/1472-6750-5-10
130. Lee, J.; Kim, J.; Kim, S.; Min, D.-H. *Adv. Drug Delivery Rev.* **2016,** *105,* 275–287. doi:10.1016/j.addr.2016.06.001
131. Liu, Z.; Liu, B.; Ding, J.; Liu, J. *Anal. Bioanal. Chem.* **2014,** *406,* 6885–6902. doi:10.1007/s00216-014-7888-3
132. Zeng, S.; Chen, L.; Wang, Y.; Chen, J. *J. Phys. D* **2015,** *48,* 275402. doi:10.1088/0022-3727/48/27/275402
133. Lu, C.-H.; Yang, H.-H.; Zhu, C.-L.; Chen, X.; Chen, G.-N. *Angew. Chem., Int. Ed.* **2009,** *48,* 4785–4787. doi:10.1002/anie.200901479
134. He, S.; Song, B.; Li, D.; Zhu, C.; Qi, W.; Wen, Y.; Wang, L.; Song, S.; Fang, H.; Fan, C. *Adv. Funct. Mater.* **2010,** *20,* 453–459. doi:10.1002/adfm.200901639
135. Liu, B.; Sun, Z.; Zhang, X.; Liu, J. *Anal. Chem.* **2013,** *85,* 7987–7993. doi:10.1021/ac401845p
136. Wu, M.; Kempaiah, R.; Huang, P.-J. J.; Maheshwari, V.; Liu, J. *Langmuir* **2011,** *27,* 2731–2738. doi:10.1021/la1037926
137. Park, J.; Goo, N.-I.; Kim, D.-E. *Langmuir* **2014,** *30,* 12587–12595. doi:10.1021/la503401d
138. Kotikam, V.; Fernandes, M.; Kumar, V. A. *Phys. Chem. Chem. Phys.* **2012,** *14,* 15003–15006. doi:10.1039/c2cp42770b
139. Guo, S.; Du, D.; Tang, L.; Ning, Y.; Yao, Q.; Zhang, G.-J. *Analyst* **2013,** *138,* 3216–3220. doi:10.1039/c3an00266g
140. Lee, J.; Park, G.; Min, D.-H. *Chem. Commun.* **2015,** *51,* 14597–14600. doi:10.1039/C5CC04706D
141. Lee, J.; Park, I.-S.; Jung, E.; Lee, Y.; Min, D.-H. *Biosens. Bioelectron.* **2014,** *62,* 140–144. doi:10.1016/j.bios.2014.06.028
142. Lee, J.; Park, I.-S.; Kim, H.; Woo, J.-S.; Choi, B.-S.; Min, D.-H. *Biosens. Bioelectron.* **2015,** *69,* 167–173. doi:10.1016/j.bios.2015.02.030
143. Hong, C.; Baek, A.; Hah, S. S.; Jung, W.; Kim, D.-E. *Anal. Chem.* **2016,** *88,* 2999–3003. doi:10.1021/acs.analchem.6b00046
144. Park, J. S.; Baek, A.; Park, I.-S.; Jun, B.-H.; Kim, D.-E. *Chem. Commun.* **2013,** *49,* 9203–9205. doi:10.1039/c3cc45750h
145. Ryoo, S.-R.; Lee, J.; Yeo, J.; Na, H.-K.; Kim, Y.-K.; Jang, H.; Lee, J. H.; Han, S. W.; Lee, Y.; Kim, V. N.; Min, D.-H. *ACS Nano* **2013,** *7,* 5882–5891. doi:10.1021/nn401183s
146. Hwang, D. W.; Kim, H. Y.; Li, F.; Park, J. Y.; Kim, D.; Park, J. H.; Han, H. S.; Byun, J. W.; Lee, Y.-S.; Jeong, J. M.; Char, K.; Lee, D. S. *Biomaterials* **2017,** *121,* 144–154. doi:10.1016/j.biomaterials.2016.12.028
147. Wang, S.; Zhang, Y.; Ning, Y.; Zhang, G.-J. *Analyst* **2015,** *140,* 434–439. doi:10.1039/C4AN01738B
148. Liao, X.; Wang, Q.; Ji, H. *Analyst* **2015,** *140,* 4245–4252. doi:10.1039/C5AN00128E
149. Wu, Y.; Han, J.; Xue, P.; Xu, R.; Kang, Y. *Nanoscale* **2015,** *7,* 1753–1759. doi:10.1039/C4NR05447D
150. Stanisavljevic, M.; Krizkova, S.; Vaculovicova, M.; Kizek, R.; Adam, V. *Biosens. Bioelectron.* **2015,** *74,* 562–574. doi:10.1016/j.bios.2015.06.076
151. Lu, W.; Qin, X.; Luo, Y.; Chang, G.; Sun, X. *Microchim. Acta* **2011,** *175,* 355–359. doi:10.1007/s00604-011-0657-5
152. Algar, W. R.; Massey, M.; Krull, U. J. *TrAC, Trends Anal. Chem.* **2009,** *28,* 292–306. doi:10.1016/j.trac.2008.11.012
153. Noor, M. O.; Petryayeva, E.; Tavares, A. J.; Uddayasankar, U.; Algar, W. R.; Krull, U. J. *Coord. Chem. Rev.* **2014,** *263-264,* 25–52. doi:10.1016/j.ccr.2013.08.010
154. Lee, J.; Choi, Y.; Kim, J.; Park, E.; Song, R. *ChemPhysChem* **2009,** *10,* 806–811. doi:10.1002/cphc.200800504
155. Noor, M. O.; Shahmuradyan, A.; Krull, U. J. *Anal. Chem.* **2013,** *85,* 1860–1867. doi:10.1021/ac3032383
156. Yang, L.-H.; Ahn, D. J.; Koo, E. *Mater. Sci. Eng., C* **2016,** *69,* 625–630. doi:10.1016/j.msec.2016.07.021
157. Svanvik, N.; Westman, G.; Wang, D.; Kubista, M. *Anal. Biochem.* **2000,** *281,* 26–35. doi:10.1006/abio.2000.4534
158. Nygren, J.; Svanvik, N.; Kubista, M. *Biopolymers* **1998,** *46,* 39–51. doi:10.1002/(SICI)1097-0282(199807)46:1<39::AID-BIP4>3.0.CO;2-Z

159. Svanvik, N.; Nygren, J.; Westman, G.; Kubista, M. *J. Am. Chem. Soc.* **2001,** *123,* 803–809. doi:10.1021/ja002294u
160. Isacsson, J.; Cao, H.; Ohlsson, L.; Nordgren, S.; Svanvik, N.; Westman, G.; Kubista, M.; Sjöback, R.; Sehlstedt, U. *Mol. Cell. Probes* **2000,** *14,* 321–328. doi:10.1006/mcpr.2000.0321
161. Wolffs, P.; Knutsson, R.; Sjöback, R.; Rådström, P. *BioTechniques* **2001,** *31,* 766–771.
162. Svanvik, N.; Ståhlberg, A.; Sehlstedt, U.; Sjöback, R.; Kubista, M. *Anal. Biochem.* **2000,** *187,* 179–182. doi:10.1006/abio.2000.4824
163. Tanaka, M.; Shigi, N.; Sumaoka, J.; Komiyama, M. *RSC Adv.* **2014,** *4,* 63533–63538. doi:10.1039/C4RA13780A
164. Sato, T.; Sato, Y.; Iwai, K.; Kuge, S.; Nishizawa, S.; Teramae, N. *Chem. Commun.* **2015,** *51,* 1421–1424. doi:10.1039/C4CC08800J
165. Sato, T.; Sato, Y.; Iwai, K.; Kuge, S.; Teramae, N.; Nishizawa, S. *Anal. Sci.* **2015,** *31,* 315–320. doi:10.2116/analsci.31.315
166. Englund, E. A.; Appella, D. H. *Angew. Chem., Int. Ed.* **2007,** *46,* 1414–1418. doi:10.1002/anie.200603483
167. Ditmangklo, B.; Boonlua, C.; Suparpprom, C.; Vilaivan, T. *Bioconjugate Chem.* **2013,** *24,* 614–625. doi:10.1021/bc3005914
168. Englund, E. A.; Appella, D. H. *Org. Lett.* **2005,** *7,* 3465–3467. doi:10.1021/ol051143z
169. Reenabthue, N.; Boonlua, C.; Vilaivan, C.; Vilaivan, T.; Suparpprom, C. *Bioorg. Med. Chem. Lett.* **2011,** *21,* 6465–6469. doi:10.1016/j.bmcl.2011.08.079
170. Boonlua, C.; Ditmangklo, B.; Reenabthue, N.; Suparpprom, C.; Poomsuk, N.; Siriwong, K.; Vilaivan, T. *RSC Adv.* **2014,** *4,* 8817–8827. doi:10.1039/C3RA47997H
171. Dangsopon, A.; Poomsuk, N.; Siriwong, K.; Vilaivan, T.; Suparpprom, C. *RSC Adv.* **2016,** *6,* 74314–74322. doi:10.1039/C6RA15861G
172. Yotapan, N.; Charoenpakdee, C.; Wathanathavorn, P.; Ditmangklo, B.; Wagenknecht, H.-A.; Vilaivan, T. *Beilstein J. Org. Chem.* **2014,** *10,* 2166–2174. doi:10.3762/bjoc.10.224
173. Yamana, K.; Iwase, R.; Furutani, S.; Tsuchida, H.; Zako, H.; Yamaoka, T.; Murakami, A. *Nucleic Acids Res.* **1999,** *27,* 2387–2392. doi:10.1093/nar/27.11.2387
174. Nakamura, M.; Fukunaga, Y.; Sasa, K.; Ohtoshi, Y.; Kanaori, K.; Hayashi, H.; Nakano, H.; Yamana, K. *Nucleic Acids Res.* **2005,** *33,* 5887–5895. doi:10.1093/nar/gki889
175. Okamoto, A.; Kanatani, K.; Saito, I. *J. Am. Chem. Soc.* **2004,** *126,* 4820–4827. doi:10.1021/ja039625y
176. Hwang, G. T.; Seo, Y. J.; Kim, B. H. *J. Am. Chem. Soc.* **2004,** *126,* 6528–6529. doi:10.1021/ja049795q
177. Wilhelmsson, L. M. *Q. Rev. Biophys.* **2010,** *43,* 159–183. doi:10.1017/S0033583510000090
178. Wilson, J. N.; Kool, E. T. *Org. Biomol. Chem.* **2006,** *4,* 4265–4274. doi:10.1039/b612284c
179. Gangamani, B. P.; Kumar, V. A. *Chem. Commun.* **1997,** 1913–1914. doi:10.1039/a705539k
180. Amant, A. H. S.; Hudson, R. H. E. *Org. Biomol. Chem.* **2012,** *10,* 876–881. doi:10.1039/C1OB06582C
181. Müllar, S.; Strohmeier, J.; Diederichsen, U. *Org. Lett.* **2012,** *14,* 1382–1385. doi:10.1021/ol3000603
182. Lin, K.-Y.; Jones, R. J.; Matteucci, M. *J. Am. Chem. Soc.* **1995,** *117,* 3873–3874. doi:10.1021/ja00118a026
183. Wilhelmsson, L. M.; Holmén, A.; Lincoln, P.; Nielsen, P. E.; Nordén, B. *J. Am. Chem. Soc.* **2001,** *123,* 2434–2435. doi:10.1021/ja0025797
184. Eldrup, A. B.; Nielsen, B. B.; Haaima, G.; Rasmussen, H.; Kastrup, J. S.; Christensen, C.; Nielsen, P. E. *Eur. J. Org. Chem.* **2001,** 1781–1790. doi:10.1002/1099-0690(200105)2001:9<1781::AID-EJOC1781>3.0.CO;2-K
185. Rajeev, K. G.; Maier, M. A.; Lesnik, E. A.; Manoharan, M. *Org. Lett.* **2002,** *4,* 4395–4398. doi:10.1021/ol027026a
186. Ausin, C.; Ortega, J.-A.; Robles, J.; Grandas, A.; Pedroso, E. *Org. Lett.* **2002,** *4,* 4073–4075. doi:10.1021/ol026815p
187. Ortega, J.-A.; Blas, J. R.; Orozco, M.; Grandas, A.; Pedroso, E.; Robles, J. *Org. Lett.* **2007,** *9,* 4503–4506. doi:10.1021/ol701826x
188. Wahba, A. S.; Esmaeili, A.; Damha, M. J.; Hudson, R. H. E. *Nucleic Acids Res.* **2010,** *38,* 1048–1056. doi:10.1093/nar/gkp1022
189. Zhang, H.; Wang, M.; Gao, Q.; Qi, H.; Zhang, Q. *Talanta* **2011,** *84,* 771–776. doi:10.1016/j.talanta.2011.02.005
190. Wojciechowski, F.; Hudson, R. H. E. *J. Am. Chem. Soc.* **2008,** *130,* 12574–12575. doi:10.1021/ja804233g
191. Wojciechowski, F.; Hudson, R. H. E. *Org. Lett.* **2009,** *11,* 4878–4881. doi:10.1021/ol9019474
192. Hu, J.; Dodd, D. W.; Hudson, R. H. E.; Corey, D. R. *Bioorg. Med. Chem. Lett.* **2009,** *19,* 6181–6184. doi:10.1016/j.bmcl.2009.09.004
193. Suchý, M.; Hudson, R. H. E. *J. Org. Chem.* **2014,** *79,* 3336–3347. doi:10.1021/jo402873e
194. Cheruiyot, S. K.; Rozners, E. *ChemBioChem* **2016,** *17,* 1558–1562. doi:10.1002/cbic.201600182
195. Sabale, P. M.; Srivatsan, S. G. *ChemBioChem* **2016,** *17,* 1665–1673. doi:10.1002/cbic.201600192
196. Sabale, P. M.; George, J. T.; Srivatsan, S. G. *Nanoscale* **2014,** *6,* 10460–10469. doi:10.1039/C4NR00878B
197. Boonlua, C.; Vilaivan, C.; Wagenknecht, H.-A.; Vilaivan, T. *Chem. – Asian J.* **2011,** *6,* 3251–3259. doi:10.1002/asia.201100490
198. Mayer-Enthart, E.; Wagenknecht, H.-A. *Angew. Chem., Int. Ed.* **2006,** *45,* 3372–3375. doi:10.1002/anie.200504210
199. Wanninger-Weiss, C.; Wagenknecht, H.-A. *Eur. J. Org. Chem.* **2008,** 64–71. doi:10.1002/ejoc.200700818
200. Nim-anussornkul, D.; Vilaivan, T. *Bioorg. Med. Chem.* **2017,** *25,* 6388–6397. doi:10.1016/j.bmc.2017.10.008
201. Seitz, O.; Bergmann, F.; Heindl, D. *Angew. Chem., Int. Ed.* **1999,** *38,* 2203–2206. doi:10.1002/(SICI)1521-3773(19990802)38:15<2203::AID-ANIE2203>3.0.CO;2-2
202. Hövelmann, F.; Seitz, O. *Acc. Chem. Res.* **2016,** *49,* 714–723. doi:10.1021/acs.accounts.5b00546
203. Köhler, O.; Jarikote, D. V.; Singh, I.; Parmar, V. S.; Weinhold, E.; Seitz, O. *Pure Appl. Chem.* **2005,** *77,* 327–338. doi:10.1351/pac200577010327
204. Köhler, O.; Seitz, O. *Chem. Commun.* **2003,** 2938–2939. doi:10.1039/B308299G
205. Loakes, D. *Nucleic Acids Res.* **2001,** *29,* 2437–2447. doi:10.1093/nar/29.12.2437
206. Köhler, O.; Jarikote, D. V.; Seitz, O. *Chem. Commun.* **2004,** 2674–2675. doi:10.1039/B411877D
207. Jarikote, D. V.; Krebs, N.; Tannert, S.; Röder, B.; Seitz, O. *Chem. – Eur. J.* **2006,** *13,* 300–310. doi:10.1002/chem.200600699
208. Jarikote, D. V.; Köhler, O.; Socher, E.; Seitz, O. *Eur. J. Org. Chem.* **2005,** 3187–3195. doi:10.1002/ejoc.200500201
209. Köhler, O.; Jarikote, D. V.; Seitz, O. *ChemBioChem* **2005,** *6,* 69–77. doi:10.1002/cbic.200400260

210. Socher, E.; Jarikote, D. V.; Knoll, A.; Röglin, L.; Burmeister, J.; Seitz, O. *Anal. Biochem.* **2008,** *375,* 318–330. doi:10.1016/j.ab.2008.01.009
211. Tonelli, A.; Tedeschi, T.; Germini, A.; Sforza, S.; Corradini, R.; Medici, M. C.; Chezzi, C.; Marchelli, R. *Mol. BioSyst.* **2011,** *7,* 1684–1692. doi:10.1039/c0mb00353k
212. Sato, T.; Sato, Y.; Nishizawa, S. *J. Am. Chem. Soc.* **2016,** *138,* 9397–9400. doi:10.1021/jacs.6b05554
213. Sato, T.; Sato, Y.; Nishizawa, S. *Chem. – Eur. J.* **2017,** *23,* 4079–4088. doi:10.1002/chem.201604676
214. Kummer, S.; Knoll, A.; Socher, E.; Bethge, L.; Herrmann, A.; Seitz, O. *Angew. Chem., Int. Ed.* **2011,** *50,* 1931–1934. doi:10.1002/anie.201005902
215. Kam, Y.; Rubinstein, A.; Nissan, A.; Halle, D.; Yavin, E. *Mol. Pharmaceutics* **2012,** *9,* 685–693. doi:10.1021/mp200505k
216. Torres, A. G.; Fabani, M. M.; Vigorito, E.; Williams, D.; Al-Obaidi, N.; Wojciechowski, F.; Hudson, R. H. E.; Seitz, O.; Gait, M. J. *Nucleic Acids Res.* **2012,** *40,* 2152–2167. doi:10.1093/nar/gkr885
217. Tilsner, J.; Flors, C. *ChemBioChem* **2011,** *12,* 1007–1009. doi:10.1002/cbic.201100139
218. Kam, Y.; Rubinstein, A.; Naik, S.; Djavsarov, I.; Halle, D.; Ariel, I.; Gure, A. O.; Stojadinovic, A.; Pan, H.; Tsivin, V.; Nissan, A.; Yavin, E. *Cancer Lett.* **2014,** *352,* 90–96. doi:10.1016/j.canlet.2013.02.014
219. Bethge, L.; Jarikote, D. V.; Seitz, O. *Bioorg. Med. Chem.* **2008,** *16,* 114–125. doi:10.1016/j.bmc.2006.12.044
220. Kummer, S.; Knoll, A.; Socher, E.; Bethge, L.; Herrmann, A.; Seitz, O. *Bioconjugate Chem.* **2012,** *23,* 2051–2060. doi:10.1021/bc300249f
221. Kolevzon, N.; Hashoul, D.; Naik, S.; Rubinstein, A.; Yavin, E. *Chem. Commun.* **2016,** *52,* 2405–2407. doi:10.1039/C5CC07502E
222. Socher, E.; Bethge, L.; Knoll, A.; Jungnick, N.; Hermann, A.; Seitz, O. *Angew. Chem., Int. Ed.* **2008,** *47,* 9555–9559. doi:10.1002/anie.200803549
223. Socher, E.; Knoll, A.; Seitz, O. *Org. Biomol. Chem.* **2012,** *10,* 7363–7371. doi:10.1039/c2ob25925g
224. MacKinnon, K. F.; Qualley, D. F.; Woski, S. A. *Tetrahedron Lett.* **2007,** *48,* 8074–8077. doi:10.1016/j.tetlet.2007.09.019
225. Tedeschi, T.; Tonelli, A.; Sforza, S.; Corradini, R.; Marchelli, R. *Artif. DNA PNA XNA* **2010,** *1,* 83–89. doi:10.4161/adna.1.2.13899
226. Mansawat, W.; Boonlua, C.; Siriwong, K.; Vilaivan, T. *Tetrahedron* **2012,** *68,* 3988–3995. doi:10.1016/j.tet.2012.03.062
227. Li, J. D.; Chen, M.; Liu, S.; Zhang, H. B.; Liu, Z. L. *Chin. Chem. Lett.* **2008,** *19,* 807–810. doi:10.1016/j.cclet.2008.05.023
228. Ikeda, H.; Yoshida, K.; Ozeki, M.; Saito, I. *Tetrahedron Lett.* **2001,** *42,* 2529–2531. doi:10.1016/S0040-4039(01)00228-3
229. Ikeda, H.; Nakamura, Y.; Saito, I. *Tetrahedron Lett.* **2002,** *43,* 5525–5528. doi:10.1016/S0040-4039(02)01164-4
230. Okamoto, A.; Tanabe, K.; Saito, I. *Org. Lett.* **2001,** *3,* 925–927. doi:10.1021/ol015549x

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Preparation of trinucleotide phosphoramidites as synthons for the synthesis of gene libraries

Ruth Suchsland, Bettina Appel and Sabine Müller*

Review

Open Access

Address:
Institut für Biochemie, Ernst-Moritz-Arndt-Universität Greifswald, Felix-Hausdorff-Str. 4, D-17489 Greifswald, Germany

Email:
Sabine Müller* - smueller@uni-greifswald.de

* Corresponding author

Keywords:
gene library; protein engineering; soluble support; synthesis; trinucleotide

Beilstein J. Org. Chem. **2018,** *14,* 397–406.
doi:10.3762/bjoc.14.28

Received: 28 November 2017
Accepted: 24 January 2018
Published: 13 February 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

The preparation of protein libraries is a key issue in protein engineering and biotechnology. Such libraries can be prepared by a variety of methods, starting from the respective gene library. The challenge in gene library preparation is to achieve controlled total or partial randomization at any predefined number and position of codons of a given gene, in order to obtain a library with a maximum number of potentially successful candidates. This purpose is best achieved by the usage of trinucleotide synthons for codon-based gene synthesis. We here review the strategies for the preparation of fully protected trinucleotides, emphasizing more recent developments for their synthesis on solid phase and on soluble polymers, and their use as synthons in standard DNA synthesis.

Introduction

Protein engineering is a highly actual research area with a number of potential applications [1-4]. The construction, adaptation and optimization of proteins can proceed by two major strategies: (i) rational design or (ii) directed evolution. The rational design is based on the introduction of point mutations, insertions or deletions at a defined position of the protein sequence, and requires detailed knowledge of the protein structure and the mechanism of action. On the opposite, directed evolution relies on the selection of a mutant with predefined properties from a random protein library. This strategy is advantageous over the rational design; whenever molecular properties of proteins are investigated that are not yet sufficiently understood, if properties like solvent or temperature stability need to be optimized, or regio-, chemo- or enantioselectivity and substrate specificity shall be changed. Thus, the optimization and variation of proteins, in particular of enzymes, by random mutagenesis and subsequent selection and identification of mutants with improved properties is a favoured method in the field of white biotechnology and biocatalysis, to improve the fitness of enzymes for industrial application [5].

In general, directed evolution may be summarized as an iterative two-step process which involves the generation of protein mutant libraries and high throughput screening processes to select for variants with improved traits. Protein mutant libraries are produced from gene libraries, which are generated by random mutagenesis at DNA level. Often polymerase chain raction (PCR)-based methods like error-prone PCR are used for this purpose as well as recombinant methods like DNA shuffling and related strategies [6,7]. One of the major challenges in gene library production is to generate libraries with a high number of promising candidates to enhance the chance of selecting functional protein variants. The methods mentioned above allow the degree and localization of randomization to be adjusted to a certain degree, however, full control over mutagenesis is still rather limited. Oligonucleotide-based methods with a number of sophisticated techniques [8] are advantageous here, as they offer a better possibility to control randomization. The basic principle consists of using chemically synthesized primers of mixed composition for introducing subsets of the 20 canonical amino acids at a defined position of the protein [9]. In the simplest way, a mixture of the four standard nucleotides is used for coupling at each randomized position of the primer in DNA synthesis. For a primer with 9 randomized positions (corresponding to three randomized amino acids in the resulting protein) this would lead to $4^9 = 262144$ sequence variants including stop codons and codons of undesired amino acids, and a bias towards amino acids encoded by multiple codons. Moreover, it is impossible to restrict randomization to a defined subset of amino acids at a desired position. Thus, the result is a rather large library, however, with only a small number of potentially successful candidates. There are strategies to at least partially circumvent this problem, like using NNS instead of NNN codons (with N = A, C, G, T; S = C, G) taking advantage of redundancy of the third nucleotide positions in the majority of codons [10], or using spiked oligonucleotides [11], which are synthesized from solutions of the four nucleotide building blocks, each of those contaminated with a "spiking mix" consisting of equal aliquots of each of the four building blocks [9,12]. The required volume of the spiking mix to achieve a desired amount of nucleotide replacements at a defined position of the oligonucleotide can be calculated, such that library size and degree of randomization can be restricted [13,14]. Nevertheless, although those methods and sophisticated variations of them [14-17] have improved library design and synthesis, full control over randomization is not possible. This can be achieved only by the usage of trinucleotide synthons for codon-based synthesis of a desired primer [18]. Taking the example from above, for a DNA fragment encoding three randomized amino acids, instead of nine nucleotide positions to be randomized, variation of trinucleotides (codons for the 20 amino acids) at only three positions is required. Therefore, the number of possible sequence variants in the gene library decreases from $4^9 = 262144$ to $20^3 = 8000$, if the full set of the 20 amino acids is desired at each of the three randomized positions. The library size can be even further decreased by using subsets of amino acids (e.g., only basic or only acidic amino acids) at the individual positions. Furthermore, stop codons as well as bias to amino acids with codon redundancy are completely prevented. Not at last, the coupling efficiency of individual trinucleotide synthons in chemical DNA synthesis can be considered when preparing the trinucleotide mixture, to ensure that each of the trinucleotides is coupled with identical statistical probability, or alternatively, to adjust the trinucleotide mixture to a desired amino acid distribution at the respective position. Thus, the application of trinucleotide building blocks for the synthesis of gene libraries stands out as facilitating fully controlled total or partial randomization at any predefined number and position of codons of a given gene. Trinucleotide synthons need to be chemically synthesized. Here, the challenge has been to find a suitable set of orthogonal protecting groups that allows the preparation of the trinucleotide, its conversion into a coupling competent building block, and its subsequent use in chemical DNA synthesis. Trinucleotides have been prepared in solution [19], on solid phase [20], and more recently on soluble polymers [21-23] (Figure 1), followed by phosphitylation to be used in standard DNA synthesis.

The preparation of mixed oligonucleotides for random mutagenesis including the strategy of using trinucleotide synthons has been reviewed recently [19,24]. Therefore, herein we will concentrate on more recent developments in trinucleotide synthesis.

Review

1. Preparation of trinucleotides in solution

Over the years, a number of methodologies has been published, varying in the protecting group for the phosphate moiety being methyl [25], ethyl [26], cyanoethyl [27] or *ortho*-chlorophenyl [28,29], and for the 3'-OH-group being phenoxyacetyl [25], dimethoxytrityl (DMTr) [26], *tert*-butyldimethylsilyl (TBDMS) [27,30], levulinoyl [26], or 2-azidomethylbenzoyl [27] (Figure 2), and applying either phosphotriester chemistry [28,29,31,32] or phosphite triester chemistry [25-27,30] in solution.

In general, trinucleotides can be assembled through the reaction of two suitably protected monomers to generate a dinucleotide, which then can be extended in either 5'- or 3'-direction (Figure 3).

Surprisingly, only one report has made use of this "economy", first coupling a 5'-*O*-DMTr-protected nucleoside-3'-*ortho*-

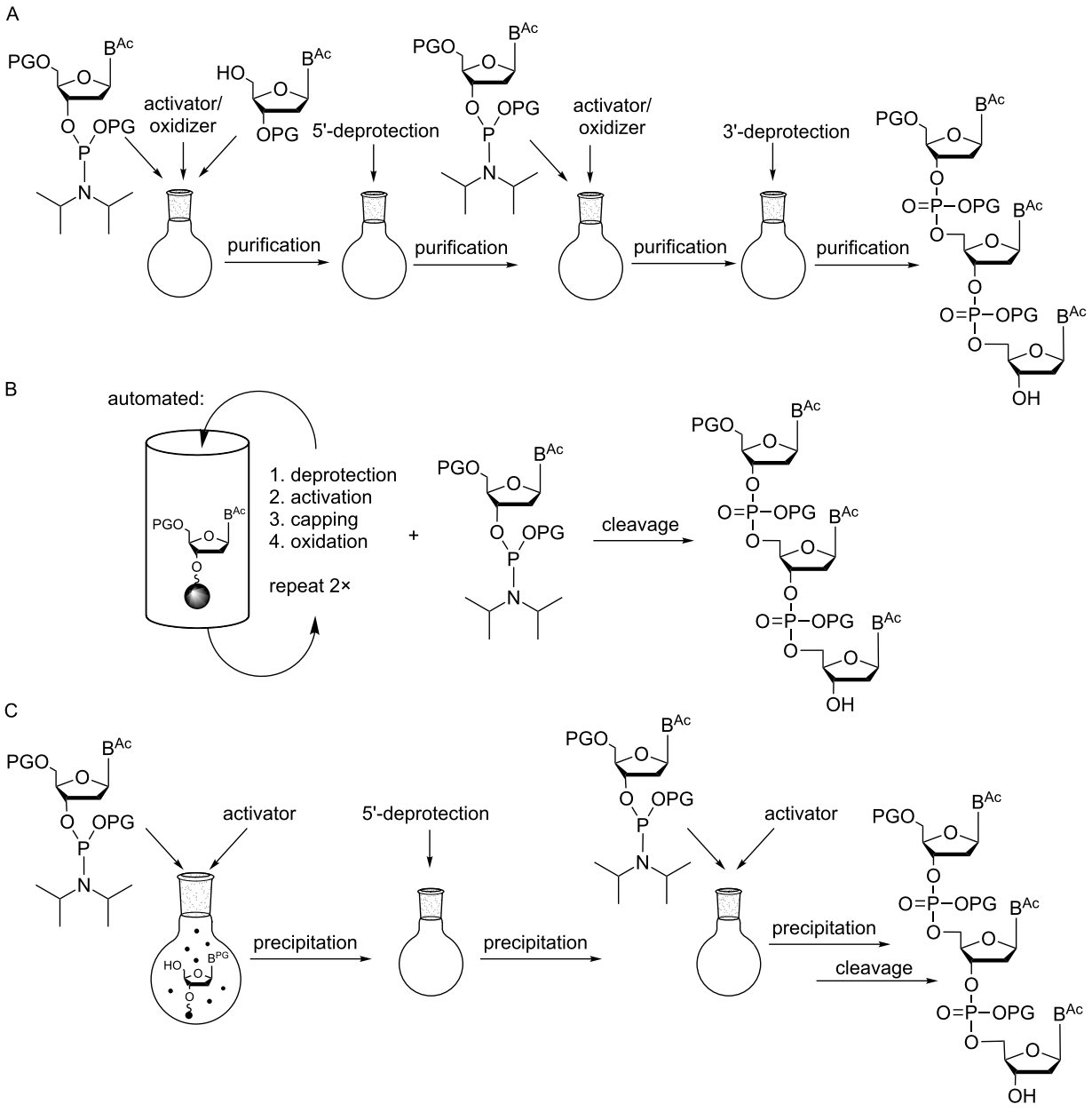


Figure 1: Preparation of fully protected trinucleotides in solution (A), on solid phase (B) and on soluble polymers (C).

chlorophenylphosphotriester to a 3'-*O*-levulinoyl-protected monomer. Upon selective removal of either the 5'-*O*-DMTr group or the 3'-*O*-levulinoyl group, the dimer was extended in 5' or 3' direction [26]. All other reports describe strategies, where the dimers are extended unidirectional, either in 5'-direction [25-27,29,30] or 3'-direction [31,32]. A key issue in all these methodologies is that the 5'- or the 3'-O-protecting group is selectively cleaved, whereas all other protecting groups (at the nucleobases, the phosphorous and the 5'- or alternatively 3'-OH group) remain intact.

Basically, this aim has been achieved, although in particular in earlier reports a number of problems associated with insufficient stability of protecting groups under synthesis conditions, as well as restricted orthogonality have been described, which was mirrored in the sometimes severely limited quality of the trinucleotide synthons and accordingly of the prepared oligonucleotide libraries [14,15,25,26,28,30,31,33]. Among the described procedures the use of *tert*-butyldimethylsilyl [25] and 2-azidomethylbenzoyl groups [29] for 3'-O-protection stands out as being the most successful in terms of high quality tri-

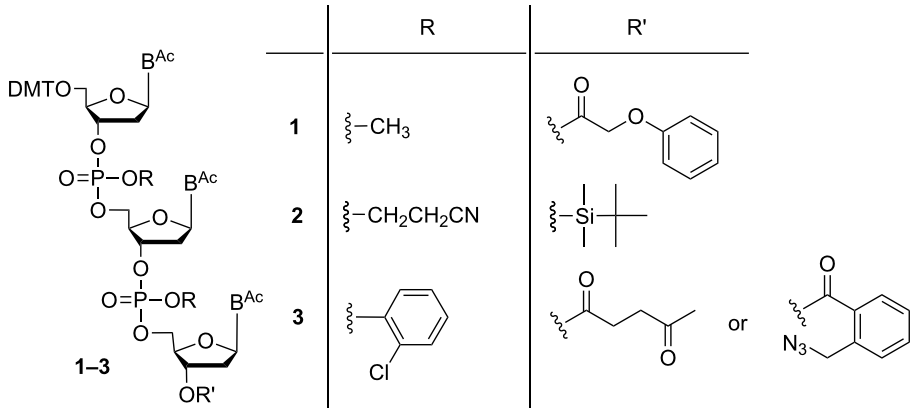


Figure 2: Strategies for trinucleotide synthesis using different pairs of orthogonal groups for protection of the phosphates and the 3'-OH-function.

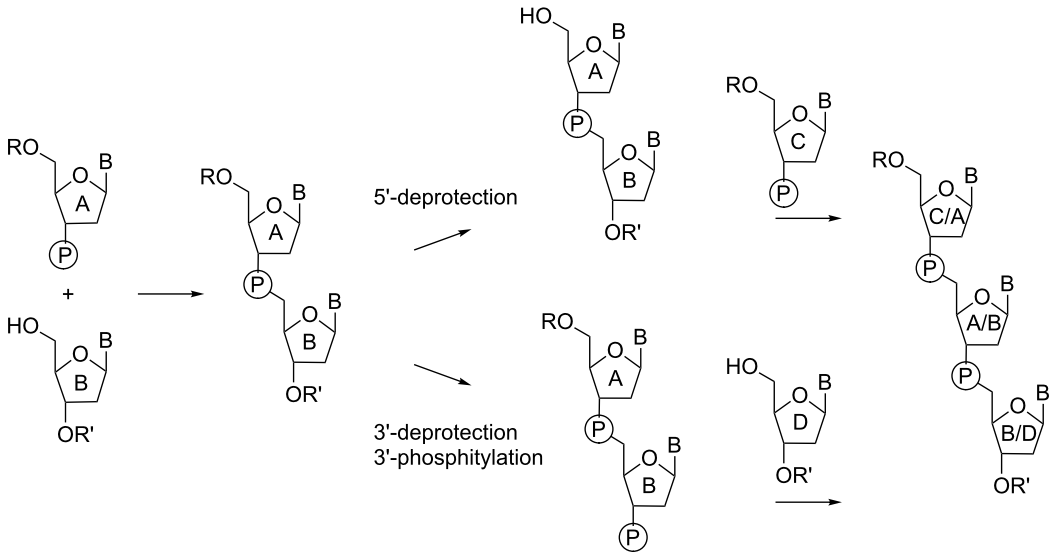


Figure 3: Strategy for the synthesis of nucleotide dimers and extension to the trimer in either 5'- or 3'-direction.

nucleotides. Both protecting groups, under the applied conditions, can be efficiently cleaved, at the same time leaving all other protecting groups intact. Thus, a full set of all 20 trimers was synthesized by phosphotriester chemistry starting with the condensation of *N*-acyl-3'-*O*-(*o*-chlorophenylphosphate)nucleosides to 3'-*O*-(2-azidomethylbenzoyl)-protected nucelosides, followed by removal of the 5'-*O*-DMTr group and extension of the dimer to the trimer by coupling of another *N*-acyl-3'-*O*-(*o*-chlorophenylphosphate)nucleoside. The final removal of the 2-azidomethylbenzoyl group occurred by reduction of the azide with triphenylphosphine in aqueous dioxane and subsequent spontaneous intramolecular cyclization leading to cleavage of the ester bond and release of the free 3'-OH group [29] (Figure 4A).

Also with 3'-*O*-TBDMS-protected monomers as mentioned above, a full set of trimers representing codons of all 20 amino acids was synthesized, although using phosphite triester chemistry [27]. In this case, the synthesis started with the coupling of an *N*-acyl-5'-*O*-DMTr-protected nucleoside-3'-*O*-phosphoramidite to an *N*-acyl-3'-*O*-TBDMS-protected nucleoside, followed by oxidation of the internucleotide phosphorous. Upon cleavage of the 5'-*O*-DMTr group, the dimer was reacted with another *N*-acyl-5'-*O*-DMTr-protected nucleoside-3'-*O*-phosphoramidite to afford the trimer. The 3'-*O*-TBDMS group was selectively removed under mild conditions with trimethylamine/3HF (Figure 4B) with strict control of pH to leave the β-cyanoethyl groups at the internucleotide phosphates intact [27]. With both procedures (3'-*O*-(2-azidomethylbenzoyl) and 3'-*O*-

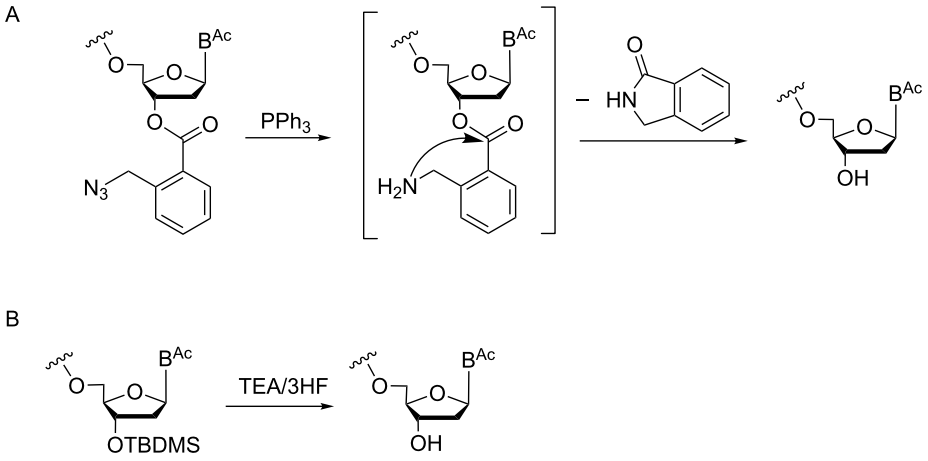


Figure 4: Removal of the 3'-O-protecting group under conditions that leave all other protecting groups at 5'-OH, nucleobases and internucleotide phosphates intact.

TBDMS protection), 20 trinucleotides of high purity were prepared and upon phosphitylation used as synthons in oligonucleotide synthesis [27,29].

In general, the reported syntheses of trinucleotides in solution proceed by either phosphite triester chemistry or phosphotriester chemistry with the latter being the more robust method. Also H-phosphonate chemistry has been used for assembling short oligomers in solution [34], although not with the aim of generating trinucleotide synthons for gene synthesis.

2. Preparation of trinucleotides on solid phase

Given the fact that trinucleotide synthesis in solution requires tedious purification and isolation of the products after each step of the synthesis, the assembly of trimers on a solid phase appears to be an attractive alternative. However, it has to be taken into account that the 3'-start nucleoside is required to be linked to the solid phase in a way that allows the cleavage of the trimer from the solid support, but leaves all other protecting groups intact. Therefore, the routinely used succinate linkage for immobilization of the start nucleotide cannot be used. Instead, linkers that allow a release of the trimers by a non-nucleophilic and/or non-basic treatment are required. In terms of trimer synthesis only one report in the literature describes such a strategy: The start nucleoside was loaded onto controlled pore glass (CPG) via an oxalyl anchor (Figure 5A), which after the synthesis was cleaved with a 5% solution of 25% aqueous ammonia in methanol, or with 20% pyridine in methanol [20].

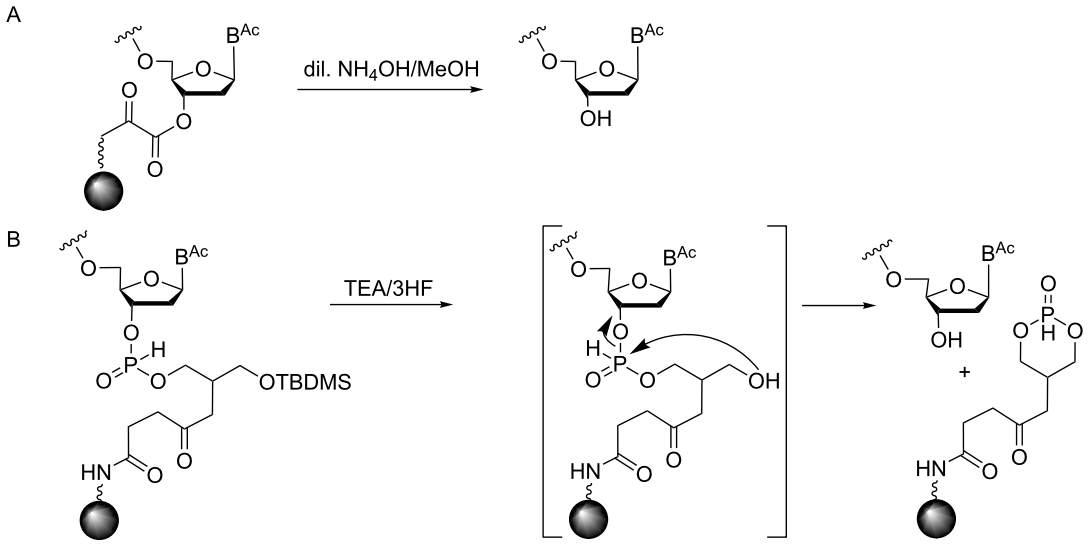


Figure 5: Release of trinucleotide blocks from the solid support by cleavage of an oxalyl anchor (A) and by a transesterification mechanism (B).

Combined with phosphotriester chemistry for trimer assembly, this treatment did not cause damage of the phosphotriester linkages and the nucleobase *N*-acyl groups. Using this strategy the large scale synthesis (5 g) of 3'-unprotected trinucleotides proceeded with a total 75–90% yield [20].

Other strategies with potential for the solid-phase synthesis of protected trinucleotides might rely on a universal solid support, from which oligomers with free 3'-OH function are released by a transesterification mechanism [35]. The 3'-start nucleoside is bound to one of the primary hydroxy groups of CPG-linked glycerol via an H-phosphonate linkage (Figure 5B). The removal of the TBDMS group from the remaining primary alcohol of glycerol induces the spontaneous cleavage of the H-phosphonate and the release of the oligomer with the free 3'-OH group leaving all other protecting groups intact. This strategy has been shown to be compatible with phosphoramidite chemistry and β-cyanoethyl protection of the internucleotide phosphates [33].

A more recent report describes the preparation of a polystyrene support decorated with a photolabile linker and its potential use for the synthesis of siRNA duplexes under mild and neutral conditions [36]. A similar strategy was used for the synthesis of partially 2'/3'-*O*-acetylated RNA oligonucleotides [37]. A photo-cleavable linker would also have potential for the synthesis of protected trinucleotides, as it would allow the cleavage of the trimer from the support by irradiation with UV light, without harming nucleobase and internucleotide phosphate protection. Nevertheless, photo-induced formation of byproducts may be an issue to be considered.

In our lab, we have been developing a strategy for solid-phase trinucleotide synthesis involving a disulfide linkage to the support (CPG or polystyrene), which can be cleaved under reductive conditions without harming nucleobase and phosphate protecting groups. The disulfide bridge is generated through the reaction of a 3'-*O*-methylthiomethyl-functionalized nucleoside with 2-mercaptopropionic acid and subsequent coupling to amino-functionalized CPG or polystyrene. After assembly of the trinucleotide on the support, the disulfide bridge is cleaved by treatment with dithiothreitol (DTT) [38] or tris-(2-carboxyethyl)phosphine (TCEP, Figure 6) leaving all other protecting groups intact.

The resulting hemi-(*S,O*)-acetal at the nucleotiode 3'-terminus is spontaneously degraded into the alcohol and thioformaldehyde, thus delivering the trimer with free 3'-OH group for subsequent phosphitylation. The detailed strategy and syntheses will be described elsewhere.

3. Preparation of trinucleotides by inverse solid-phase synthesis

Interestingly, also the use of polymer-supported reagents for H-phosphonate or phosphoramidite activation and phosphite oxidation has been described [34,39], thereby combining the advantages of solution chemistry and solid-phase methods. Thus, solid-supported acyl chloride or pyridinium tosylate as the activator of nucleoside-3'-*O*-H-phosphonates/phosphoramidites, and polystyrene-bound trimethylammonium periodate as oxidation reagent have been demonstrated to be superior for dimer and trimer synthesis, as complicated purification steps can be avoided, and excess reagents are easily removed by filtration. Compared with standard phosphotriester and phosphite triester chemistry, the limitations of this approach are lower coupling yields and side reactions hampering the yield and quality of the desired products [34,39].

4. Preparation of trinucleotides on soluble supports

Another strategy of combining the advantages of solution chemistry and solid-phase methods is the assembly of oligonucleotides on soluble supports. Among the supports used for this purpose, polyethylene glycol (PEG) has a prominent position, appearing as the routinely used polymer [40-44]. The isolation of intermediate and final products from the reaction mixture proceeds by precipitation from diethyl ether and filtration, thus significantly speeding up the process. In addition, the method is

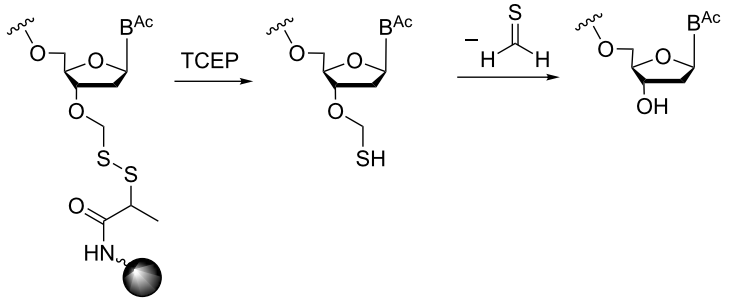


Figure 6: Release of the trinucleotide from the support under reductive conditions.

favorable in terms of producing oligonucleotides at a larger scale, since the reaction proceeds in homogeneous solution on a rather cheap polymer. The synthesis of oligonucleotides on soluble supports has been reviewed recently [45], showing that a variety of soluble polymers and precipitative supports are well suited to it. Also the solution-phase synthesis of protected trinucleotide building blocks has been described in the literature [21-23]. In an initial attempt, thymidine as a start nucleoside was tethered to a precipitative tetrapodal soluble support via a disulfide-linker [21] (Table 1, entry 1).

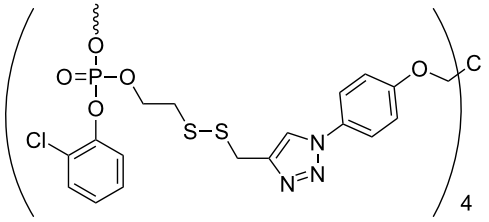
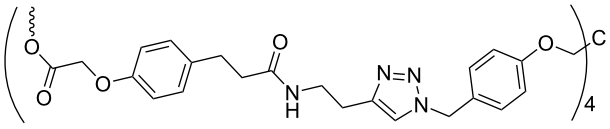
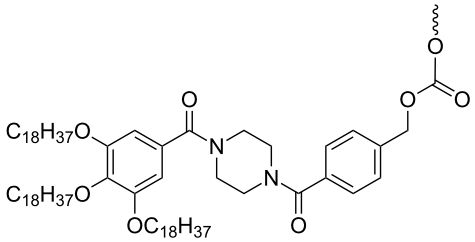
Upon detritylation, the support carrying the start nucleoside now having a free 5'-OH group was precipitated from methanol, followed by coupling with a 5'-*O*-DMTr-protected nucleoside-3'-*O*-(*o*-chlorophenyl)phosphate activated as benzotriazol and renewed precipitation with methanol. The resulting dimer was then extended to the trimer by another cycle of detritylation, precipitation, coupling and precipitation. During reductive cleavage of the disulfide bond to release the fully protected trimer from the support, unfortunately the loss of the 5'-DMTr group was observed. To overcome this hurdle, the disulfide tether was replaced in a following-up study with a Q-linker (hydroquinone-*O,O'*-diacetic acid), to be cleaved with dilute methanolic K_2CO_3 for the release of trimers in fully protected form. Five different trimers were assembled at 0.5 mmol scale and released form the support as described [22] (Table 1, entry 2). Thus, the fully protected trinucleotide building blocks were obtained with 65 to 70% yield from three coupling cycles, each containing two precipitations.

Yet another method for the synthesis of oligonucleotide blocks has been developed using a Cbz-type alkyl-chain-soluble support [23]. The support was attached via the benzyloxycarbonyl (Cbz) group to the 3'-OH of the starting nucleoside being adenosine, cytidine, guanosine or thymidine, and trimers were assembled by phosphoramidite chemistry (Table 1, entry 3). The support was found to disperse homogenously in the reaction solvents and to precipitate upon the addition of a polar solvent, typically methanol. After coupling of a standard phosphoramidite building block followed by oxidation with 2-butanone peroxide in dichloromethane, the resulting dimer on the support was again precipitated with methanol and filtered, before detritylation and coupling of the third monomer. The release of the trimer in fully protected form from the support was achieved by hydrogenation with Pd/C (10%) in tetrahydrofurane (THF) for 40 h at room temperature. Three fully protected trimers were prepared this way with isolated yields in the range of 44 to 49% [23].

5. Phosphitylation and coupling of trinucleotide synthons in solid phase DNA synthesis

To be used as building blocks in standard phosphoramidite synthesis, fully protected trimers need to be converted in phosphoramidites (Figure 7).

Table 1: Assembly of trimers on soluble supports.

entry	soluble support	5'-O-PG	chemistry	release conditions
1		H	phosphotriester	TCEP, NEt_3, MeOH, 3 h, 57%
2		DMTr	phosphotriester	K_2CO_3,DCM/MeOH/ dioxane, 30 min, 88–99%
3		DMTr	phosphoramidite	H_2/Pd, THF, 40 h, 44–49%

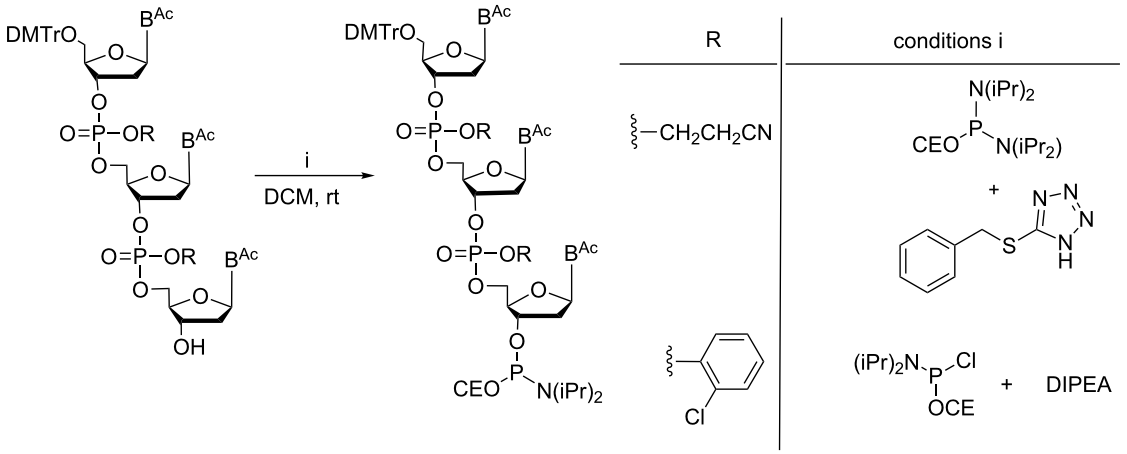


Figure 7: Phosphitylation of trimers. Reaction conditions, in particular the choice of the phosphitylation reagent, are dependent on the nature of the protecting group at the internucleotide phosphates.

This has been described in a number of reports [19,22,27,29], and is easily achieved with trimers having *o*-chlorophenyl groups for protection of the phosphate moiety [22,29]. However, phosphitylation becomes a crucial step, if β-cyanoethyl is used as the phosphate protecting group [27]. Using 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite for phosphitylation requires the presence of *N,N*-diisopropylethylamine (DIPEA) to neutralize HCl that is generated during the reaction. This, however, would lead to the removal of the β-cyanoethyl group at the phosphate moieties, which, due to the phosphorous atom in the oxidized state, is highly sensitive to basic agents and readily undergoes β-elimination [27].

An alternative reagent is 2-cyanoethyl-*N,N,N′,N′*-tetraisopropylphosphordiamidite in combination with tetrazole derivatives such as benzylmercaptotetrazole. Under those conditions, the phosphitylation proceeds with the production of one equivalent of diisopropylamine, which is neutralized by benzylmercaptotetrazole released back after the reaction. The tetrazole derivative is sufficiently acidic to act as a scavenger for diisopropylamine converting it into the ammonium salt. Thus, fully protected trimers can be converted to phosphoramidites without the loss of the β-cyanoethyl groups at the internucleotide phosphate linkages [27].

For the use in standard oligonucleotide synthesis, trinucleotide phosphoramidites have been dissolved in a mixture of acetonitrile and dichloromethane to a concentration of 0.1–0.15 M. The coupling yields are typically between 70–95%, preferentially with double or triple couplings, and a coupling time of 120 to 300 s [22,27,29].

Conclusion

The synthesis of fully protected trimers can be achieved in solution, on a solid phase or on soluble supports. The key element is the choice of a suitable set of orthogonal protecting groups to allow the selective deprotection of the functionality required for the reaction, while leaving all other protecting groups intact. The first trinucleotide synthesis was performed in solution using phosphotriester or phosphoramidite chemistry. More recently strategies for trimer assembly on a solid phase or soluble supports have been developed. Here, release of the synthesized trimer in fully protected form from the support is the crucial step. This has been convincingly achieved by using molecular entities linking the trimer to the support, which can be selectively cleaved either under reductive conditions (disulfide cleavage or hydrogenation) or under mild basic conditions leaving all protecting groups at the trimer undamaged.

In particular, soluble support strategies have great potential for an efficient large scale synthesis of fully protected trinucleotides. The essential feature here is that small molecular reagents can be easily removed after coupling and 5'-*O*-deprotection, by quantitative precipitation of the soluble support in a polar solvent, such as methanol.

With the developments in the field of biotechnology and protein engineering, the preparation of gene libraries has become a major issue. In this regard, the use of trinucleotide synthons for codon-based gene synthesis has high potential, as it allows the fully controlled total or partial randomization at any predefined number and position of codons of a given gene. Methods for their large scale preparation are available now.

ORCID® iDs

Sabine Müller - https://orcid.org/0000-0002-8385-7274

References

1. Chen, Z.; Zeng, A.-P. *Curr. Opin. Biotechnol.* **2016,** *42,* 198–205. doi:10.1016/j.copbio.2016.07.007
2. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. *Nature* **2012,** *485,* 185–194. doi:10.1038/nature11117
3. Truppo, M. D. *ACS Med. Chem. Lett.* **2017,** *8,* 476–480. doi:10.1021/acsmedchemlett.7b00114
4. Ernst, P.; Plückthun, A. *Biol. Chem.* **2017,** *398,* 23–29. doi:10.1515/hsz-2016-0233
5. Porter, J. L.; Rusli, R. A.; Ollis, D. L. *ChemBioChem* **2016,** *17,* 197–203. doi:10.1002/cbic.201500280
6. Packer, M. S.; Liu, D. R. *Nat. Rev. Genet.* **2015,** *16,* 379–394. doi:10.1038/nrg3927
7. Sullivan, B.; Walton, A. Z.; Stewart, J. D. *Enzyme Microb. Technol.* **2013,** *53,* 70–77. doi:10.1016/j.enzmictec.2013.02.012
8. Sauer, N. J.; Mozoruk, J.; Miller, R. B.; Warburg, Z. J.; Walker, K. A.; Beetham, P. R.; Schöpke, C. R.; Gocal, G. F. *Plant Biotechnol. J.* **2016,** *14,* 496–502. doi:10.1111/pbi.12496
9. Derbyshire, K. M.; Salvo, J. J.; Grindley, N. D. *Gene* **1986,** *46,* 145–152. doi:10.1016/0378-1119(86)90398-7
10. Isalan, M. *Nat. Protoc.* **2006,** *1,* 468–475. doi:10.1038/nprot.2006.68
11. Stemmer, W. P. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91,* 10747–10751. doi:10.1073/pnas.91.22.10747
12. Cárcamo, E.; Roldán-Salgado, A.; Osuna, J.; Bello-Sanmartin, I.; Yáñez, J. A.; Saab-Rincón, G.; Viadiu, H.; Gaytán, P. *ACS Omega* **2017,** *2,* 3183–3191. doi:10.1021/acsomega.7b00508
13. Dale, S. J.; Belfield, M.; Richardson, T. C. *Methods* **1991,** *3,* 145–153. doi:10.1016/S1046-2023(05)80167-7
14. Gaytán, P.; Yáñez, J.; Sánchez, F.; Soberón, X. *Nucleic Acids Res.* **2001,** *29,* E9. doi:10.1093/nar/29.3.e9
15. Arkin, A. P.; Youvan, D. C. *BioTechnology* **1992,** *10,* 297–300. doi:10.1038/nbt0392-297
16. Jensen, L. J.; Andersen, K. V.; Svendsen, A.; Kretzschmar, T. *Nucleic Acids Res.* **1998,** *26,* 697–702. doi:10.1093/nar/26.3.697
17. Tomandl, D.; Schober, A.; Schwienhorst, A. *J. Comput.-Aided Mol. Des.* **1997,** *11,* 29–38. doi:10.1023/A:1008071310472
18. Popova, B.; Schubert, S.; Bulla, I.; Buchwald, D.; Kramer, W. *PLoS One* **2015,** *10,* e0136778. doi:10.1371/journal.pone.0136778
19. Arunachalam, T. S.; Wichert, C.; Appel, B.; Müller, S. *Org. Biomol. Chem.* **2012,** *10,* 4641–4650. doi:10.1039/c2ob25328c
20. Kayushin, A.; Korosteleva, M.; Miroshnikov, A. *Nucleosides, Nucleotides Nucleic Acids* **2000,** *19,* 1967–1976. doi:10.1080/15257770008045471
21. Jabgunde, A. M.; Molina, A. G.; Virta, P.; Lönnberg, H. *Beilstein J. Org. Chem.* **2015,** *11,* 1553–1560. doi:10.3762/bjoc.11.171
22. Kungurtsev, V.; Lonnberg, H.; Virta, P. *RSC Adv.* **2016,** *6,* 105428–105432. doi:10.1039/C6RA22316H
23. Matsuno, Y.; Shoji, T.; Kim, S.; Chiba, K. *Org. Lett.* **2016,** *18,* 800–803. doi:10.1021/acs.orglett.6b00077
24. Raetz, R.; Appel, B.; Müller, S. *Chim. Oggi* **2016,** *34,* 14–17.
25. Virnekas, B.; Ge, L.; Plückthun, A.; Schneider, K. C.; Wellnhofer, G.; Moroney, S. E. *Nucleic Acids Res.* **1994,** *22,* 5600–5607. doi:10.1093/nar/22.25.5600
26. Yáñez, J.; Argüello, M.; Osuna, J.; Soberón, X.; Gaytán, P. *Nucleic Acids Res.* **2004,** *32,* e158. doi:10.1093/nar/gnh156
27. Janczyk, M.; Appel, B.; Springstubbe, D.; Fritz, H.-J.; Müller, S. *Org. Biomol. Chem.* **2012,** *10,* 1510–1513. doi:10.1039/c2ob06934b
28. Zehl, A.; Starke, A.; Cech, D.; Hartsch, T.; Merkl, R.; Fritz, H.-J. *Chem. Commun.* **1996,** 2677–2678. doi:10.1039/cc9960002677
29. Yagodkin, A.; Azhayev, A.; Roivainen, J.; Antopolsky, M.; Kayushin, A.; Korosteleva, M.; Miroshnikov, A.; Randolph, J.; Mackie, H. *Nucleosides, Nucleotides Nucleic Acids* **2007,** *26,* 473–497. doi:10.1080/15257770701426260
30. Lyttle, M. H.; Napolitano, E. W.; Calio, B. L.; Kauvar, L. M. *BioTechniques* **1995,** *19,* 274–281.
31. Ono, A.; Matsuda, A.; Zhao, J.; Santi, D. V. *Nucleic Acids Res.* **1995,** *23,* 4677–4682. doi:10.1093/nar/23.22.4677
32. Kayushin, A. L.; Korosteleva, M. D.; Miroshnikov, A. I.; Kosch, W.; Zubov, D.; Piel, N. *Nucleic Acids Res.* **1996,** *24,* 3748–3755. doi:10.1093/nar/24.19.3748
33. Sondek, J.; Shortle, D. *Proc. Natl. Acad. Sci. U. S. A.* **1992,** *89,* 3581–3585. doi:10.1073/pnas.89.8.3581
34. Adamo, I.; Dueymes, U.; Schönberger, A.; Navarro, A.-E.; Meyer, A.; Lange, M.; Imbach, J.-L.; Link, F.; Morvan, F.; Vasseur, J.-J. *Eur. J. Org. Chem.* **2006,** 436–448. doi:10.1002/ejoc.200500547
35. Ferreira, F.; Meyer, A.; Vasseur, J.-J.; Morvan, F. *J. Org. Chem.* **2005,** *70,* 9198–9206. doi:10.1021/jo051172n
36. Lietard, J.; Hassler, M. R.; Fakhoury, J.; Damha, M. J. *Chem. Commun.* **2014,** *50,* 15063–15066. doi:10.1039/C4CC07153K
37. Xu, J.; Duffy, C. D.; Chan, C. K. W.; Sutherland, J. D. *J. Org. Chem.* **2014,** *79,* 3311–3326. doi:10.1021/jo5002824
38. Semenyuk, A.; Földesi, A.; Johansson, T.; Estmer-Nilsson, C.; Blomgren, P.; Brännvall, M.; Kirsebom, L. A.; Kwiatkowski, M. *J. Am. Chem. Soc.* **2006,** *128,* 12356–12357. doi:10.1021/ja0636587
39. Dueymes, C.; Schönberger, A.; Adamo, I.; Navarro, A.-E.; Meyer, A.; Lange, M.; Imbach, J.-L.; Link, F.; Morvan, F.; Vasseur, J.-J. *Org. Lett.* **2005,** *7,* 3485–3488. doi:10.1021/ol0511777
40. Bonora, G. M.; Scremin, C. L.; Colonna, F. P.; Garbesi, A. *Nucleic Acids Res.* **1990,** *18,* 3155–3159. doi:10.1093/nar/18.11.3155
41. Bonora, G. M.; Biancotto, G.; Maffini, M.; Scremin, C. L. *Nucleic Acids Res.* **1993,** *21,* 1213–1217. doi:10.1093/nar/21.5.1213
42. Bonora, G. M.; De Franco, A. M.; Rossin, R.; Veronese, F. M.; Ferruti, P.; Plyasunova, O.; Vorobjev, P. E.; Pyshnyi, D. V.; Komarova, N. I.; Zarytova, V. F. *Nucleosides, Nucleotides Nucleic Acids* **2000,** *19,* 1281–1288. doi:10.1080/15257770008033051
43. Padiya, K. J.; Salunkhe, M. M. *Bioorg. Med. Chem.* **2000,** *8,* 337–342. doi:10.1016/S0968-0896(99)00287-4
44. Ballico, M.; Drioli, S.; Morvan, F.; Xodo, L.; Bonora, G. M. *Bioconjugate Chem.* **2001,** *12,* 719–725. doi:10.1021/bc010034b
45. Lönnberg, H. *Beilstein J. Org. Chem.* **2017,** *13,* 1368–1387. doi:10.3762/bjoc.13.134

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Stimuli-responsive oligonucleotides in prodrug-based approaches for gene silencing

Françoise Debart*, Christelle Dupouy and Jean-Jacques Vasseur*

Review

Open Access

Address:
IBMM, Université de Montpellier, CNRS, ENSCM, Montpellier, France

Email:
Françoise Debart* - francoise.debart@umontpellier.fr;
Jean-Jacques Vasseur* - jean-jacques.vasseur@umontpellier.fr

* Corresponding author

Keywords:
enzymolabile group; light-responsive group; oligonucleotide prodrugs; reduction-responsive; stimuli-responsive nucleic acids; thermolytic prodrugs

Beilstein J. Org. Chem. **2018,** *14,* 436–469.
doi:10.3762/bjoc.14.32

Received: 16 November 2017
Accepted: 26 January 2018
Published: 19 February 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

Oligonucleotides (ONs) have been envisaged for therapeutic applications for more than thirty years. However, their broad use requires overcoming several hurdles such as instability in biological fluids, low cell penetration, limited tissue distribution, and off-target effects. With this aim, many chemical modifications have been introduced into ONs definitively as a means of modifying and better improving their properties as gene silencing agents and some of them have been successful. Moreover, in the search for an alternative way to make efficient ON-based drugs, the general concept of prodrugs was applied to the oligonucleotide field. A prodrug is defined as a compound that undergoes transformations in vivo to yield the parent active drug under different stimuli. The interest in stimuli-responsive ONs for gene silencing functions has been notable in recent years. The ON prodrug strategies usually help to overcome limitations of natural ONs due to their low metabolic stability and poor delivery. Nevertheless, compared to permanent ON modifications, transient modifications in prodrugs offer the opportunity to regulate ON activity as a function of stimuli acting as switches. Generally, the ON prodrug is not active until it is triggered to release an unmodified ON. However, as it will be described in some examples, the opposite effect can be sought.

This review examines ON modifications in response to various stimuli. These stimuli may be internal or external to the cell, chemical (glutathione), biochemical (enzymes), or physical (heat, light). For each stimulus, the discussion has been separated into sections corresponding to the site of the modification in the nucleotide: the internucleosidic phosphate, the nucleobase, the sugar or the extremities of ONs. Moreover, the review provides a current and detailed account of stimuli-responsive ONs with the main goal of gene silencing. However, for some stimuli-responsive ONs reported in this review, no application for controlling gene expres-

sion has been shown, but a certain potential in this field could be demonstrated. Additionally, other applications in different domains have been mentioned to extend the interest in such molecules.

Introduction

For past decades, oligonucleotide-based therapies have been widely developed using short synthetic oligonucleotides (ONs) and their chemically modified mimics as powerful tools to block mRNA function, inhibit protein function or induce an immune response [1,2]. Among these ON therapeutic strategies, ON-based gene silencing, which involves mRNAs as specific targets, has been largely investigated, and several promising ONs have been under clinical development [3]. Gene silencing strategies include antisense oligonucleotides (AONs), ribozymes, DNAzymes, small interfering RNAs (siRNAs) and micro RNAs (miRNAs) that specifically target the complementary mRNA sequence of the relevant undesired gene before translation.

AONs are single-stranded DNA of 15 to 25 nucleotides in length that bind to mRNA targets through Watson–Crick base pairing and form a RNA/DNA duplex [4]. This can result in either mRNA cleavage mediated by RNase H or mRNA translational arrest through steric blocking. Another strategy for gene inhibition involves ribozymes [5] and DNAzymes [6], which are nucleic acid molecules with enzymatic activity. These catalytic RNAs and DNAs trigger the cleavage of RNA substrates at a specific position. Additionally, ribozymes can catalyze the ligation of target mRNA, extending their therapeutic potential to RNA repair applications. Finally, another promising ON-based therapy, more potent than AONs or ribozymes for gene knockdown, is centered on the RNA interference (RNAi) mechanism, which uses two natural pathways for gene silencing. One is guided by double-stranded siRNAs of 19–23 nucleotides in length that are fully complementary to the mRNA targets, and the other is guided by miRNAs (22 nucleotides in length) that bind incorrectly within the 3'-untranslated region (3'-UTR) of the target mRNAs [7]. miRNAs also represent interesting targets, and inhibition of their function was obtained using anti-miRNA AONs via an antisense approach or via the blocking of the mRNA binding site (miRNA masking) [8].

Although many ONs are under investigation for clinical use, several hurdles remain to be overcome for the exploitation of ONs as therapeutic compounds. Among the major limitations of unmodified ONs, poor stability in vivo, low delivery and lack of specificity to target cells or tissues, off-target effects and toxicity hamper the path to success of ON-based therapeutics and need to be solved. Fortunately, various chemical modifications of ONs have been designed to address these issues [9]. The most common modification in AONs and siRNAs is the phosphorothioate (PS) backbone in the replacement of the phosphate ester internucleotide linkages. This modification provides nuclease stability and favorable pharmacokinetic properties but can lead to some toxicity. In addition, the most extensively used sugar modifications are represented by the 2'-modifications: 2'-*O*-methyl (2'-OMe), 2'-fluoro (2'-F), and 2'-*O*-(2-methoxyethyl) (MOE) [9,10]. Some examples of the combination of 2'-OMe and 2'-F modified nucleotides in siRNAs were reported, and the potency of the modified siRNA was increased compared to unmodified siRNA. Many chemical modifications have been introduced in ONs definitively as a means of modifying and better improving their properties as gene silencing agents [11]. However, an alternative way to make efficient ON-based drugs is to apply the general concept of prodrugs to the oligonucleotide field. Based on the definition of a prodrug given by Albert in 1958 [12], a prodrug is an agent that undergoes chemical or enzymatic transformations in vivo to yield the active parent drug. The prodrug approach is used to optimize the physicochemical properties of the drug and to improve its pharmacological and toxicological profile.

Oligonucleotide prodrugs that could be defined as caged oligonucleotides are transiently modified ONs with non-permanent chemical modifications (responsive units) that can be removed in response to appropriate stimuli, producing the native oligonucleotide. The aim of the prodrug strategy for nucleic acid therapeutic applications such as gene regulation is to circumvent the poor chemical stability of nucleic acids in biological media due to their low resistance to nucleases and to overcome their low cell uptake due to their polyanionic nature. In the present review, we aimed to identify various ON prodrugs that are responsive to various stimuli and evaluate their applications, mainly focusing on the control of gene expression. The use of ON prodrugs as aptamers, decoys or immunostimulatory ONs in other ON-based therapeutic strategies is marginally mentioned.

Two classes of stimuli can trigger inactive ON prodrugs in active biomolecules. Here, we summarize the chemically modified ONs that are responsive to either internal biochemical regulatory stimuli such as glutathione or enzymes (reductases, carboxyesterases), or external physical stimuli such as heat or light (photoirradiation). The transient responsive units may be

attached at different positions of the ON: the internucleotide linkage, the ribose, the nucleobase, or the 5' or 3' extremities. For simplicity, each section corresponding to one class of stimulus has been divided into sub-sections related to the site of the modification in the ON when the subject was thoroughly documented.

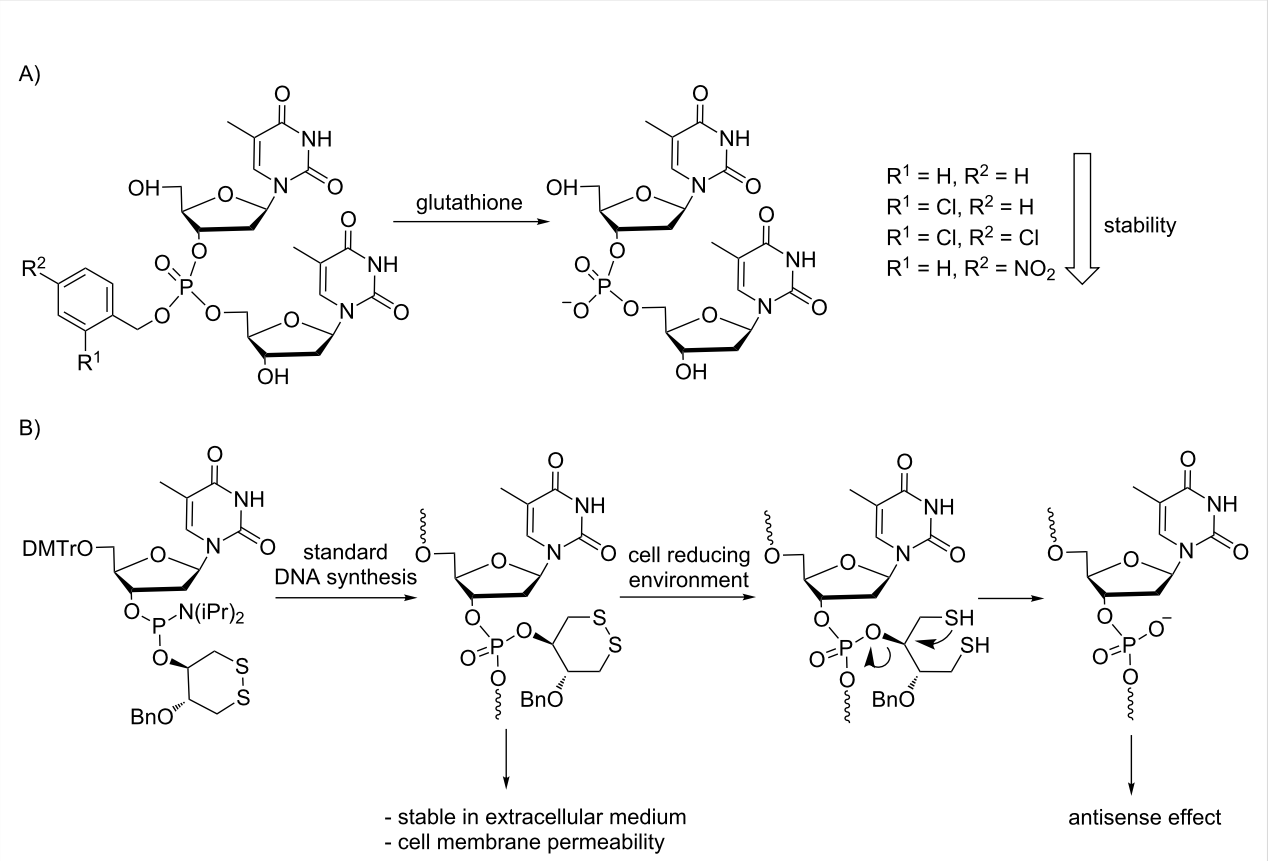
Review

Reduction-responsive ONs

These modified ONs are responsive to the reducing environment inside cells due to the natural presence of glutathione (GSH) as a conversion trigger. ONs that are responsive to the action of reductases under hypoxic conditions will be discussed vide infra in a separate section. The intracellular concentration of GSH ranges from 1 mM to 10 mM, which is 10–100 times higher than its extracellular concentration. Consequently, ON prodrugs should be stable outside the cell and, after cellular uptake, would be converted into the native ONs by intracellular abundant GSH. In this context, two classes of reduction-responsive units, disulfide-bond and benzyl-containing groups, were mainly introduced in prodrug-based ONs.

Modifications at the internucleotide linkage

Masking the negative charges of native phosphates typically improves cell penetration of the modified ONs in addition to an increase in their nuclease resistance. Thus, two Japanese groups have proposed prodrug-type phosphotriester ONs responsive to GSH (Scheme 1) [13,14]. Ono presented a preliminary study on a model of a thymidine dimer with differently substituted benzyl groups at the internucleotide linkage [13]. It was shown that the stability in aqueous buffer and deprotection rates in the presence of GSH were influenced by the nature of substituents (Cl, NO_2) on the benzene ring. More recently, Urata et al. reported a reduction-responsive modification containing a typical disulfide bond within a robust cyclic disulfide moiety [14]. Several modified ONs containing the cyclic disulfide *trans*-5-benzyl-1,2-dithiane-4-yl moiety have been synthesized using the corresponding thymidine phosphoramidite. Although they exhibited strong stability in serum and penetrated cells more efficiently, their gene silencing effects were weaker than those of PS AONs when tested using the same model assay. It seems that the conversion of the modified ONs into native ONs might occur too slowly inside cells to improve gene silencing.



Scheme 1: Demasking under reducing agents of ON prodrugs modified as phosphotriesters with A) benzyl groups [13] and B) a cyclic disulfide *trans*-5-benzyl-1,2-dithiane-4-yl moiety [14].

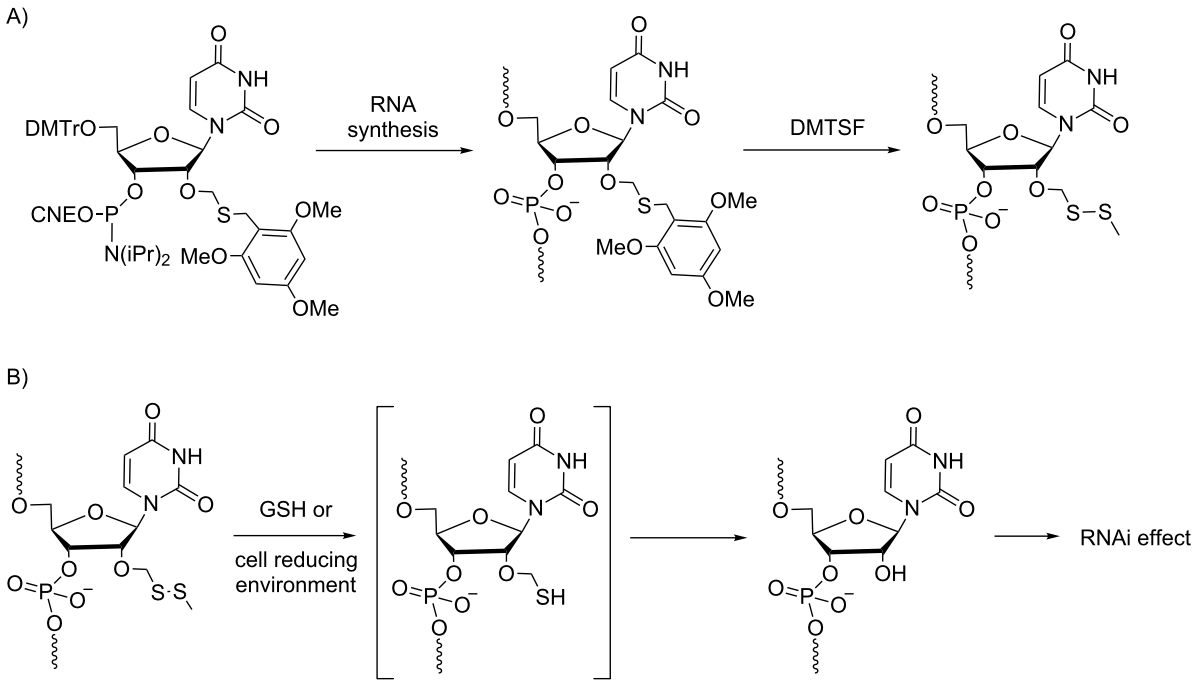
Modifications at the sugar 2'-OH

Several permanent 2'-O-modifications (2'-F, 2'-OMe) have been proposed to increase the nuclease resistance of ONs, but most of them have decreased gene silencing potential. To overcome this drawback, novel prodrug-type RNAs containing a disulfide bridge at the 2'-position have been designed, and in 2016, Urata and our group reported on the synthesis and properties of 2'-*O*-alkyldithiomethyl-modified RNAs [15,16]. Previously, Urata had described a post-synthetic approach for the synthesis of 2'-*O*-methyldithiomethyl (MDTM) ONs [17] that was more practical than the phosphoramidite approach used initially for the chemical synthesis of RNAs using the 2'-*O*-*tert*-butyldithiomethyl-protecting group [18]. In the recent approach, the MDTM modification was obtained in excellent yield after conversion of the 2,4,6-trimethoxybenzylthiomethyl precursor group by treatment with dimethyl(methylthio)sulfonium tetrafluoroborate (DMTSF, Scheme 2). First, ONs containing 2'-*O*-MDTM modifications have shown greater nuclease resistance, and they were rapidly and efficiently converted into 2'-OH ONs under reducing conditions (10 mM 1,4-dithiothreitol or 10 mM glutathione, pH 7) [17]. In a subsequent report [16], the unmasking of double-stranded 2'-*O*-MDTM siRNAs into 2'-OH siRNAs was similarly demonstrated in the presence of 10 mM GSH. Furthermore, firefly luciferase expression in A549-Luc cells was inhibited by 2'-*O*-MDTM siRNAs to a higher extent than the unmodified siRNA regardless of the modification site (5'-end and/or the seed region of the antisense strand). These results suggest that 2'-*O*-MDTM siRNAs fulfill some features of typical prodrug-type siRNAs.

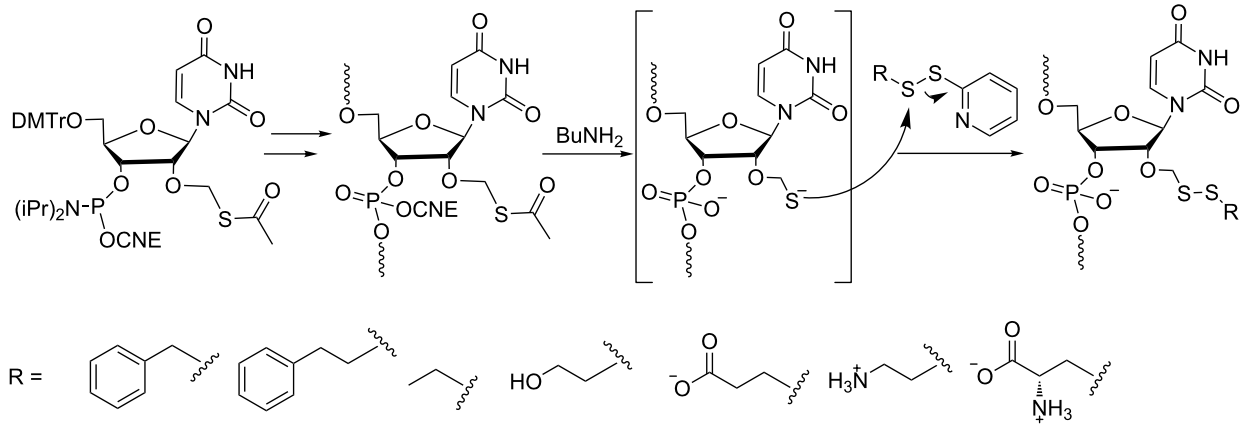
Similarly, our group has developed a post-synthetic method on a solid support to introduce various disulfide bond-containing groups at the 2'-OH of RNAs [15]. Using this versatile method, one precursor, 2'-*O*-acetylthiomethyl-containing RNA, produces various 2'-*O*-alkyldithiomethyl (RSSM)-modified RNAs bearing lipophilic or polar groups through a thiol disulfide exchange reaction with alkyldisulfanyl-pyridine derivatives (Scheme 3). In a preliminary evaluation, the RSSM modifications were shown to increase RNA resistance against 3'-exonuclease and not disturb the duplex stability too much while maintaining an A-form conformation. In addition, glutathione treatment under physiological conditions rapidly and efficiently reduced all the RSSM groups releasing 2'-OH RNA. These properties are promising for the use of 2'-*O*-RSSM-modified RNAs as prodrugs of siRNAs.

Modifications at the extremities

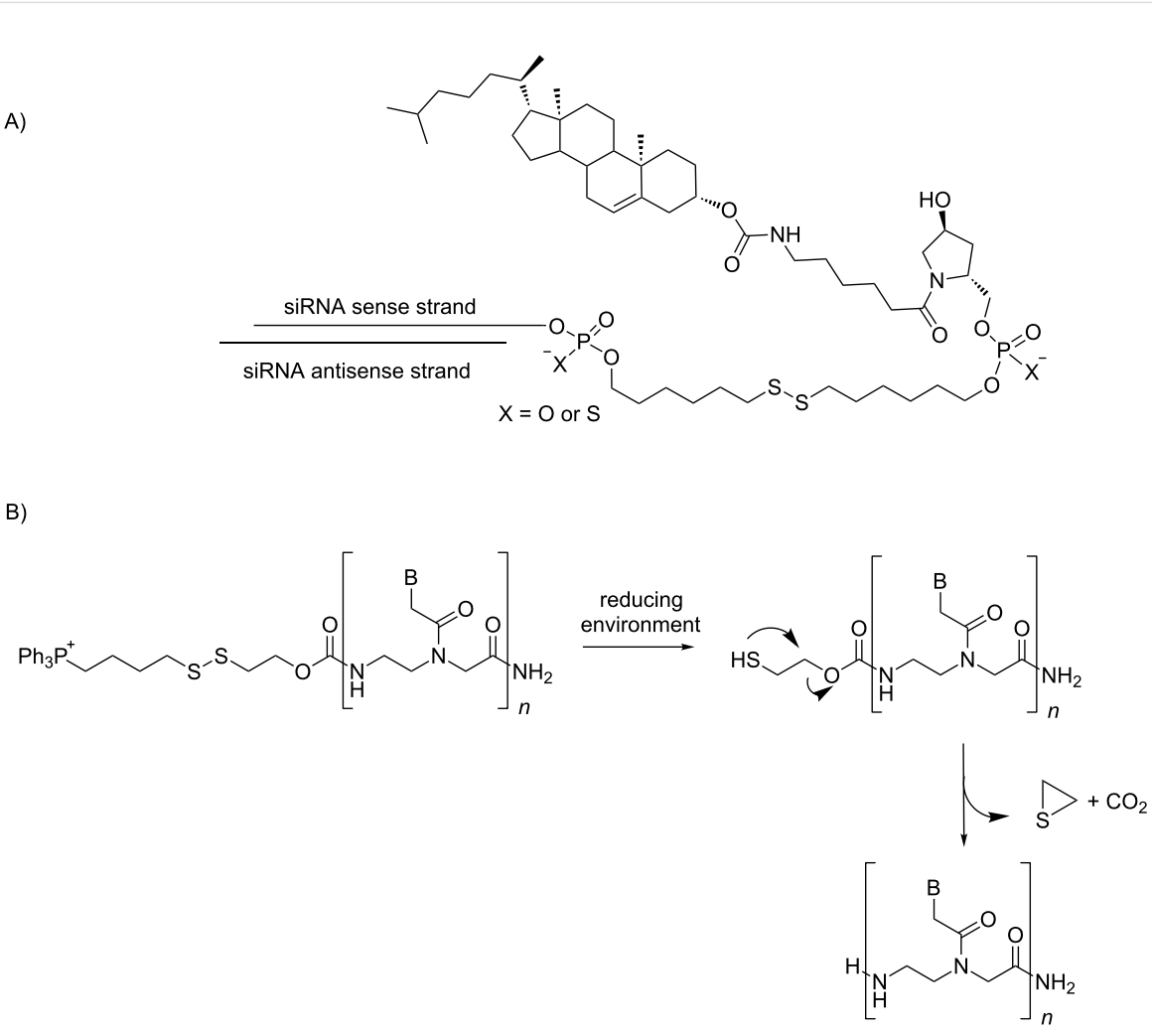
Disulfide bonds are attractive in designing drug-delivery systems. Indeed, lipophilic moieties may be attached to ONs to enhance cellular uptake. In particular, a cleavable disulfide linker has been used at the 3'-end of the sense strand to prepare cholesterol-conjugated siRNAs that were efficiently delivered to rat oligodendrocytes in vivo and achieved significant specific gene knockdown in these cells (Scheme 4A) [19]. The com-



Scheme 2: A) Synthesis via phosphoramidite chemistry and B) demasking under the reducing environment of 2'-*O*-MDTM-modified siRNA prodrugs [17].



Scheme 3: Synthesis via phosphoramidite chemistry of various 2'-*O*-alkyldithiomethyl (RSSM)-modified RNAs bearing lipophilic or polar groups (R) involving post-elongation conjugation through a thiol disulfide exchange reaction [15].



Scheme 4: A) siRNA conjugates to cholesterol [19] and B) PNA conjugates to a triphenylphosphonium [20] through a disulfide linkage.

parison with a non-cleavable alkyl linker suggests that a lipophilic siRNA conjugate with a disulfide linker is favorable to improve the suppression of 2',3'-cyclic nucleotide 3'-phosphodiesterase mRNA in oligodendrocytes in vivo. This result may be attributable to increased bioavailability of siRNA in the cytoplasm.

Similarly, regarding the intracellular delivery of naked peptide nucleic acids (PNAs), a lipophilic triphenylphosphonium (TPP) cation was attached to the N-terminal extremity of a PNA through a biodegradable carbamate linker containing a disulfide bridge (Scheme 4B) [20]. It was shown that such PNA conjugates entered cells rapidly and efficiently. Furthermore, a 16-mer PNA_{TAR} fragment directed against the TAR RNA region of the HIV genome conjugated to TPP inhibited HIV replication in CEM cell lines with an IC_{50} of 1 μM, while the unconjugated 16-mer PNA_{TAR} was inactive in these tests. The anti-HIV activity confirmed that the PNA_{TAR} was not sequestered in mitochondria; consequently, the disulfide bond was reduced into the cytoplasm.

Enzyme-responsive ONs

A control of gene expression using cellular enzymes as triggers of the activity of ON prodrugs is very attractive because this approach is based on the difference in the extra- and intracellular contents of the enzymes. Therefore, the biodegradable modification present in the prodrug could not be removed in extracellular media but only inside the cells. Two approaches have been reported using reductases or carboxyesterases to trigger transformation of ON prodrugs in native ONs. Although a post-synthesis introduction of the enzymolabile groups into phosphorothioate ONs by the reaction with alkyl iodides has been considered since the mid 90's [21-24], the use of phosphoramidite building blocks bearing the enzymocleavable group is the method of choice for synthesizing ON prodrugs regardless of the protected function (phosphate, nucleobase, sugar hydroxy groups).

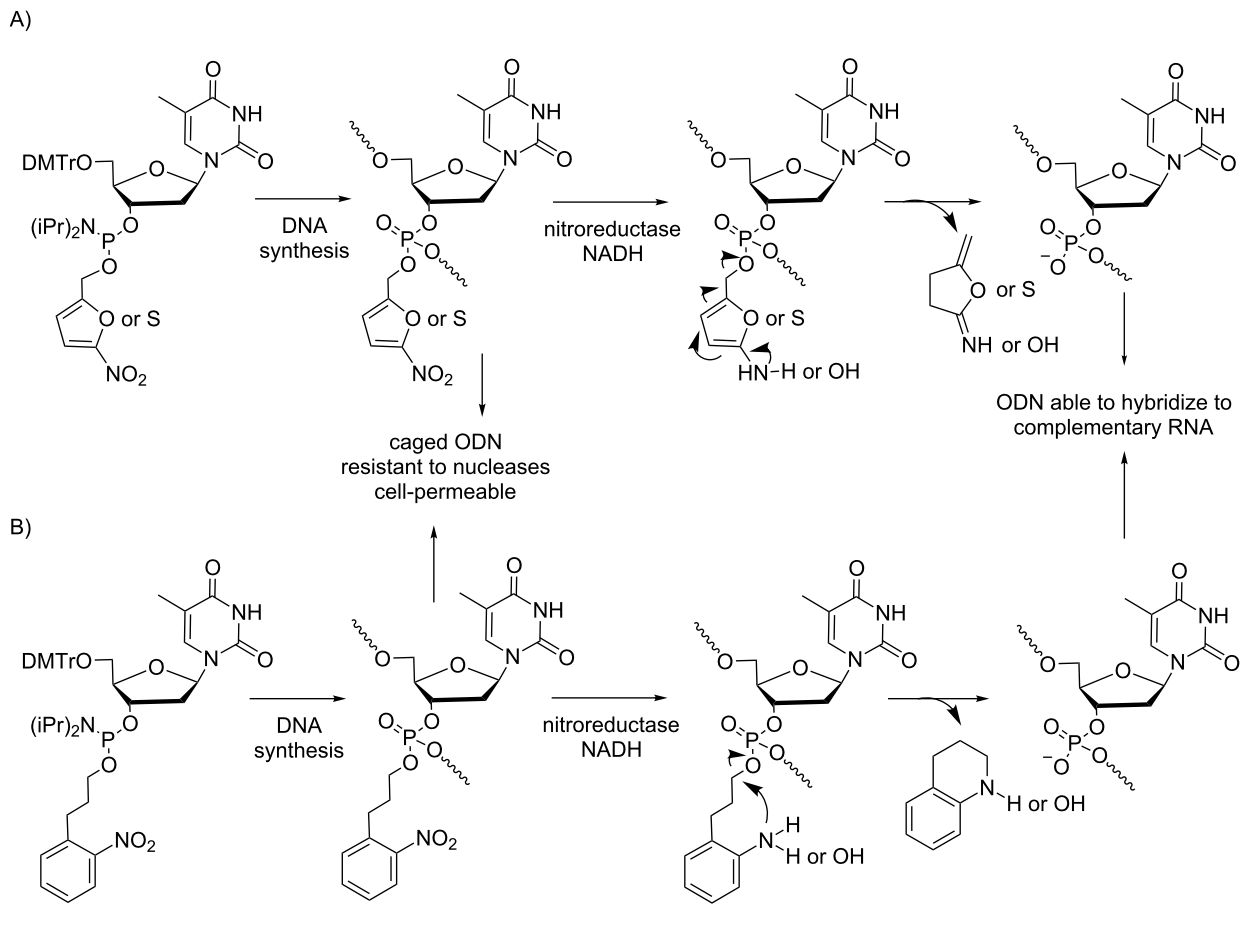
Reductase-responsive ONs

Hypoxic conditions that are characteristic of solid tumors represent a remarkable stimulus to convert non-active prodrugs into active drugs under reductase action. Three examples of hypoxia-activated ONs have been reported thus far, with a hypoxia-labile modification either in the phosphate backbone to mask the negative charge and provide better tumor selectivity [25,26] or at the nucleobase to modulate the hybridization properties with the target [27]. In all cases, a nitro-derivative-modified thymidine phosphoramidite was prepared and incorporated into oligothymidylates $(dT)_n$ or heterosequences at different sites. Actually, the nitro-derivative modifications (nitrobenzyl, nitrofuryl or nitrothienyl) can be reduced by reductases to form the corresponding amino (or hydroxylamino) derivatives, followed by a cleavage of the benzyl or heterocycle groups and release of the unmodified sequences.

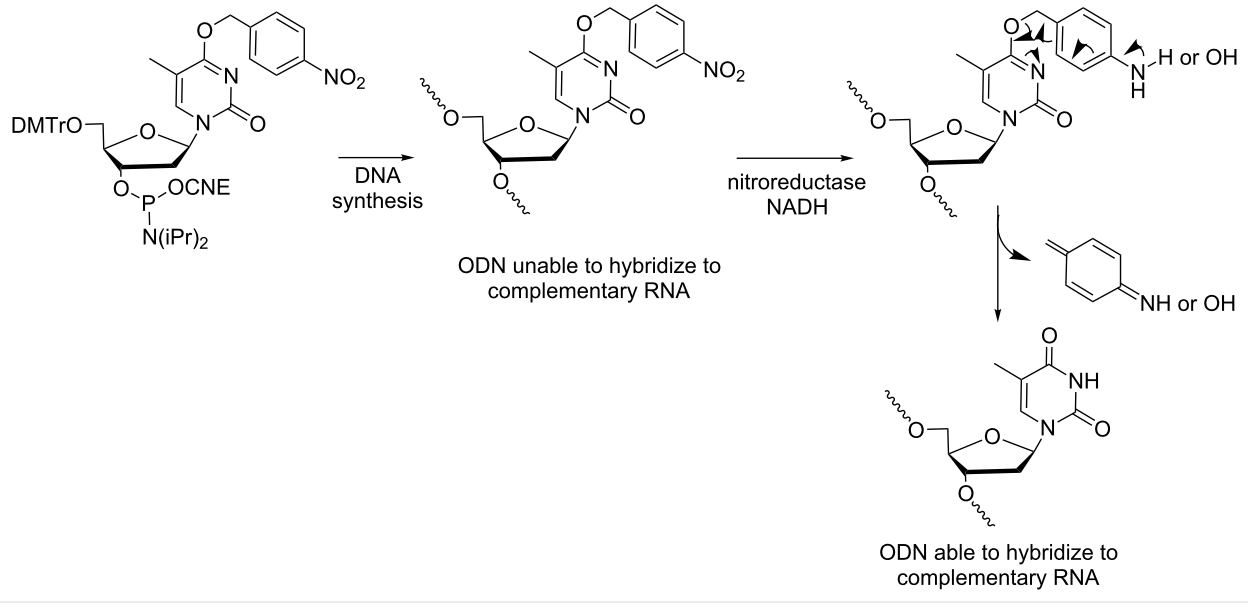
Modifications at the internucleotide linkage: ONs containing either 5-nitro-2-furylmethyl or 5-nitro-2-thiophenylmethyl modifications at some internucleoside phosphates were converted to native $(dT)_n$ with good hypoxia selectivity in vitro by nitroreductases as well as in tumor cell extract by cellular reductases (Scheme 5A) [25]. Furthermore, such nitrofuryl and nitrothienyl modifications improved nuclease resistance and cellular uptake of ONs in proportion to the number of lipophilic groups. In another study, a series of ONs with mixed sequences bearing some nitrophenylpropyl modifications were synthesized and exhibited good resistance toward nucleases and stability in human serum (Scheme 5B) [26]. Their cellular uptake in HeLa cells was greater than that of the naked ON and increased with the number of labile groups masking the phosphates. As expected, the nitrophenylpropyl groups were readily cleaved by nitroreductase in the presence of NADH. Such modified ONs could be used as prodrugs for the delivery of ON-based therapeutics in hypoxic cells.

Modifications at the nucleobase: The third example reported by Saneyoshi and Ono refers to ONs containing the hypoxia-labile group on the nucleobase. It was shown that $(dT)_5$ with one 4-nitrobenzylthymine was deprotected in vitro by nitroreductase in the presence of NADH to produce $(dT)_5$ with native thymine (Scheme 6) [27]. In addition, thermal stabilities of the duplexes formed with thymine-modified ONs and their complementary sequences were evaluated; the nucleobase modifications induced an important destabilization of the duplexes. This result suggests that 4-NO_2-benzylthymine-modified ONs cannot hybridize to their targets and consequently should be inactive in normal cells. However, in hypoxic cells after removal of the 4-nitrobenzyl groups, the resulting native ONs should form stable active duplexes with their targets. These hypoxia-labile modifications seem promising for the development of ON therapeutics with specific activity in hypoxic tumor cells and low toxicity in normal cells.

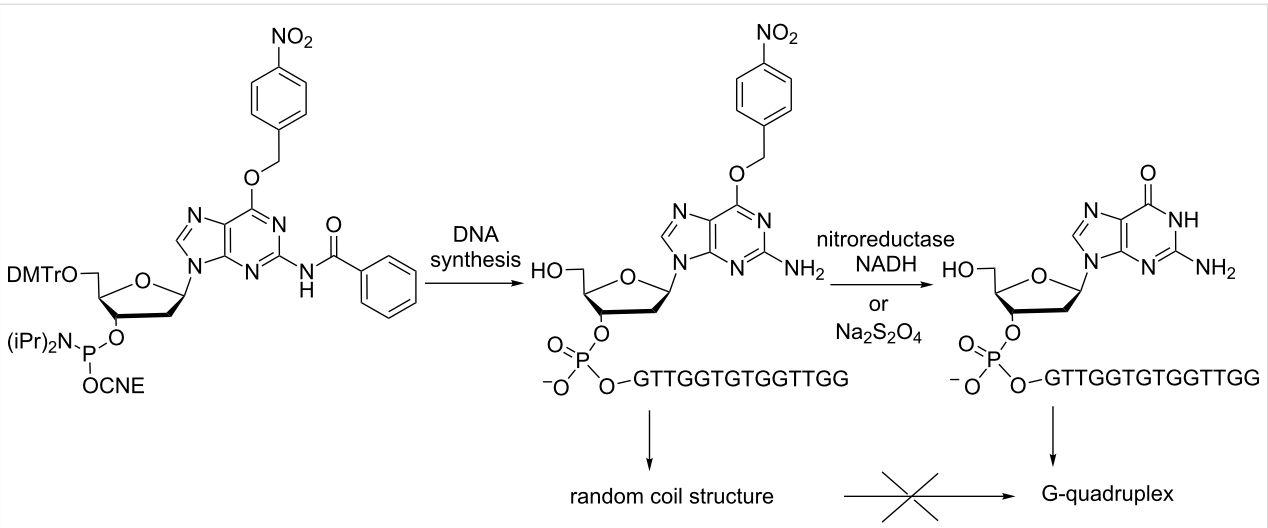
A nitrobenzyl (NB) group has also been introduced at O6 of a guanine to modulate the conformational properties of a G-quadruplex structure-forming single-stranded DNA [28]. The dG^{NB} phosphoramidite was synthesized and incorporated into the sequence of a thrombin-binding DNA aptamer (TBA, at the 5'-end) prone to form a G-quadruplex structure (Scheme 7). Circular dichroism studies have indicated that TBA^{NB} adopts a random coil structure while after reduction caused by chemical ($Na_2S_2O_4$) or enzymatic (nitroreductase with NADH) stimuli, the formation of a G-quadruplex structure was evidenced due to



Scheme 5: Synthesis via phosphoramidite chemistry and deprotection mediated by nitroreductase/NADH of hypoxia-activated prodrugs of ONs containing A) 5-nitro-2-furylmethyl or 5-nitro-2-thiophenylmethyl [25] and B) 3-(2-nitrophenylpropyl)phosphotriester internucleoside linkages [26].



Scheme 6: Synthesis via phosphoramidite chemistry and conversion mediated by nitroreductase/NADH of hypoxia-activated prodrugs of ONs containing O^4-(4-nitrobenzyl)thymidine [27].

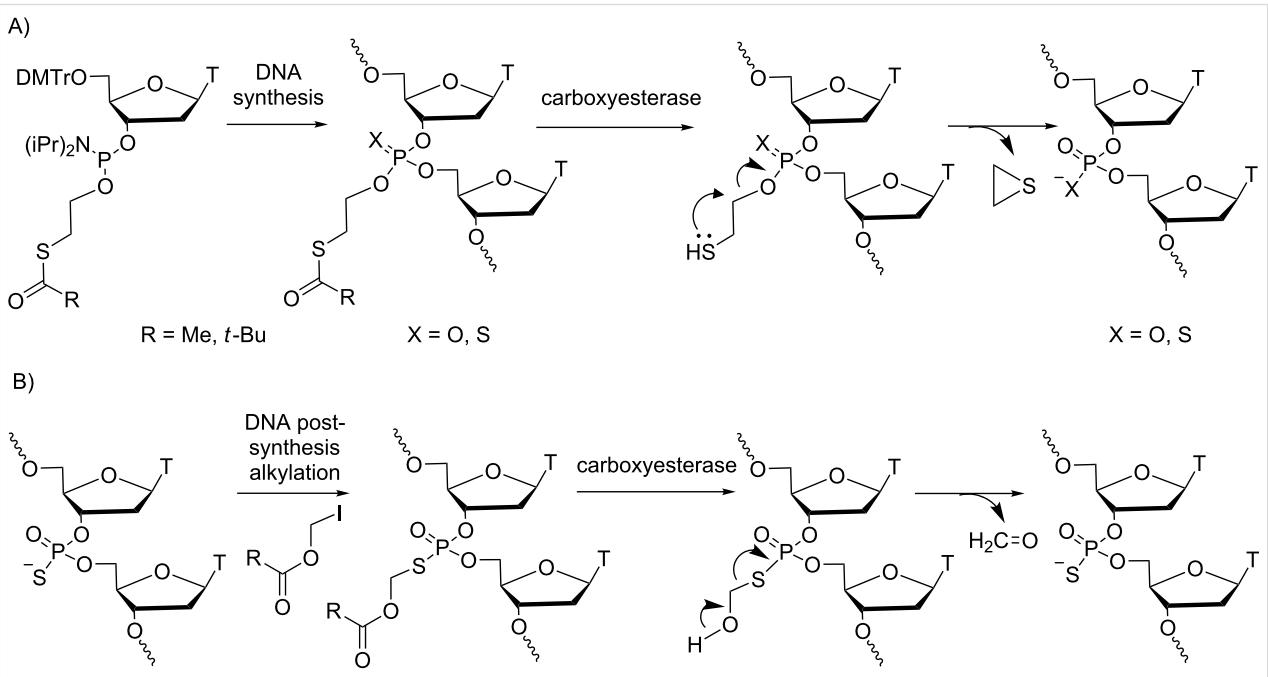


Scheme 7: Incorporation of O^6-(4-nitrobenzyl)-2'-deoxyguanosine into an ON prone to form a G-quadruplex structure, preventing it from forming this quadruplex when protected and allowing it under reducing conditions [28].

the conversion of TBA^{NB} into TBA. The modulation of the secondary structure transition of an ON in a reduction-responsive manner appears to be beneficial to understand biomolecule behavior and biological phenomena.

Esterase-responsive ONs

Modifications at the internucleotide linkage: The use of phosphate modifications cleaved under carboxyesterase mediation was envisaged for ONs more than 20 years ago and was extensively studied by Imbach's group [29] and others [22,30,31]. Ten years ago, Lönnberg summarized the chemical aspects of prodrug strategies at the nucleotide and oligonucleotide levels and particularly focused on esterase-responsive modified-phosphate ONs [32]. The most studied masking groups have been the methyl-SATE (*S*-acetylthioethyl) and *tert*-butyl SATE (*S*-pivaloylthioethyl) developed by Imbach (Scheme 8A) [29], whereas *S*-acyloxymethyl groups were studied by Agrawal (Scheme 8B) [22]. The fundamental advantage of using en-



Scheme 8: Synthesis and mechanism for the demasking of ON prodrugs from A) *S*-acylthioethyl phosphotriester [29] and B) *S*-acyloxymethyl phosphotriester [22].

zyme-cleavable modifications of the phosphodiester backbone in ONs is to transitorily mask the negative charges of the phosphate by neutral phosphotriesters. Consequently, the backbone is less prone to nuclease degradation, and the lipophilicity of the pro-ON increases cell permeation [33]. The uptake was proportional to the number of SATE groups and probably proceeded through a passive diffusion mechanism [34]. Furthermore, it was shown that SATE-protected phosphates were selectively demasked in cell extracts [35-37]. SATE thionophosphotriester ONs were quantitatively converted to phosphorothioate ONs by carboxyesterase-mediated deacylation followed by the removal of the resulting *S*-(2-mercaptoethyl) group by cyclization to episulfide. For *S*-acyloxymethyl phosphorothiolates, hydrolysis of the ester catalyzed by the enzymes was followed by release of formaldehyde to produce the phosphorothioate ON.

Despite these promising results, further studies on the use of these prodrugs to control genetic expression have not been carried out. Thus far, most of these results were obtained for thymidine homopolymers [32]. The reason is that the synthesis of ON prodrugs is incompatible with the standard deprotection treatment under basic conditions (generally aqueous ammonia) used to cleave other common base-labile acyl protection groups from nucleobases and release ON from the solid support. Furthermore, as the aqueous solubility of fully modified SATE phosphotriester ONs is rather poor [29], the design of ONs combining phosphodiester and phosphotriester linkages is required to ensure aqueous solubility and sufficient lipophilicity for cell uptake. Several attempts to obtain such chimeras were made in Imbach's laboratory in the early 2000s. In particular, the use of photolabile protecting groups [38] of allyloxycarbonyl groups deprotected by Pd(0) [39] and of fluoride-labile groups [40] in place of the standard acyl protection of nucleobases has made possible the acquisition of short sequences of heteropolymer pro-oligonucleotides. However, none of these methods led to ON prodrugs of therapeutic interest in the antisense approach. A similar conclusion can be drawn from Lönnberg's work reported in 2005 that described the synthesis of homothymidylates and phosphorothioate analogs protected by the biodegradable 2,2-bis(ethoxycarbonyl)-3-(pivaloyloxy)-propyl and 2-cyano-2-(2-phenylethylaminocarbonyl)-3-(pivaloyloxy)propyl groups (Figure 1A and 1B) [41]. Indeed, this work also did not lead to ONs for use in control of gene expression.

In addition, Lönnberg described the 4-acetylthio-2,2-dimethyl-3-oxobutyl group as another phosphate protecting group that should be removed by both, esterases and heat (Figure 2) [42]. The resulting phosphotriesters of short oligothymidylates were successfully converted into phosphodiesters at 37 °C, but some cleavage of internucleosidic bonds also occurred. The slow conversion could be accelerated upon the addition of hog liver esterase, but the accumulation of negative charge slowed down the enzymatic hydrolysis. These preliminary data did not provoke further development of such an approach.

Unfortunately, despite many strategies, all attempts to synthesize DNA ONs with SATE-phosphotriesters resulted in poor synthetic yields that made biological evaluation impossible. Consequently, for about ten years, research in the field of carboxyesterase-responsive ONs protected at the phosphate backbone had waned until Dowdy reported on the synthesis, delivery and in vivo activity of siRNA prodrugs containing charge-neutralizing phosphotriester linkages [43]. This recent publication, which was twice highlighted by C. Ducho [44] and A. Khvorova [45], is a reference in the field of ON prodrugs because, for the first time, a biological effect was measured in mice. Indeed, Dowdy's group succeeded in the synthesis of a library of more than 40 phosphotriester groups on ribonucleic neutral (RNN) phosphoramidite building blocks containing 2'-modifications (2'-F, 2'-OMe) to avoid 2'-OH nucleophilic attack on the phosphotriester linkage. Moreover, they used

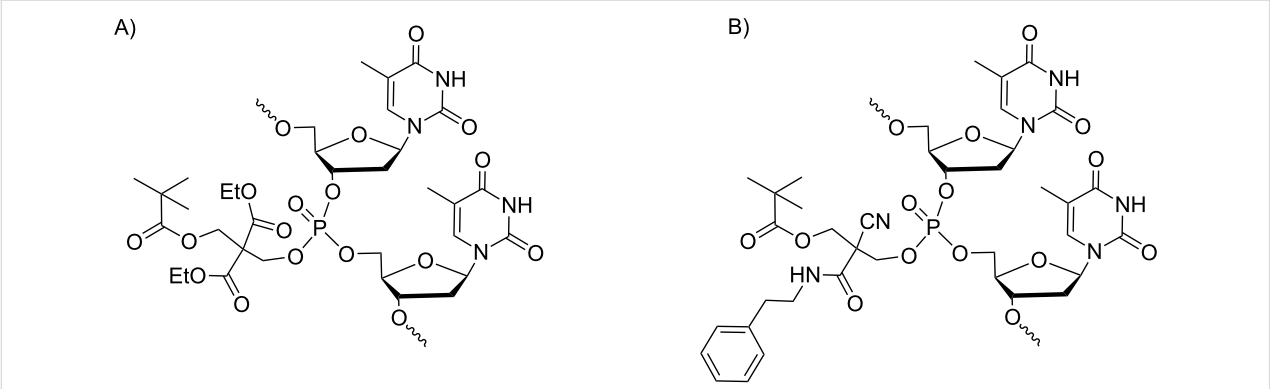


Figure 1: Oligothymidylates bearing A) 2,2-bis(ethoxycarbonyl)-3-(pivaloyloxy)propyl- and B) 2-cyano-2(2-phenylethylaminocarbonyl)-3-(pivaloyloxy)propyl phosphate protecting groups [41].

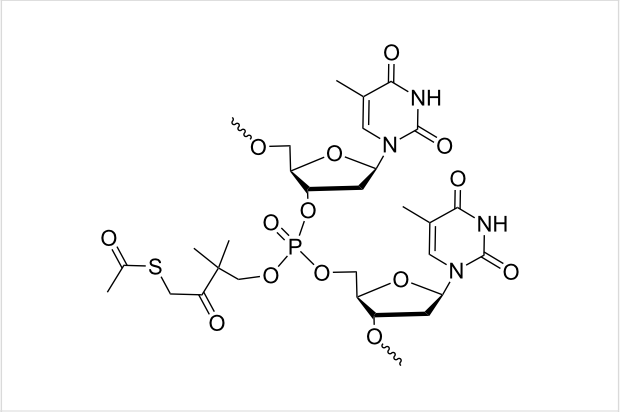
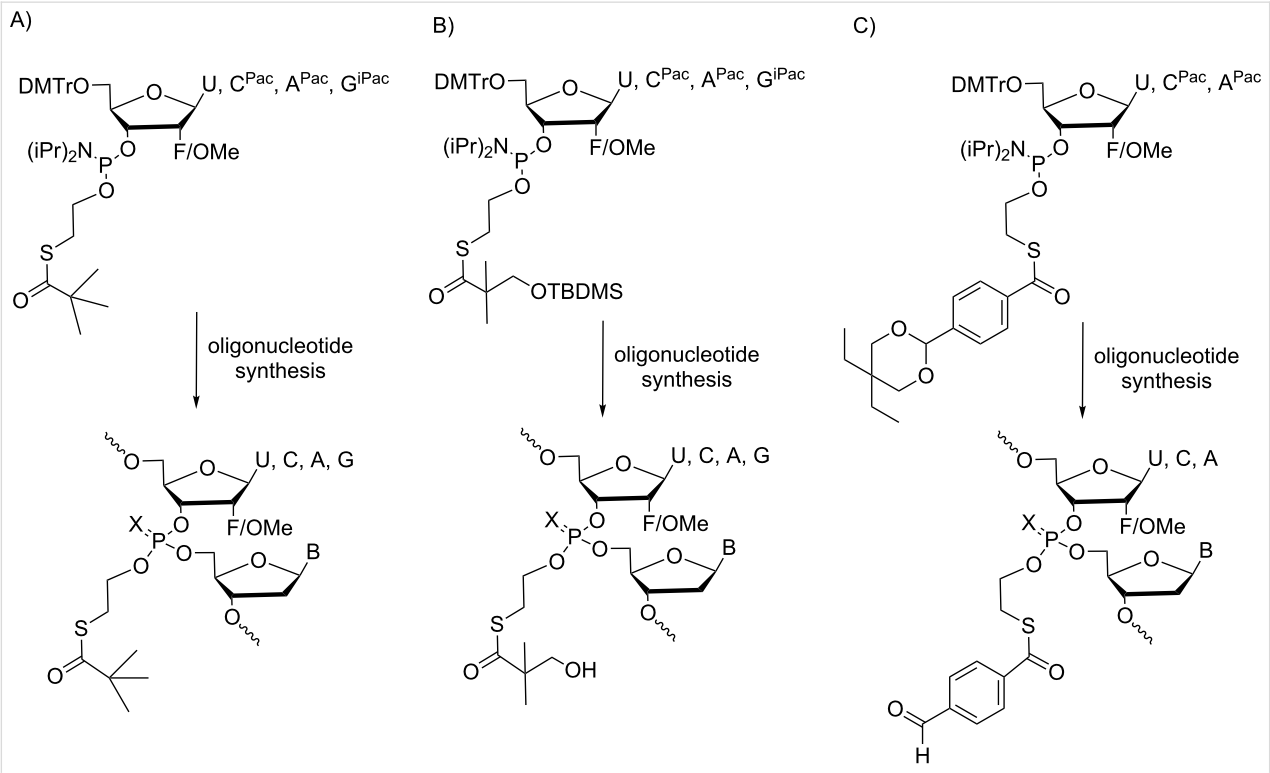


Figure 2: Oligothymidylates containing esterase and thermo-labile (4-acetylthio-2,2-dimethyl-3-oxobutyl) phosphate protecting groups [42].

extremely mild basic diisopropylamine in methanol to deprotect nucleobases containing phenoxyacetyl (for A and C) or isopropylphenoxyacetyl (for G) groups on exocylic amines. These deprotection conditions prevent base-mediated phosphotriester cleavage. Finally, to address the synthetic issue completely, they stabilized the thioester bond to diisopropylamine/methanol by substituting electron-donating groups at the distal α-carbon or lengthening the proximal ethyl linker to a butyl linker. With such RNN phosphoramidite building blocks >3000, RNN ONs have been synthesized with high yields comparable to those of RNA synthesis, demonstrating the robustness and versatility of the chemical method. Three enzymolabile phosphotriester groups, namely, *t*-Bu-SATE, OH-SATE and a conjugable aldehyde A-SATE for conjugation to delivery and targeting domains, have been selected for complete evaluation (Scheme 9A, 9B, and 9C, respectively). The optimum phosphotriester placement and number of phosphotriester groups were shown to have an important impact on the siRNA solubility and duplex stability. Such designed siRNNs showed a high solubility and serum stability and are not recognized by the innate immune system. On the other hand, due to their large size, they do not passively cross cell membranes. Therefore, to facilitate their uptake, a TAT-peptide delivery domain was conjugated to the siRNNs via A-SATE phosphotriester groups. Hence, a chimeric passenger strand containing four A-SATE phosphotriesters duplexed with an RNN guide strand was conjugated to the delivery domain TAT peptides. The resulting conjugates possessing only ≈25% of neutralized phosphates and four TAT peptides were optimal to enter cells passively. Once inside the cells, the SATE groups were efficiently removed by esterases, leading to siRNAs that are induced according to knockdown with apparent EC_{50} values in the low nanomolar range and in a noncytotoxic fashion. Next, the authors prepared conjugates of the siRNNs via one A-SATE phosphotriester with a hepatocyte-specific tris-*N*-acetylgalactosamine targeting domain and demonstrated a stronger RNAi response in mouse liver

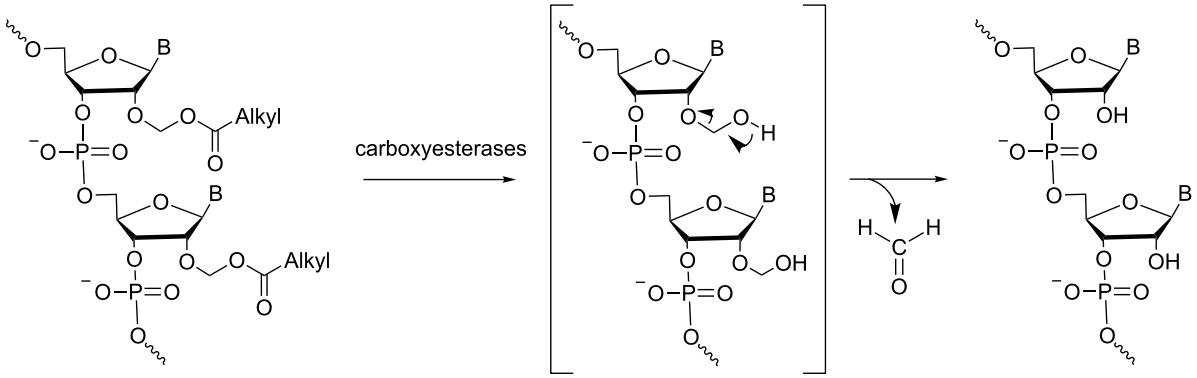


Scheme 9: Phosphoramidites and the corresponding RNA prodrugs protected as A) *t*-Bu-SATE, B) OH-SATE and C) A-SATE phosphotriesters [43].

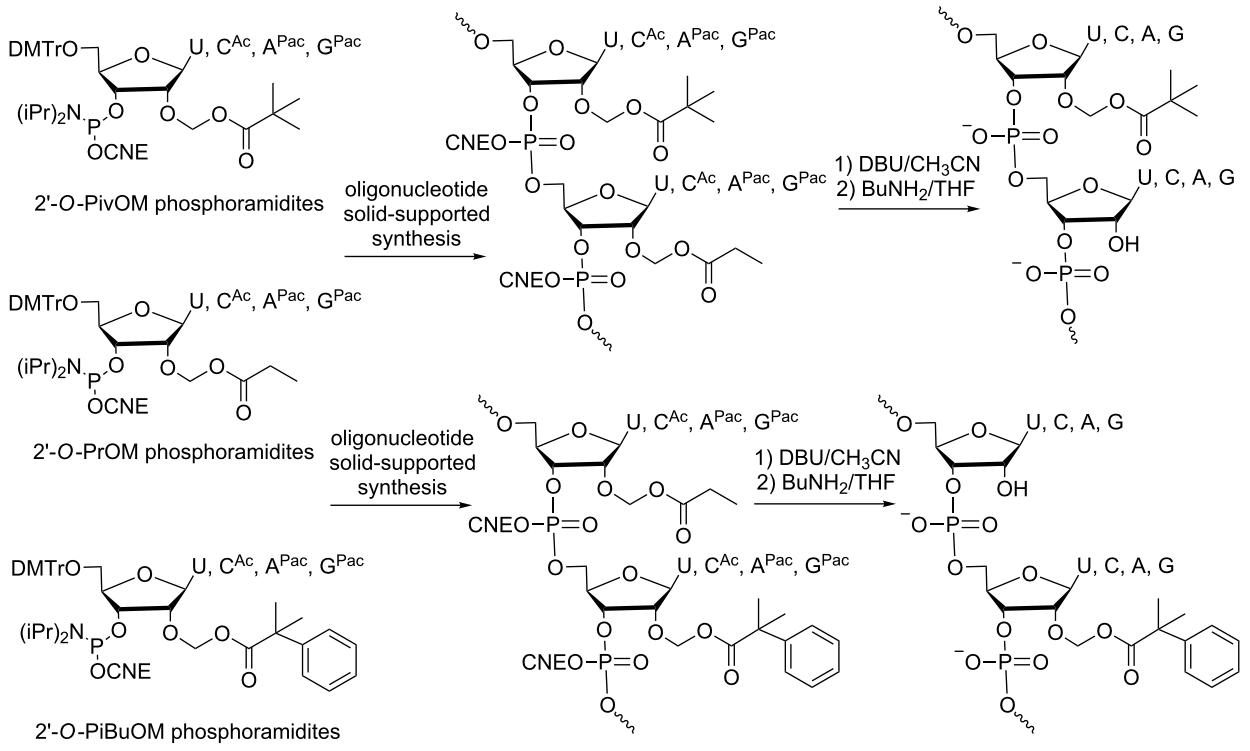
(following subcutaneous or intravenous administration) than the same conjugates with non-enzymolabile phosphotriesters as reference compounds. In conclusion, from this relevant study, it is noteworthy that for the first time, siRNA prodrugs have been synthesized by a versatile method and are intracellularly converted into natural phosphodiester siRNAs that induce robust RNAi responses in vivo. This work clearly opens the way to the new development of ON prodrugs for RNAi therapeutics.

Modifications at the sugar: For the last ten years, our group has been more interested in making RNA prodrugs with enzyme-cleavable modifications at the 2'-position. We essentially focused on several acetalester groups whose lipophilicities and stabilities were variable to tune siRNA properties, particularly their delivery. The first evaluation of biolabile 2'-O-modifications was achieved using short oligo-U sequences containing 2'-*O*-acyloxymethyl or acylthiomethyl groups [46,47]. They were shown to improve RNA nuclease resistance and not to hamper duplex dsRNA formation, and they are removed by cellular esterases. Indeed, 2'-*O*-acyloxymethyl ONs are converted to unmodified RNAs by carboxyesterase-mediated deacylation with the release of formaldehyde to produce the parent RNA (Scheme 10).

These features made 2'-*O*-acetalester modifications promising for their use in a prodrug approach; of particular interest was the pivaloyloxymethyl (PivOM) group, which completes the requirements to functionalize a potential siRNA prodrug. Therefore, for the first time, several mixed-nucleobase RNAs partially 2'-O-masked with PivOM groups were synthesized via a solid-phase method involving silyl-based protections on amino functions of the nucleobases combined to CNE on phosphates and Q-linker between pro-RNA and the solid support [48]. One of them with five PivOM groups at the 5'-end was active in a human cell culture-based RNA interference assay, and it exerted improved cellular uptake. These preliminary data provided a proof-of-concept for a prodrug-based approach for the delivery of siRNA to living human cells. The next report described a more convenient and straightforward method to synthesize partially modified 2'-*O*-PivOM RNAs (Scheme 11) [49]. The strategy involves standard labile acyl groups for nucleobases, cyanoethyl groups for phosphates, a Q-linker to the solid support [50] and two acetal ester groups for 2'-OH, namely, propionyloxymethyl (PrOM) and PivOM exhibiting different stability under deprotection conditions. Indeed, a specific treatment with butylamine in anhydrous THF [51] selectively removes the PrOM groups while the PivOM groups stay attached. Thus, partially PivOM-modified siRNAs with a different design have been evaluated. No serious thermal destabilization of the siRNA duplex was observed and the A-form duplex was maintained [52]. Moreover, all PivOM-modified siRNAs (1 nM) showed control of gene expression activity after transfection into ECV304 cells expressing the firefly luciferase gene. Nevertheless, the RNAi activity of such 2'-*O*-acetal ester siRNAs taken up by cells in the absence of any carriers remained to be demonstrated. The robust synthetic method developed in 2014 [49] made 2'-PivOM-modified siRNAs readily available. To improve their lipophilic features, one methyl of the *tert*-butyl moiety in the PivOM groups was replaced by one phenyl, resulting in the phenylisobutyryloxymethyl (PiBuOM) modification, which was introduced into siRNAs for investigation (Scheme 11) [53]. Indeed, we provided evidence of improved spontaneous cellular uptake of naked PiBuOM-modified siRNAs compared to unmodified or PivOM-modified siRNAs. Consequently, a substantial inhibition (90% at 1 μM concentration) of EWS-Fli1 expression in A673 cells in serum-containing medium was observed. It is noteworthy that this PiBuOM modification is efficient in assisting siRNAs to enter cells and promote gene inhibition without the use of transfecting agents. Furthermore, even if the intended prodrug



Scheme 10: Mechanism of the hydrolysis of 2'-*O*-acyloxymethyl ONs mediated by carboxyesterases [46]. The hydrolysis of the ester functions yields an unstable 2'-hemiacetal, affording the free RNA through the release of formaldehyde.



Scheme 11: Synthesis of partially 2'-*O*-PivOM-modified RNAs [49] and 2'-*O*-PiBuOM-modified RNAs [53] using their corresponding phosphoramidites and 2'-*O*-PrOM phosphoramidites to generate 2'-OH.

strategy was not validated with PiBuOM modification because of a certainly too slow esterase cleavage, its use in the sense strand as permanent lipophilic modification has been relevant to facilitating the cellular uptake of siRNAs and subsequent gene inhibition.

Beside it is known that cellular internalization properties can be improved by adding positive charges to ONs to counterbalance the overall negative charge of these compounds. In this context and in extension of the previous work with the 2'-*O*-acetal ester modifications cited above, new modified ONs were designed with amino or guanidino-containing 2'-*O*-acetal ester groups bearing positive charges: 2-amino-2-methylpropionyloxymethyl (AMPrOM), 2-aminomethyl-2-ethylbutyryloxymethyl (AMEBuOM) or 2-guanidinomethyl-2-ethylbutyryloxymethyl (GMEBuOM, Figure 3A) [54]. The two modifications with a guanidinium and an ammonium moiety, GMEBuOM and AMPrOM, respectively, were found to be unstable during HPLC purification and handling. Therefore, they could not be further investigated. By contrast, the AMEBuOM modification was evaluated within several 2'-OMe ONs or a fully AMEBuOM-modified ON, which was more resistant to enzymatic degradation. A slightly moderate internalization of AMEBuOM-modified ON (ammonium side chain) was observed compared to the ON with the PivOM group (*t*-Bu side chain), probably due to the instability of AMEBuOM groups in cell culture medium before internalization. Overall, these cationic acetal ester modifications are chemically too unstable for further developments as ON prodrugs. Similarly, Damha reported on the synthesis of ONs containing amino acid-acetal esters at the 2'-OH, particularly with lysine for its positive charge (Figure 3B) [55]. Unfortunately, 2'-*O*-acetal ester ONs with lysine, alanine and phenylalanine could not be isolated with good yield because they were partially degraded during HPLC purification and subsequent handling. No further study has been described in the literature with such 2'-modified ONs.

Prodrugs of conformationally constrained nucleic acids such as tricyclo-DNA (tc-DNA) deserve to be mentioned in this review as sugar-modified ONs. Indeed, tc-DNAs were evaluated as promising candidates for ON-based therapeutic applications, exhibiting increased affinity to RNA and better resistance to nucleases. The main bottleneck of their use, as for many other modified ONs, is their poor cellular uptake. Therefore, to address this issue, Leumann et al. synthesized "pro-tricyclo-ONs" bearing two different metabolically labile ethyl and hexadecyl esters at position C6' that were expected to promote cell penetration (Scheme 12) [56]. It was shown that the cellular uptake of a decamer containing five tc^{hd}-T units with a C_{16} side chain was increased in two different cell lines (HeLa and HEK

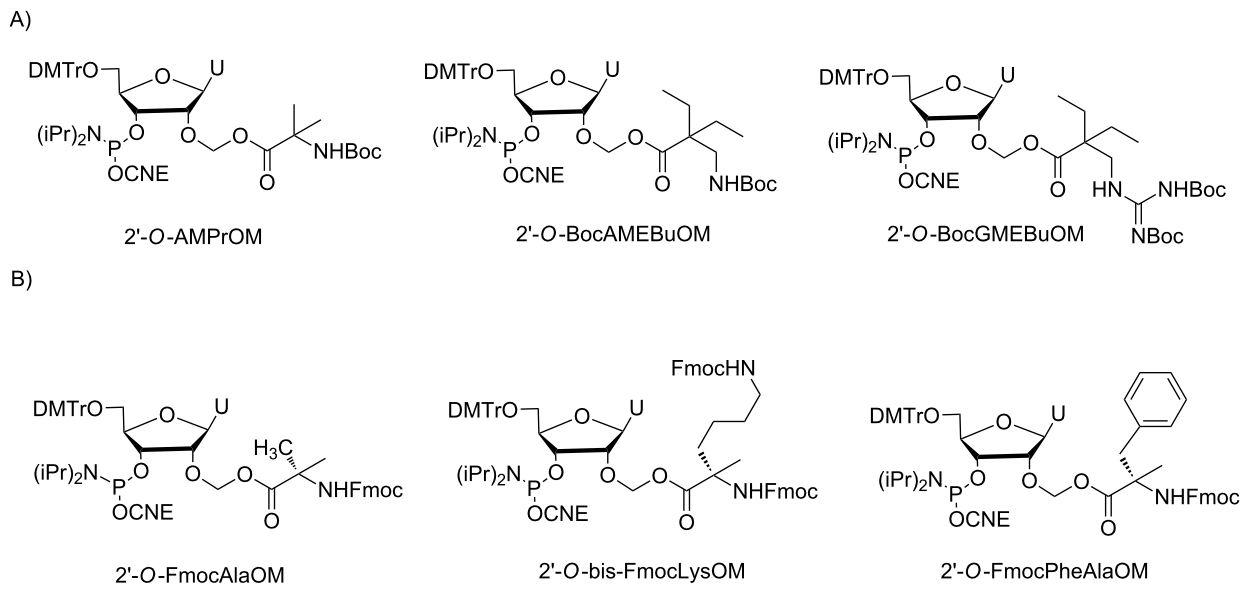
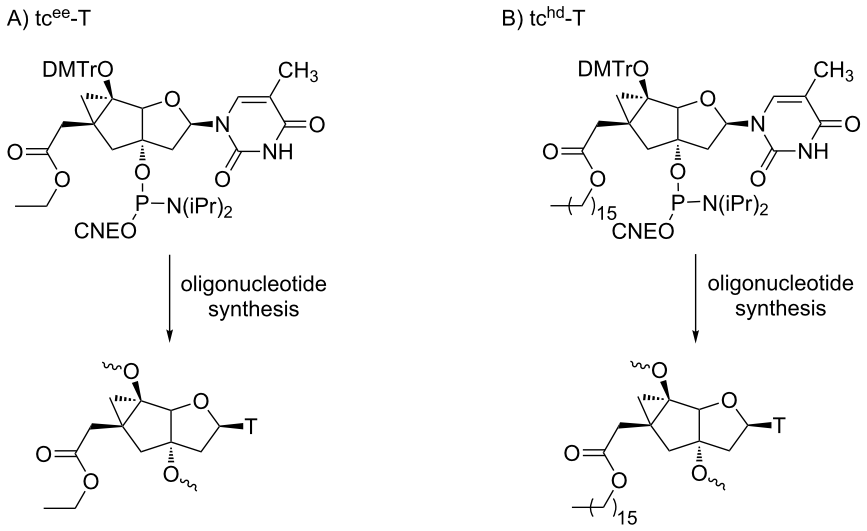


Figure 3: A) 2'-*O*-amino and guanidino-containing acetal ester phosphoramidites and B) 2'-*O*-(amino acid) acetal ester phosphoramidites reported by Debart [54] and Dahma [55], respectively.



Scheme 12: Prodrugs of tricyclo-ONs functionalized with A) ethyl (tc^{ee}-T) and B) hexadecyl (tc^{hd}-T) ester functions at C6 obtained from corresponding thymidine phosphoramidites [56].

293T) without using a transfection agent. Nevertheless, the enzymatic hydrolysis of the hexadecyl esters and some preliminary antisense activities remain to be demonstrated.

Heat-responsive ONs

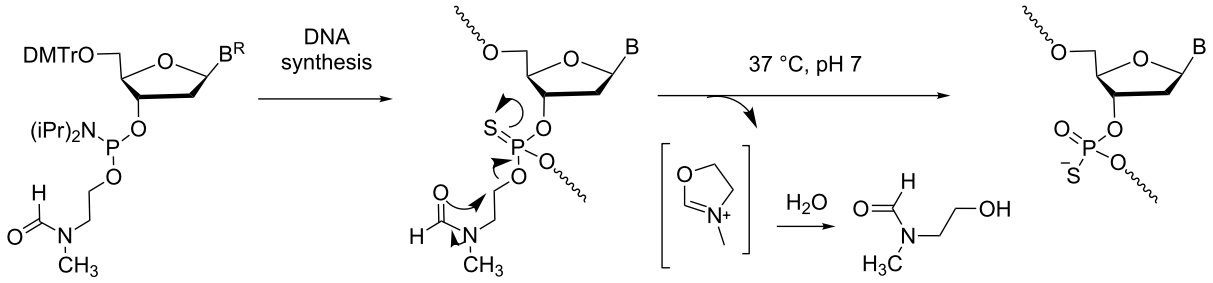
These so-called ONs contain thermolytic groups that are removed upon a 'heat-driven' process under neutral conditions.

Modifications at the internucleotide linkage

Over many years, various thermolytic groups for 5'-OH and phosphate protections have been designed and developed by Beaucage et al. to synthesize DNA ONs on microarrays due to their rapid removal under mild conditions [57]. Heat-sensitive phosphate/thiophosphate-protecting groups have been incorporated into ONs via phosphoramidite chemistry using solid-support methodology. However, some required more drastic conditions (90 °C for a long period of time) to be cleaved, and Beaucage found a potential application of such thermolytic ONs as prodrugs in the treatment of infectious diseases. Even if in this review, the applications of ON prodrugs are essentially focused on gene silencing, it seemed important to us to report on the thermolytic CpG-containing ODNs as potential

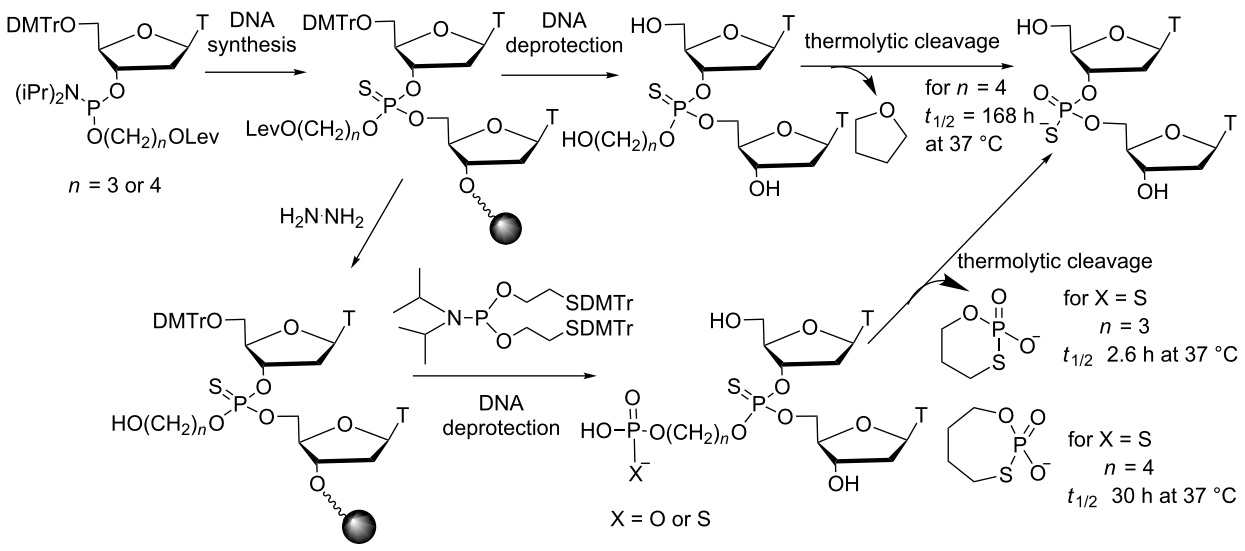
immunotherapeutic prodrugs [58]. The first impressive result was obtained in vivo with a CpG ODN (CpG ODN *fma*1555) functionalized with the 2-(*N*-formyl-*N*-methyl)aminoethyl (*fma*) thiophosphate protecting groups, which were cleaved at 37 °C to yield the well-known immunomodulatory CpG ODN 1555 (Scheme 13). When the CpG ODN *fma*1555 was administrated to newborn mice that had been infected with Tacaribe virus, 43% of mice survived [58]. Moreover, an improved immunoprotection (60–70% survival) was obtained when the CpG ODN prodrug was administered three days before infection. Interestingly, it also was shown that the combination of CpG ODN 1555 and CpG ODN *fma*1555 (more than 50% survival) increased the window for therapeutic treatment against the disease. However, the induction of the immunostimulatory effect was delayed, which is consistent with the formation of the biologically active phosphorothioate diesters from the *fma* thiophosphate triesters with a thermolytic conversion half-life of $t_{1/2}$ = 73 h at 37 °C.

Although these *fma* ODNs exhibit the features of ON prodrugs in that they are neutral to enable cellular delivery and are stable to hydrolytic nucleases, Beaucage et al. developed other thermolytic ONs with thermolabile groups displaying slower or faster removal kinetics than that of *fma* groups. In particular, the subsequent heat-sensitive groups for phosphate masking were designed with a phosphate or a thiophosphate branched to a propyl or a butyl chain connected to the internucleoside linkage (Scheme 14) [59]. Consequently, the presence of only one phos-



B^R = thymin-1-yl, 4-*N*-benzoylcytosin-1-yl, 6-*N*-benzoyladenin-9-yl, 2-*N*-isobutyrylguanin-9-yl
B = thymin-1-yl, cytosin-1-yl, adenin-9-yl, guanin-9-yl

Scheme 13: Demasking mechanism of *fma* thiophosphate triesters in CpG ODN upon heat action [58].



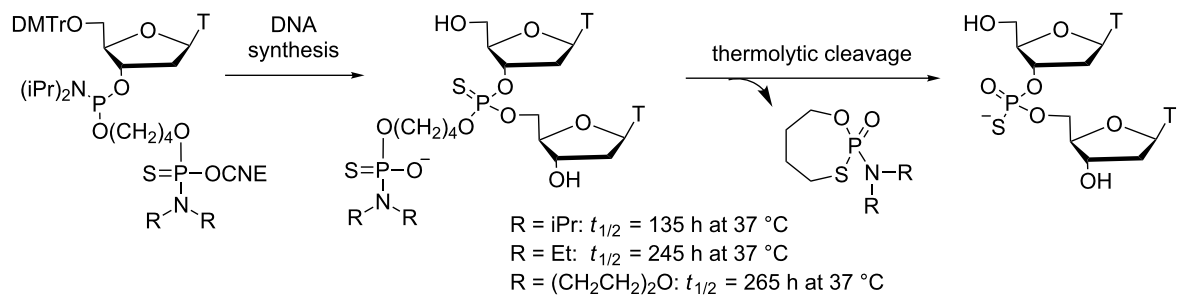
Scheme 14: Thermolytic cleavage of the hydroxy-alkylated thiophosphate and phosphato-/thiophosphato-alkylated thiophosphate protecting groups from thymidine dinucleotides [59].

phate monoester function in an *fma* ON significantly increased the solubility. Unfortunately, no biological evaluation of such modified ONs was performed, and only the complete conversion of modified CpG into unmodified CpG upon elevated temperature conditions was shown.

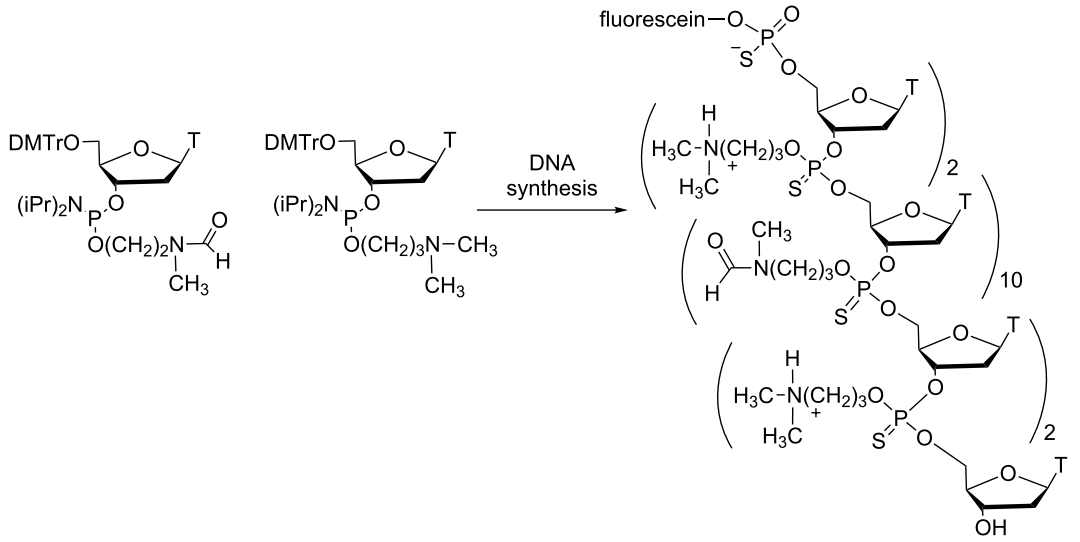
Another study in the same laboratory described new heat-sensitive thiophosphate protecting groups derived from the previously cited *fma* [58] and 4-(methylthio)butyl groups [57]. Some 20 groups, which will not be detailed here, have been assessed and were found to exhibit slower or faster thermolytic deprotection rates than those of the *fma* group at 37 °C ($t_{1/2}$ = 72 h) [60]. Typically, the thermostable groups with deprotection kinetics slower than those of the *fma* group may be used for the protection of terminal phosphodiesters of the immunomodulatory DNA sequence targeting the nuclease resistance of the ON prodrug. On the other hand, the thermosensitive groups are more suitable for the protection of the thiophosphates flanking the CpG motif of DNA prodrugs to provide both lipophilicity (better cellular uptake) and hydrophilicity (better solubility once groups are removed). Moreover, some of thermolabile groups ($t_{1/2}$ in the range of 6 h to 40 h at 37 °C) may be applicable to protect the thiophosphates of CpG motifs of immunoregulatory DNA sequences. Thus, the investigation of these different heat-sensitive groups may serve to design optimal CpG DNA prodrugs.

Similarly, in the search for thiophosphate protecting groups with deprotection half-lives in the range of 100–200 h at 37 °C for sustained CpG ODN immunostimulation in animal models, Beaucage et al. have developed a new class of thermosensitive groups that are hydroxy-alkylated phosphoramidate, phosphoramidothioate and phosphorodiamidothioate derivatives (Scheme 15) [61]. Their thermolytic deprotection rates at 37 °C have been determined in PBS (pH 7.4) from thymidine dinucleoside phosphorothioate models. It was shown that the thermolytic cleavage of alkylated (diisopropyl, diethyl, morpholino) phosphoramidothioylbutyl groups to TpsT proceeded with respective half-lives of 135 h, 245 h and 265 h at 37 °C. Therefore, these groups are appropriate for thiophosphate protection of the CpG motif of CpG ODN prodrugs, and they are complementary to those identified earlier [60]. It remains to study such thermosensitive CpG ODNs in animal models infected by viruses and/or bacteria to evaluate the correlation between extended immunostimulation and resistance.

The most recent data reported by Beaucage on thermosensitive PS DNA prodrugs were related to the assessment of their internalization in various cell lines [62]. The study was essentially performed with oligothymidylate models. First, the internalization of a 5'-fluorescein *fma* $(Tps)_{14}T$ in Vero, HeLa and GC-2 cells was poor but comparable to that of the control 5'-fluorescein $(Tps)_{14}T$. These data can be explained by the decreased solubility in aqueous medium of the uncharged ON and can be correlated with the similar abilities of CpG ODN *fma*1555 and CpG ODN 1555 to induce an immunostimulatory response in the mice mentioned above [58]. On the other hand, the introduction of four positively charged 3-(*N,N*-dimethylamino)propyl groups into an *fma*-thiophosphate oligothymidylate resulted in enhanced aqueous solubility and a 40-fold increase in the cellular uptake of the ON in Vero and GC-2 cells (Scheme 16). It is noteworthy that the presence of four positively charged groups into a negatively charged PS oligothymidylate is not sufficient for an efficient cellular internalization in Vero cells. These data support that both 3-(*N,N*-dimethylamino)propyl and *fma* groups are required for optimal internalization in the three cell lines. Of special interest was the absence of cytotoxic effects in Vero cells at a 50 μM extracellular ON concentration for 72 h. Moreover, confocal microscopy studies showed that the positively charged oligoT escaped endosomal vesicles and migrated to the nucleus of Vero or GC-2 cells. This observation may support the correlation between cellular uptake and the activity of ther-



Scheme 15: Synthesis via phosphoramidite chemistry and thermolytic cleavage of alkylated (diisopropyl, diethyl, morpholino) phosphoramidothioylbutyl internucleoside linkages [61].



Scheme 16: Synthesis of thermosensitive prodrugs of ODNs containing *fma* thiophosphate triesters combined to positively charged 3-(*N,N*-dimethylamino)propyl phosphotriesters internucleoside linkages to improve cellular uptake [62].

mosensitive DNA prodrugs. Supplementary experiments with mixed-nucleobase DNA sequences should provide more information on these thermosensitive ON prodrugs.

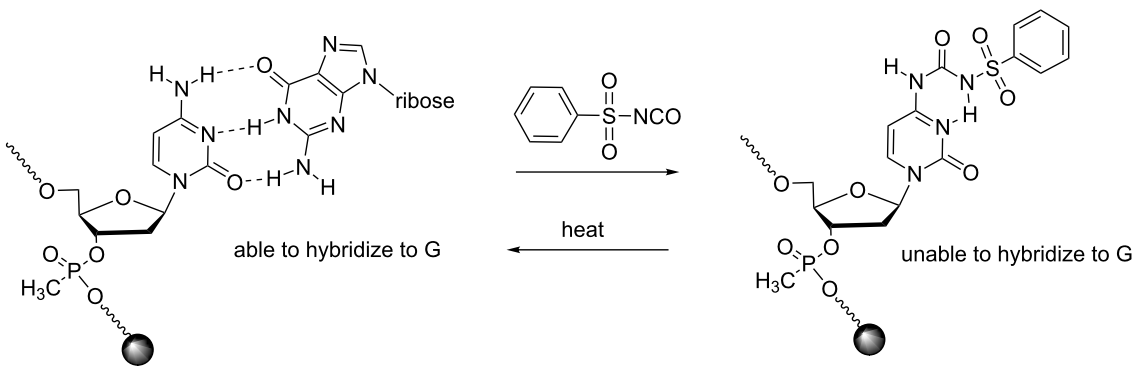
Finally, it should be mentioned that additional thermolabile protecting groups for phosphodiesters have been reported by Lönnberg [63,64]. Actually, in the search for esterase-labile protecting groups for phosphoesters, a set of 2,2-disubstituted 4-acylthio-3-oxobutyl groups was additionally thermolabile. This investigation was only achieved at the nucleotide stage and no data with ONs were reported. Consequently, these special protecting groups will not be detailed in this review.

Modifications at the nucleobase

The temporary protection of nucleobases by heat-responsive groups has not yet found applications in the field of ON prodrugs despite a certain potential. Indeed, the introduction of the phenylsulfonylcarbamoyl (psc) protection of cytosines in methylphosphonate ONs through the reaction with phenylsulfonyl isocyanate produces a caged ON unable to hybridize to its complementary RNA sequence until heat removal of the psc (Scheme 17) [65]. However, currently, this approach is limited to CPG-supported methylphosphonate ONs containing thymines and cytosines immobilized on a glass slide.

Light-responsive ONs

Compared to other stimuli used to generate ONs that act as gene regulator, light is the external physical regulatory element that is most used. Actually, photoirradiation is the major and simplest method to temporally and spatially regulate the activity of photoresponsive ONs that could be assimilated to prodrugs, although this term is not commonly used except in a



Scheme 17: Caging of deoxycytidine in methylphosphonate ONs by using the thermolabile phenylsulfonylcarbamoyl protecting group introduced through reaction with phenylsulfonyl isocyanate [65].

few reports [66]. Depending on the strategies used, the introduction of photolabile moieties into an ON renders it active or inactive and, therefore, it is turned on (light activated) or off (light deactivated) by light, respectively [67-71]. Thus, the advantages of light are to give the possibility of controlling this switch in time but also in space because photoirradiation could be performed only on a desired part of a sample, a cell, a tissue or a living organism. However, it should be noted that currently, most of the activities of photocaged ONs have been validated on reporter gene models except for a few studies on specific genes in zebrafish embryos.

Despite the advantages described above, the use of light to control gene expression has several drawbacks. Extended UV irradiation may produce side reactions, lowering the yield of active ON and inducing toxicity. Moreover, the diffusion of light resulting from long UV irradiation decreases temporal and spatial resolution for experiments in cells. Finally, because light has poor tissue diffusion, the photocaging approach may be restricted to in vitro gene-silencing interactions and of limited use for therapeutic applications.

Modifications at the phosphate moieties

The control of gene expression with photocaged phosphate-modified ONs has been mostly used for light activation of RNA interference, as commonly used by the Friedman group [72-76], and occasionally for RNA-cleaving activity with DNAzymes [77].

It is expected that phosphate-modified siRNAs sterically block the interaction of siRNA with the RISC complex and that the process is turned on upon photoirradiation [72]. Considering DNAzymes, their catalytic activity is inhibited until photoirradiation releases the native DNAzyme [77]. In phosphate-caged siRNAs, chemical derivatization of phosphates either in the phosphodiester backbone [72] or at a terminal phosphate [73,74] of ON was performed following two different approaches: a) post-functionalization of ON with a suitable reagent, which generally is a diazo derivative bearing a photoresponsive moiety, or b) incorporation of an appropriate photocaged phosphoramidite during the solid-supported ON synthesis [73,78]. The advantage of the first approach is that the functionalization results from a reaction with available unmodified ONs, while the second approach first requires the synthesis of a modified unit followed by its incorporation into ON during solid-phase synthesis. However, the first approach is far less efficient than the second one because the labeling of phosphodiester linkages with diazo compounds is not specific to a given phosphodiester in siRNA and cannot be controlled in location and the amount of caging units, yielding a random mixture of ONs. Moreover, diazo compounds exhibit certain reactivity toward nucleobases that can lead to undesired side reactions [74]. Considering their RNAi activity, these statistically phosphate-caged RNAs also have several drawbacks. Indeed, Friedman et al. have shown that low percentages of photolabile-protecting groups in siRNA only induce partial inhibition of gene silencing. Inversely, higher percentages increase the blocking of RNAi before light activation induces the release of photoresponsive moieties during photoirradiation, yielding a lower extent of GFP expression in HeLa cells [72].

Later, Mc Master showed that it is not necessary to heavily modify siRNA because a single photoresponsive unit (biotin linked to nitrophenylethyl, Figure 4) at the phosphate located at the 5'-end of the antisense strand of a siRNA decreased RNAi, although only moderate photomodulated silencing of several transfected genes in HeLa cells was observed [73]. In this work, the responsive unit was introduced into an ON using the corresponding phosphoramidite (Figure 4), but Friedman showed that this also could be done by the reaction of diazo compounds with the terminal phosphates of an ON. Indeed, the reactivity of diazo reagents with terminal phosphates (phosphomonoesters) was much greater and more specific than that with the internucleoside phosphates (phosphodiesters) [74].

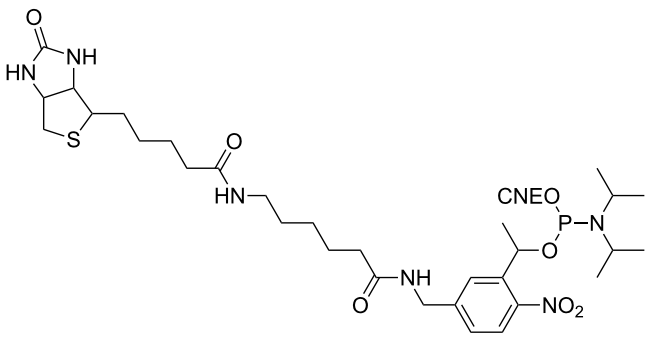
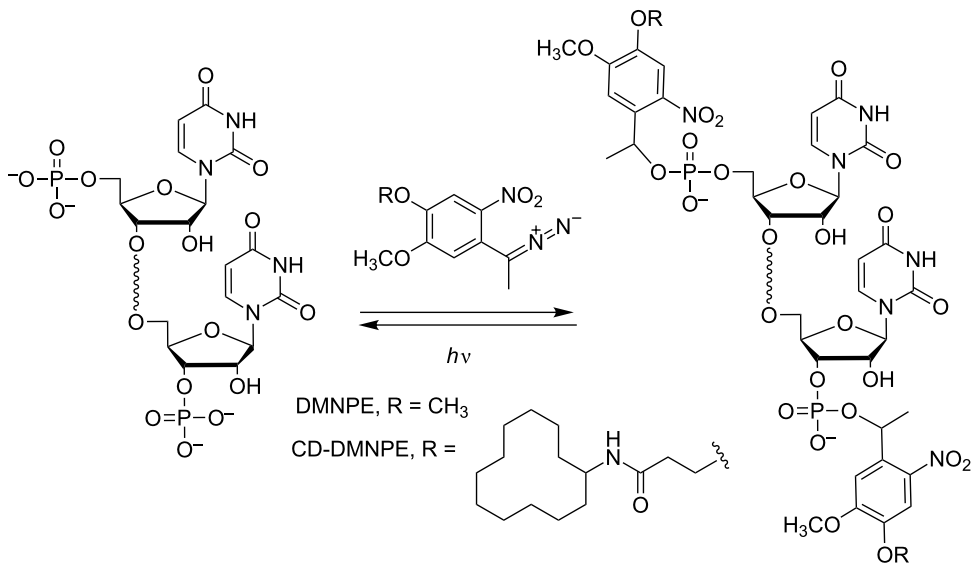


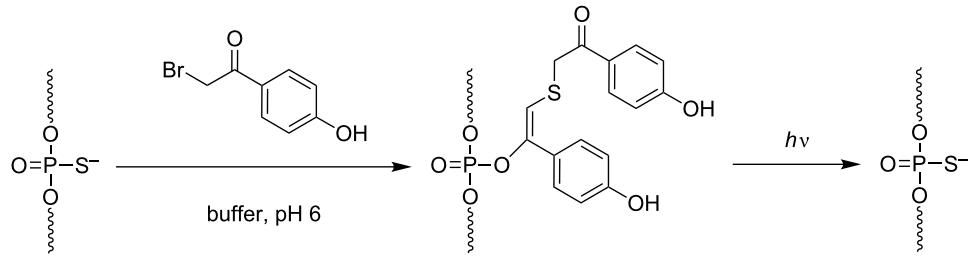
Figure 4: Biotinylated 1-(5-(aminomethyl)-2-nitrophenyl)ethyl phosphoramidite used to cage the 5'-end of a siRNA during its synthesis on solid support using phosphoramidite chemistry [73].

Friedman improved the efficiency of the phosphate caging approach by introducing photolabile moieties (dimethoxynitrophenylethyl = DMNPE) at the phosphate 5' and 3'-ends of both strands of siRNAs (Scheme 18) [74]. Here, again, the inhibition of gene silencing due to the caging moieties has not been complete, although much better than that with the backbone-modified siRNAs, in spite of the fact that the RNAi was fully restored after photoirradiation. One of the possible reasons for the partial inhibition of gene silencing by the photocaged siRNA (35% knockdown without photoirradiation) could be explained by the partial loss of terminal photoreactive units due to nuclease degradation. Friedman et al. have first improved their system using phosphorothioate (PS) internucleoside linkages to enhance nuclease resistance near the terminal caged phosphates preventing unwanted loss of the photoreactive moieties before photoirradiation [75]. This was the case when two PS linkages were introduced into each strand of caged siRNAs. Surprisingly, an increasing number of PS, up to 6 per strand, turned on the caged siRNA to an active species, probably because many PS linkages increased the affinity for DICER overcoming the blocking capacity of the caged ON. Finally, the best results were obtained when bulkier photolabile protecting groups (i.e., cyclododecyl-DMNPE = CD-DMNPE) were employed to cage siRNAs (Scheme 18) [76]. The system was efficient as the photocaged siRNA did not induce RNAi while it was fully deprotected under photolysis restoring the activity of the native siRNA.

As stated previously, the introduction of a photoreactive moiety into the phosphodiester backbone of an ON with diazo compounds is not specific. Xiang et al. developed a more efficient and specific post-synthetic method. It is based on the reaction between a phosphorothioate derivative and 2-bromo-4'-hydroxyacetophenone to produce a phosphate protected with a thioether-enol phosphotriester, phenol substituted (TEEP, Scheme 19) [77]. The TEEP modification was introduced into



Scheme 18: Introduction and cleavage of 1-(4,5-dimethoxy-2-nitrophenyl)ethyl (DMNPE) [74] and cyclododecyl-DMNPE (CD-DMNPE) [76] groups in the terminal 3' and 5'-phosphate of an RNA through reaction with a diazo reagent.



Scheme 19: Post-synthetic introduction of a thioether-enol phosphodiester (TEEP) linkage into a DNAzyme by the selective reaction of a phosphorothioate linkage with 2-bromo-4'-hydroxyacetophenone followed by photodecaging, leading to a phosphodiester internucleoside linkage [77].

"active sites" of 8–17 and 10–23 DNAzymes with good yields (>95%). The inhibition of the 8–17 DNAzyme activity by one modification was limited, whereas the photocaged ON with 3 modifications was totally inactive. Photoirradiation at 365 nm triggered the removal of the photoreactive moieties to phosphodiesters with up to 85% of activity recovery of the DNAzyme in vitro as in HeLa cells.

Modifications at the nucleobase

For selected reviews on this topic, see [79,80]. From all possible photoresponsive modifications introduced into ONs, modifications of the nucleobases are the most widely used for the regulation of gene expression under light activation. For this purpose various different approaches have been reported for the control of RNA translation (such as RNAi [81-83] and antisense [84,85], including splice switching of pre-mRNA [86] and DNAzymes [82,87]) and for the control of gene transcription (such as antigene strategy [88] and decoys [86,89] able to interact with transcription factors). Most of the photoresponsive units are introduced as protecting groups of nucleobases in the ONs. Consequently, the nucleobases cannot hybridize until photoirradiation. Another strategy much less studied than that where natural nucleobases are protected by photolabile groups is to use artificial photolabile nucleobases [90]. Generally, these modified nucleobases are introduced into ONs through their corresponding phosphoramidites.

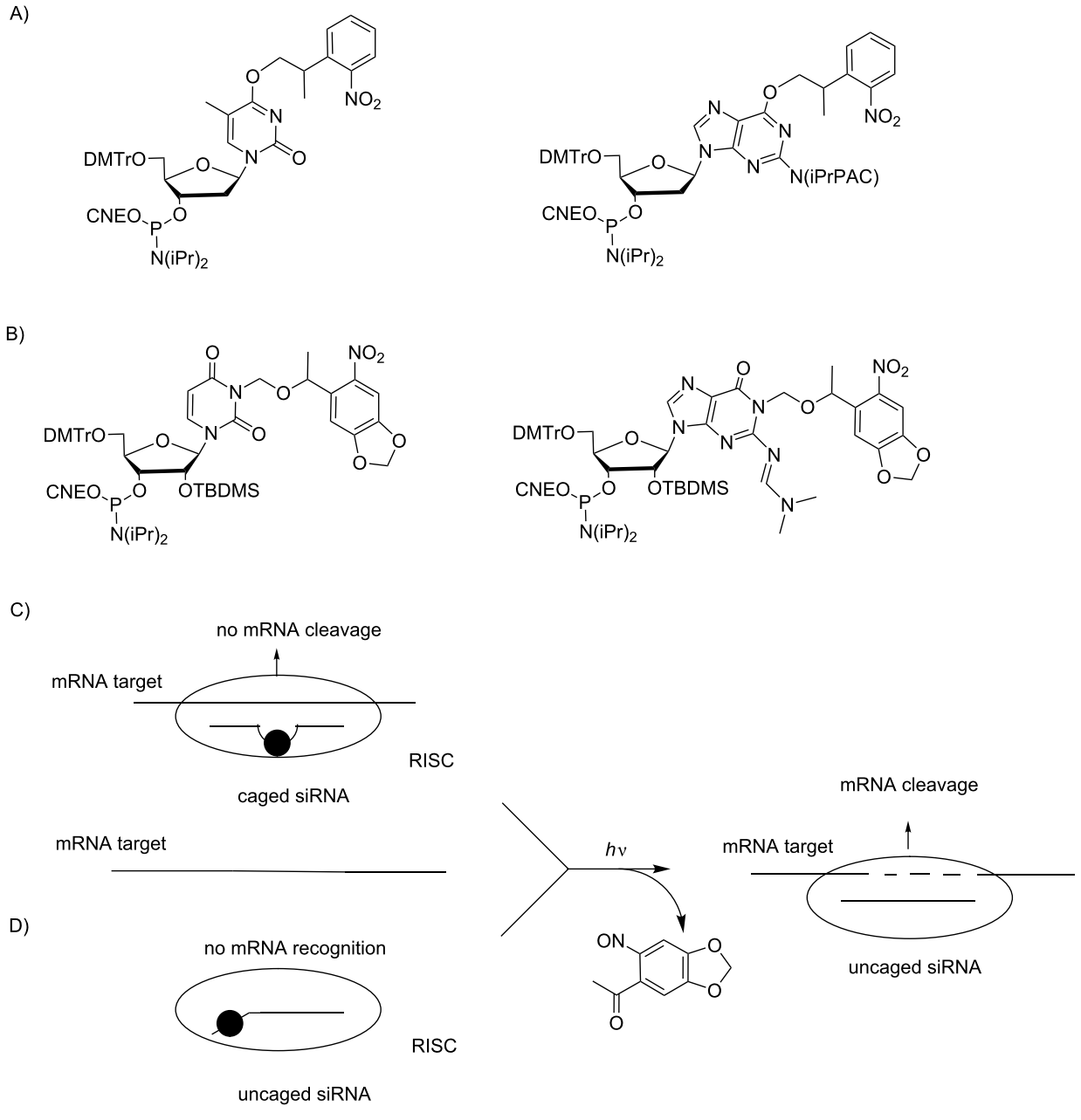
Photocaged approaches to inhibit translation: Mikat and Heckel introduced deoxyguanosine and thymidine, respectively, protected at O6 and O4 with a 2-(2-nitrophenyl)propyl (NPP) group, into siRNA (Scheme 20A) [81]. The most efficient siRNAs targeting EFGP expression in transfected HeLa cells were those modified in the central part of the siRNA – that is, in the nucleobases neighboring the argonaute cleavage site of mRNA (Scheme 20C). These caged siRNAs were completely inactive until removal of the protecting groups with UV irradiation at 366 nm, whereas modifications surrounding the central part of the siRNA were less effective. It was argued that modified nucleotides in the central part of siRNA lead to a bulge of the siRNA–mRNA hybrid, disturbing the cleavage of mRNA by the RISC. Subsequently, Deiters used the same approach with photo 6-nitropiperonyloxymethyl (NPOM)-photocaged siRNAs synthesized from phosphoramidites of the caged uridine and guanosine ribonucleotides (Scheme 20B) [83]. As previously demonstrated, light activation of RNAi was confirmed in HeLa cells transfected with a GFP reporter gene but was also demonstrated with the silencing of the endogenous gene of the mitosis motor protein Eg5. In the same article, Deiters reported the study of siRNAs with caged nucleotides at the seed region of siRNA because the seed region is crucial for the recognition of mRNA target but does not affect the cleavage site (Scheme 20D). Two protected nucleotides in a siRNA totally prevented RNAi that is "turned on" after UV irradiation. Thus, the NPOM-protecting group induces reversible inactivation of siRNAs, demonstrating the importance of hybridization in the RNAi mechanism.

Deiters et al. also applied the NPOM photosensitive group for gene silencing using antisense oligodeoxynucleotides (ODNs) in mouse fibroblast 3T3 cells transfected with the *Renilla* luciferase plasmid [84]. Three and four modifications partitioned along the sequence of the antisense ODN prevented hybridization to RNA targets and consequently inhibited the antisense activity blocking RNase H catalyzed degradation of mRNA. Upon irradiation at 365 nm, the NPOM groups were completely removed and the antisense activity was restored to the level of the uncaged ODN (Scheme 21). Photocaged NPOM thymine was further introduced into morpholino antisense ODNs [85] to block mRNA binding to the ribosome and, therefore, RNA translation. These morpholino ONs could inhibit the EGFP exogenous gene and chordin endogenous gene in zebrafish and Xenopus living embryos, only after UV photolysis at 365 nm (Scheme 21).

In the studies described above, photoirradiation "turns on" antisense activity, and ONs "turn off" gene translation. Photocaging can also be used to "turn off" antisense activity. For this purpose, the antisense ODN was linked to a complementary sequence (Scheme 22) [82]. The resulting hairpin could not associate with the mRNA. When the complementary sequence was photocaged with three NPOM thymidines, the hairpin was not formed, and the antisense hybridized with mRNA, preventing its subsequent translation by RNase H recruitment. Thus, photoirradiation causes hairpin formation and, therefore, "turns off" antisense activity.

Photocaged phosphorothioate (PS) ONs containing 2'-*O*-methyl nucleosides and two NPOM-protected 2'-OMe uridines in their sequences have also been used as splice-switching ONs (Scheme 23) [86]. The NPOM-protecting groups prevented ON hybridization with a β-globin intron aberrant splice site, inducing β-thalassemia in EFGP stably transfected HeLa cells, and the ON was not active until photoactivation.

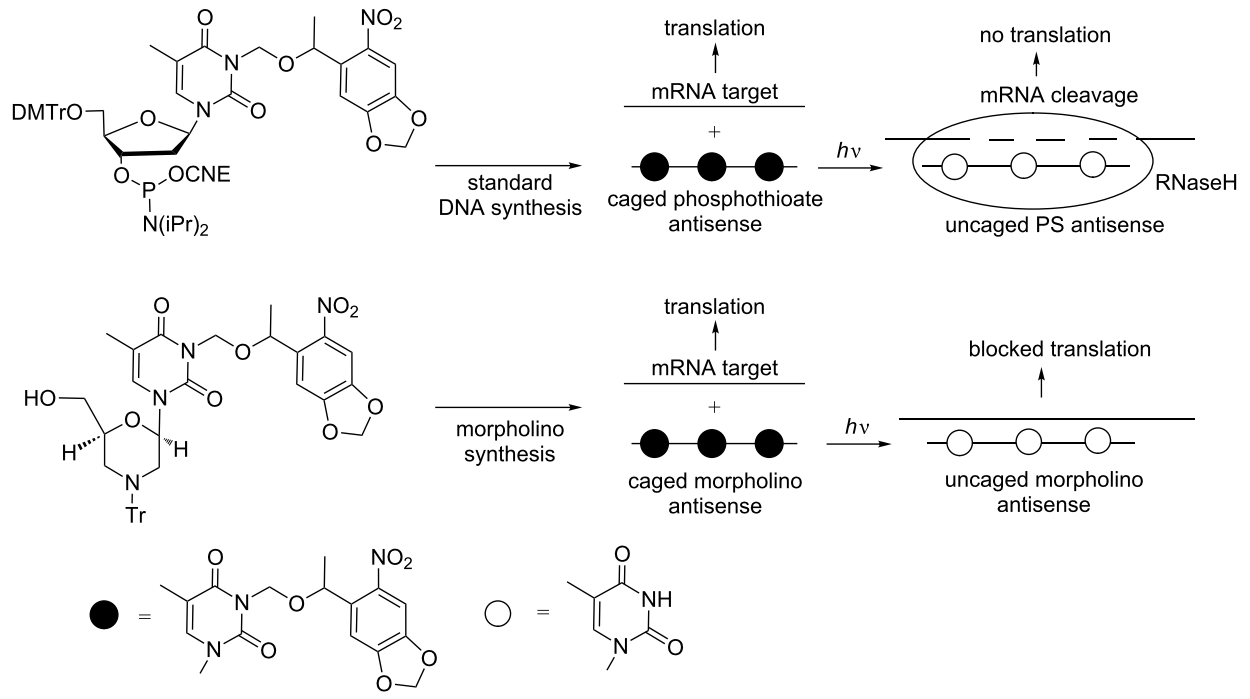
In 2007, Deiters et al. described the recovery under UV irradiation of the catalytic activity of a DNAzyme possessing in its catalytic loop a thymidine caged with the NPOM-protecting group in N3 of thymine (Scheme 24A) [87]. In this approach, the DNAzyme was light activated. Some years after, the same group showed a light deactivation process using a caged hairpin (Scheme 24B) [82]. In this case, the catalytic site of DNAzyme was not caged, but it was associated or linked to a complemen-



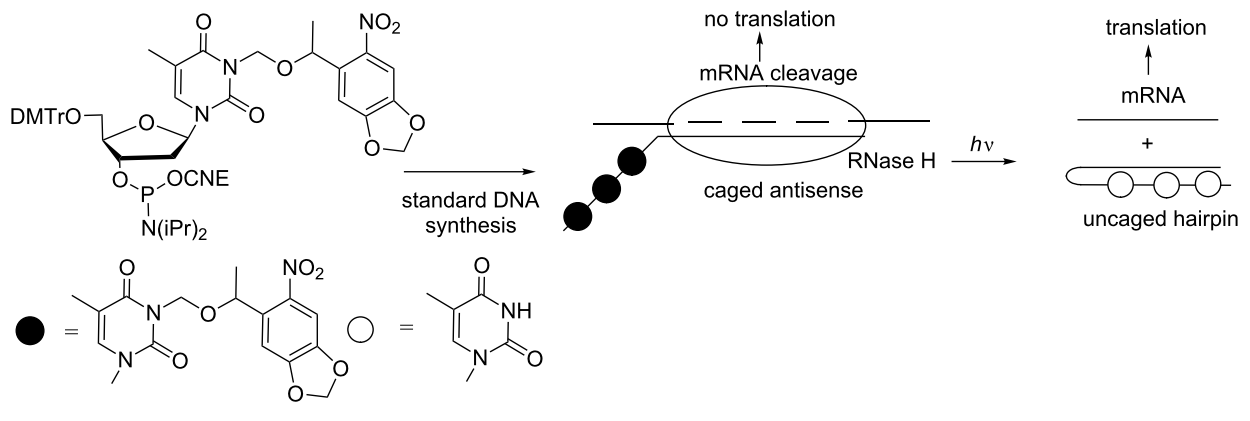
Scheme 20: A) NPP dT and dG phosphoramidites [91,92] and B) NPOM U and G phosphoramidites [83] used to introduce photocaged nucleobases into siRNAs C) close to the argonaute cleavage site to prevent siRNA cleavage [81,83] and D) in the seed region to prevent mRNA recognition by the RISC complex [83].

tary photocaged ON, and the DNAzyme could induce cleavage of a mRNA target. Once deprotected under UV light, this complementary ON hybridized to the catalytic site and inhibited the effect of DNAzyme, allowing mRNA translation.

Photocaged approaches to inhibit transcription: Similarly to antisense and DNAzymes, two similar photocaged approaches have been explored to activate or deactivate triplex-forming ONs (TFOs). These approaches inhibit or elicit gene transcription, respectively [88]. Photocaging of TFOs using NPOM-protected nucleobases prevented the formation of a triple helix with a dsDNA target, consequently permitting gene transcription (Scheme 25). Inversely, when photoirradiation removes the protecting groups, the ON creates a triple helix, hindering gene transcription. By contrast, when the TFO was linked to a caged complementary sequence, the construct could block transcription until photoirradiation led to the formation of the hairpin unable to interact with dsDNA. These photocaged DNAzymes



Scheme 21: Introduction of the photocaged 3-NPOM nucleobase into phosphorothioate antisense and morpholino antisense to inhibit RNA translation though mRNA degradation by RNase H [84] or steric blocking [85].

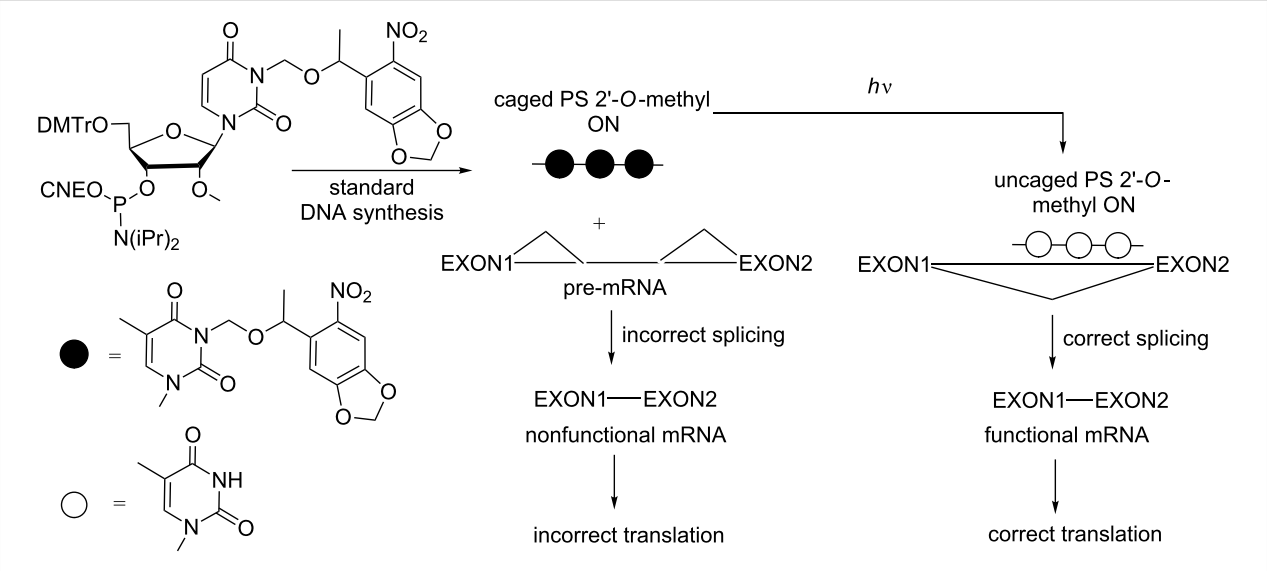


Scheme 22: Control of the activity of an antisense ODN using a photocaged hairpin [82]. Formation of the hairpin suppresses hybridization of the antisense ODN with mRNA, which could be translated.

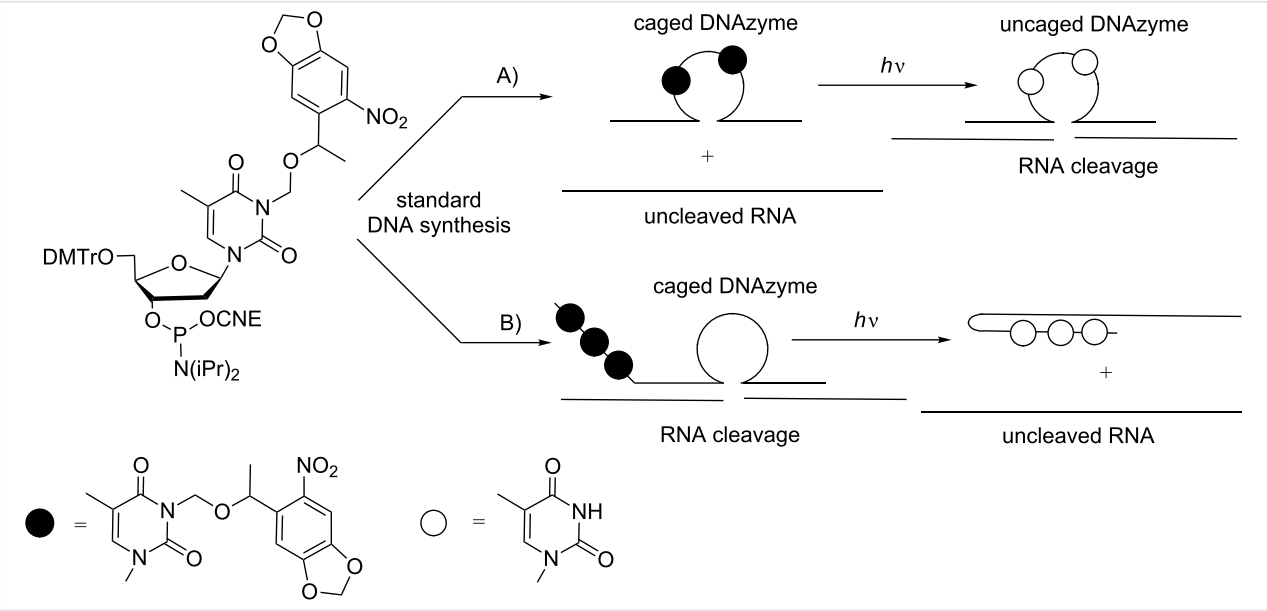
were tested as gene silencing agents to target the reporter gene DsRed in eukaryotic cells.

The first illustration of a photocaged DNA decoy used for the photocontrol of gene expression in mammalian cells was reported in 2011 by Deiters et al [89]. As generally observed, the protecting groups of the nucleobases disturb base pairing that the hairpin decoy could not be formed. The decoy is thus inactive, and the NF-κB transcription factor binds to the NF-κB binding site of an alkaline phosphatase gene to allow transcription. Photodecaging permits hairpin formation, and the active decoy can then bind to NF-κB and compete with the NF-κB binding site of the gene, leading to the inhibition of gene transcription (Scheme 26).

It is noteworthy that the photodeactivation of DNA decoys was also described using a modified photocleavable nucleobase [90]. 7-Nitroindole nucleotides incorporated in a DNA decoy did not suppress hairpin formation so that NF-κB could bind to the decoy (Scheme 27). Under UV irradiation, the nucleobase



Scheme 23: Control of alternative splicing using phosphorothioate (PS) 2'-OMe-photocaged ONs resulting from the incorporation of 3-NPOM 2'-OMe uridine phosphoramidite [86]. Photoirradiation activates the ODN, inducing a correct splicing.



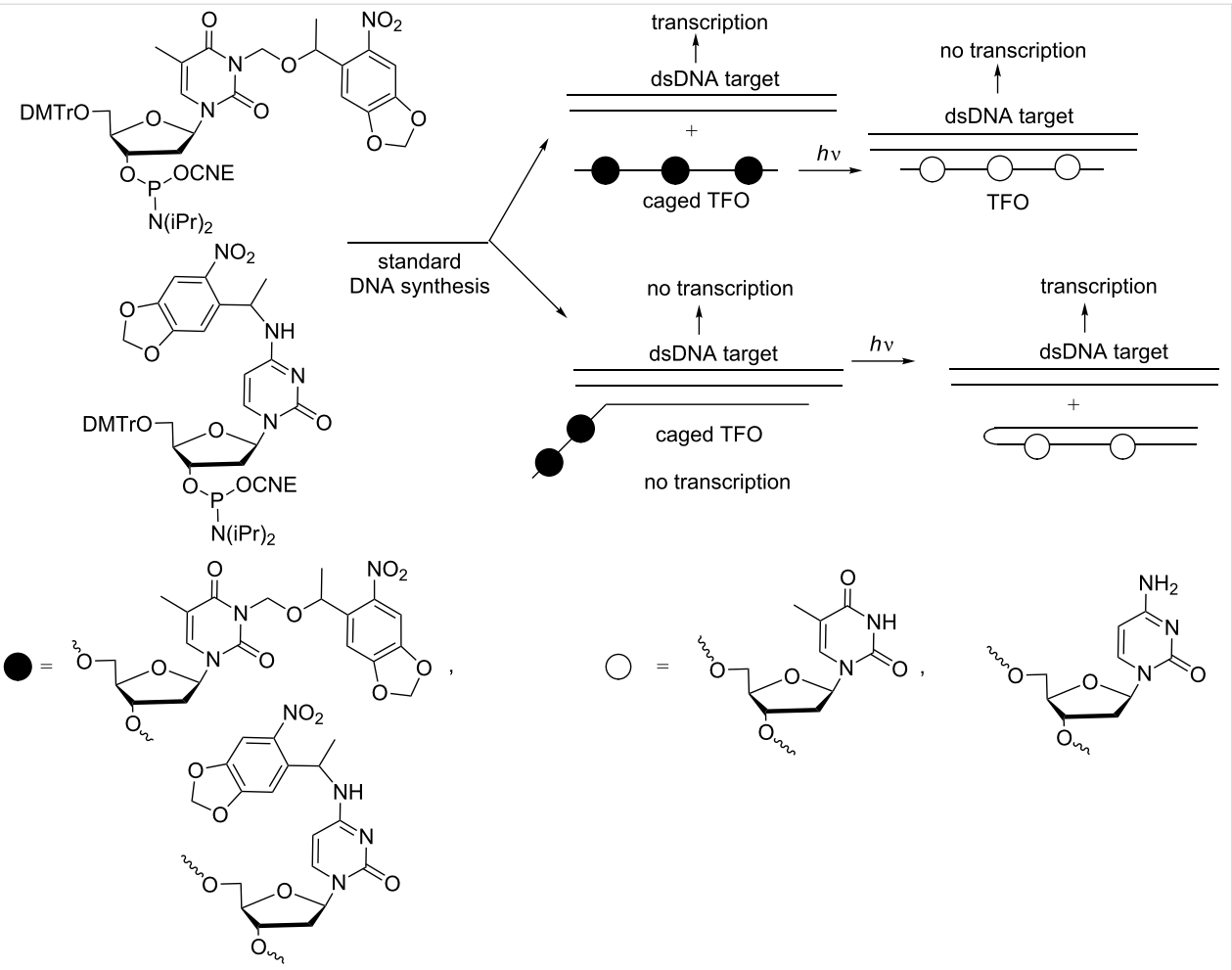
Scheme 24: A) Light activation of a photocaged DNAzyme incorporating 3-NPOM thymidine in its catalytic site [87]; B) light deactivation of a photocaged DNAzyme by formation of an inactive hairpin [82].

was photolyzed, releasing an abasic lactone and lowering the affinity for NF-κB targets. This approach is attractive to "turn on" the transcription upon UV light. However, until now, the effect on gene transcription was not reported.

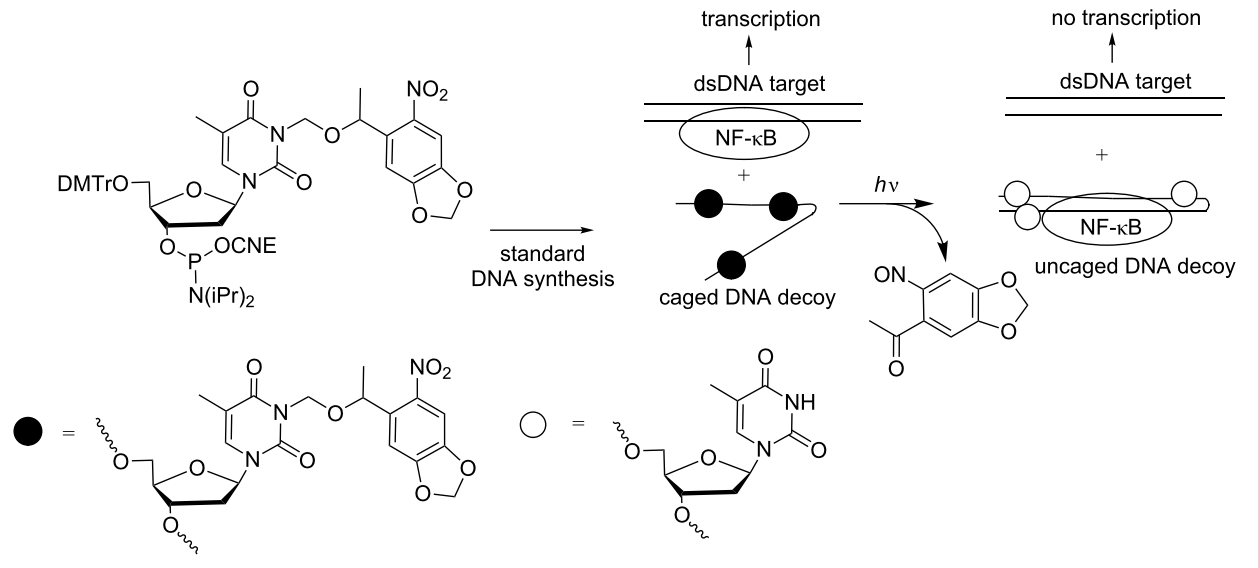
Modifications at the sugar 2'-OH

Light-dependent regulation of gene expression resulting from the interaction of 2'-O-photocaged ONs with the genetic material is not documented compared with ONs modified at phosphates or nucleobases [68]. Generally, what is sought is to suppress the chemical reactivity of this nucleophilic hydroxy function involved in a transesterification reaction that modifies the RNA substrate of the ribozyme but not the catalytic ON itself (Figure 5) [93,94]. This method is inappropriate for potential therapeutic applications. Curiously, to our knowledge, these modifications have not been exploited for the regulation of RNA interference.

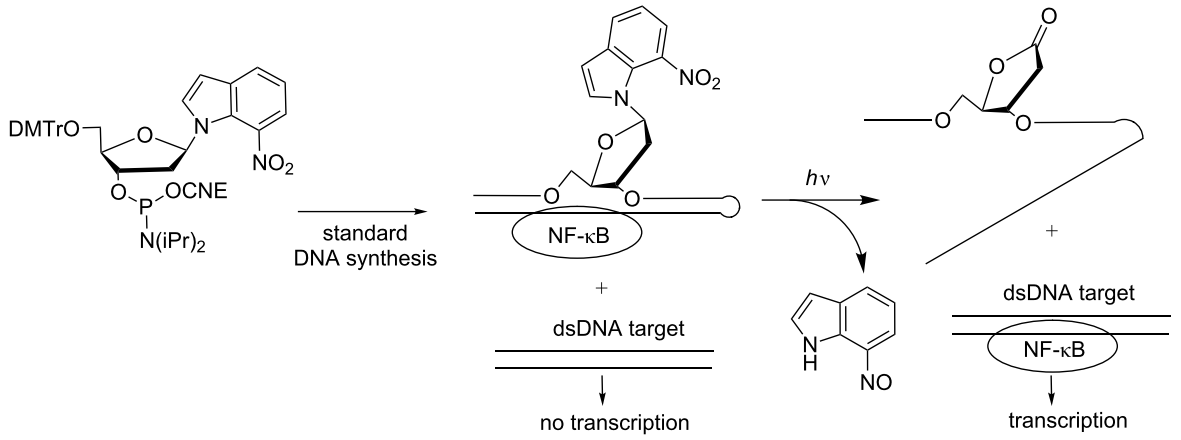
Use of photolabile linkers: For a selected review, see [96]. In this approach, the photolabile moieties are not nucleotide



Scheme 25: Incorporation of 3-(6-nitropiperonyloxymethyl) (NPOM) thymidine and 4-nitropiperonylethyl (NPE) deoxycytidine phosphoramidites into TFOs and light inhibition and light activation of gene transcription using caged TFOs and caged hairpin TFOs, respectively [88].



Scheme 26: Synthesis of a photocaged DNA decoy from a 3-NPOM thymidine phosphoramidite and release of the NPOM protecting group under photolysis, allowing the decoy to organize into a hairpin that can bind to the NF-κB transcription factor [89].



Scheme 27: Synthesis of a caged DNA decoy hairpin containing a 7-nitroindole nucleotide and release of the modified nucleobase under photolysis, leading to an abasic lactone-containing ON that cannot form a hairpin and associate with NF-κB [90].

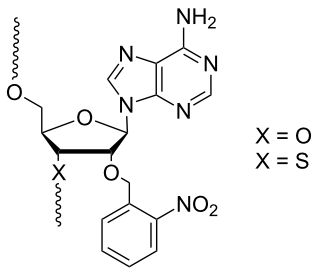
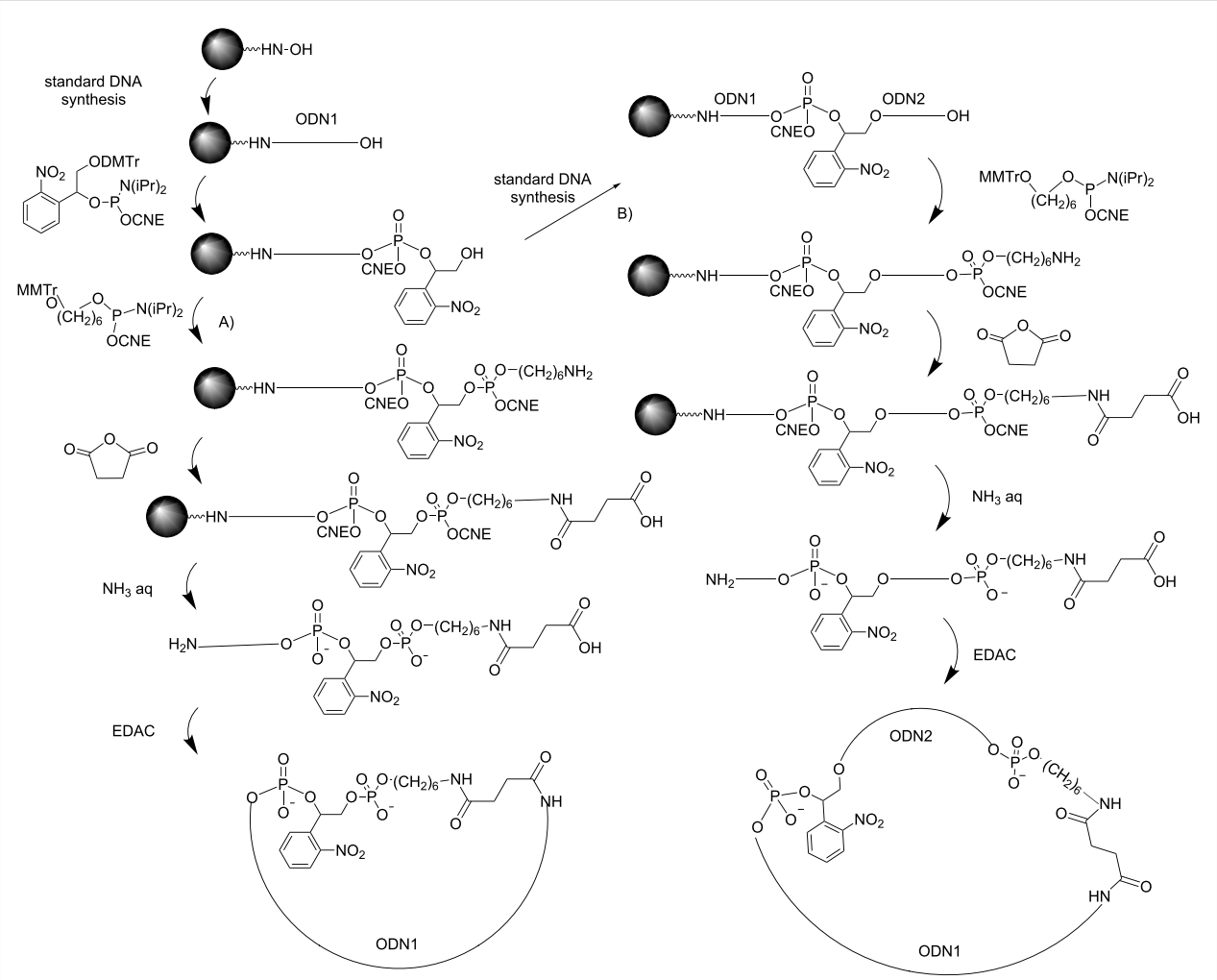


Figure 5: Caged-2'-adenosines used by MacMillan et al [93,94] (X = O) and Piccirilli et al [95] (X = S) to study RNA mechanisms.

protecting groups within ON but are non-nucleoside moieties linking different ONs or both ends of the same ON together or ONs to other molecules. Most frequently, except for circular ONs, photoirradiation cuts the construct into small fragments that induce a change in the biological activity. Photolysis of circular ONs provides linear full-length ONs. Compared to caged nucleobases that directly interact with their nucleic acid targets, photosensitive linkers do not interact but can organize the ONs into specific structures capable of or not interfering with their nucleic targets.

Control of gene expression with photocaged linker-modified ONs has been mostly used for light activation or deactivation of antisense inhibition of RNA translation by Tang and Dmochowski [97-102]. Nevertheless, they were also used to regulate the catalytic effect of DNAzymes [103] and to control alternative splicing as reported by Deiters et al [86].

Two chemical approaches exist to introduce a photoresponsive linker. The first is a post-DNA synthesis process using a heterobifunctional moiety that connects two ONs bearing complementary functionalities. The conjugation of two amino and thiol-terminated ONs with a photoresponsive 2-nitrophenylethanol unit bearing a *N*-hydroxysuccinimide ester and maleimide is an example [97,100]. In the second approach, the linker is incorporated as a phosphoramidite derivative bearing a protected hydroxy function for ON elongation using standard solid-support DNA synthesis [86,103]. This approach is beneficial because several photoactivatable phosphoramidites are commercially available. Beside these two strategies, miscellaneous processes were employed for the synthesis of circular DNA. Dmochowski used the phosphoramidite ligation method between two ONs, and then, the construct was phosphorylated at its 5'-end. After deprotection, the circularization was performed using a single-strand DNA ligase [103]. In 2010, Tang introduced a photoresponsive 1-(2-nitrophenyl)-1,2-ethanediol phosphoramidite at the end of a solid-supported 3'-amino ON (Scheme 28) [101]. This step was followed by the incorporation of an amino-C6-linker phosphoramidite. Before cleavage from the solid support, the 5'-amino functionality was reacted with succinic anhydride, yielding an ON with an amino group at the 3' end and a carboxyl group at the 5' end after deprotection and cleavage from the support (Scheme 28A). Both ends were then chemically linked using water-soluble 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, hydrochloride (EDAC, synthetic yield 20–40%). More recently, the same author followed a quite different approach (Scheme 28B) [104]. After incorporation of the photoresponsive phosphoramidite unit into a 3'-amino solid-supported ON, elongation was ongoing, and then, the aminolinker phosphoramidite was incorporated at the 5'-extremity. The reaction with succinic anhydride followed by the deprotection produced a 5'-carboxyl 3'-amino ON. Both ends, as previously described, were then connected using EDAC with isolated yields of 30–40%.

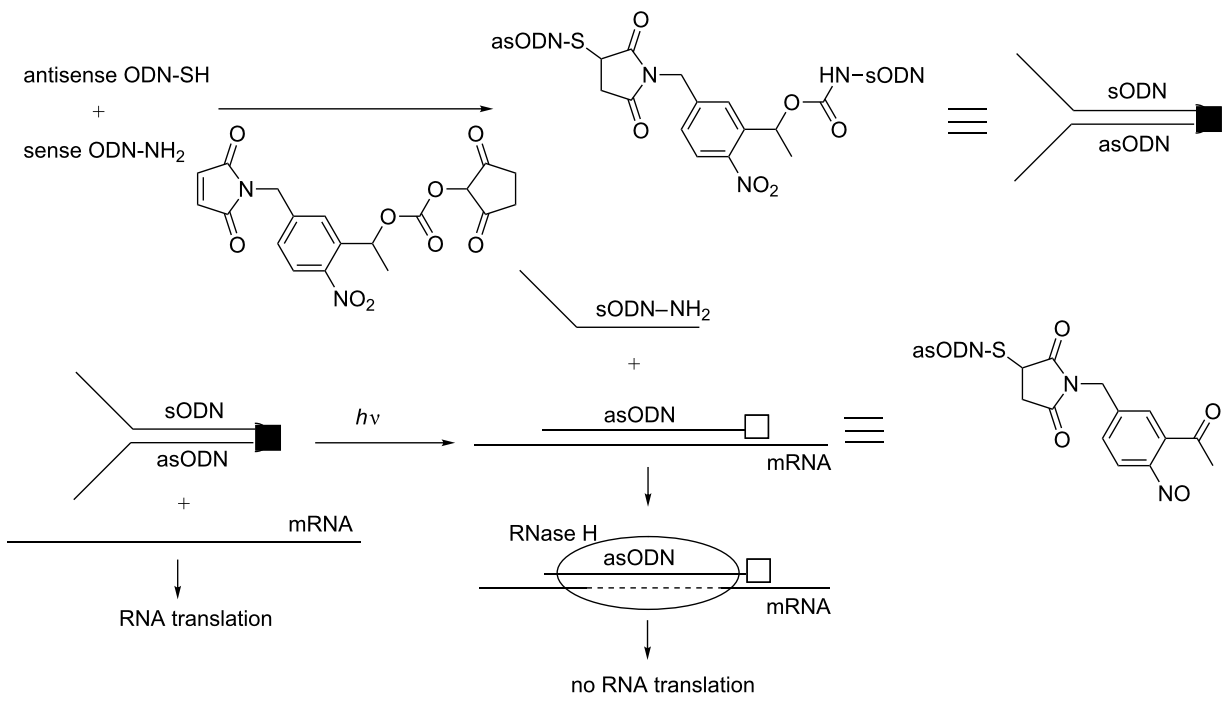


Scheme 28: Synthesis of circular ODNs containing a photolabile linker as described by Tang et al. [101,104].

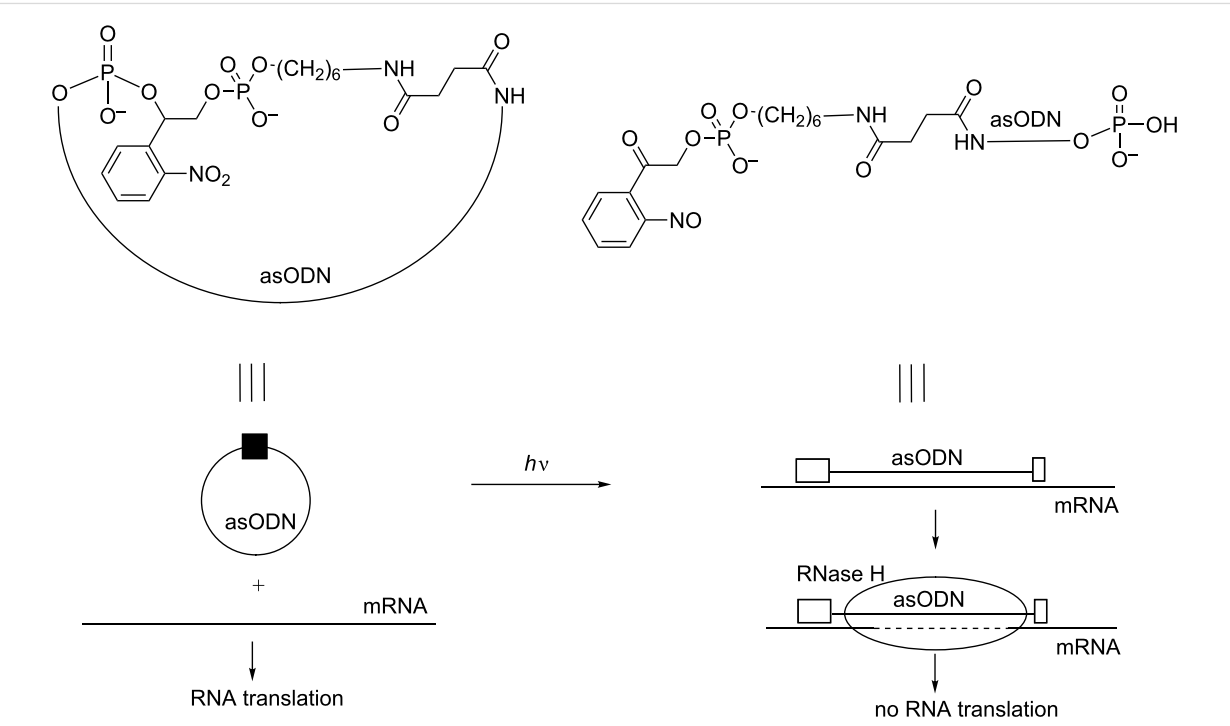
Photocleavable linkers in antisense ONs: Tang and Dmochowski have introduced the 2-nitrophenylethyl-containing linker in the loop of a DNA hairpin where an antisense DNA strand (20-mer) was linked to a shorter complementary ODN (Scheme 29, 12 base pairs). This hairpin was very stable, and the antisense ODN did not hybridize to its RNA target and could not elicit RNA degradation by RNase H. Upon UV irradiation, cleavage of the linker in the hairpin occurred, and the resulting duplex became much less stable, permitting the antisense ODN to hybridize to RNA and turn on its antisense activity [97]. Therefore, while the hairpin induced only 5% degradation of the 15-mer RNA after 1 hour, 66% of RNA degradation was observed upon UV irradiation.

The same authors applied their concept of antisense photocaged DNA hairpins to the inhibition of dC-myb expression in human leukemia cells [105]. The concept was further extended to PNA [99] and morpholino antisense ONs in zebrafish embryos [98] to block physical RNA translation by interaction with the ribosome.

Another method to cage an antisense ODN is to circularize it (Scheme 30) [101]. For this purpose, a single photocleavable linker connected both ends of the ONs as described above. The circular ONs have different lengths, and some of them have a “hairpin-like” or a “dumbbell-like” structure. The circularization of longer ONs (30–40-mers) partially prevented their hybridization to a 40-mer RNA so that RNase H degradation of the RNA target was observed. In this case, photoirradiation at 350 nm activated a 2 to 3-fold increase in RNA degradation by RNase H. A shorter circular ON produced better results because the photocaged ON did not elicit target degradation by RNase H, while photoactivation turned on the antisense activity with a 20-fold increase. The use of circular ONs was further extended by the same author to a steric block GFP RNA translation in transfected HeLa cells by 2’-*O*-methyl phosphorothioate circu-



Scheme 29: Control of RNA digestion with RNase H using light activation of a photocaged hairpin [97].



Scheme 30: Photocontrol of RNA degradation using caged circular antisense ODNs containing a photoresponsive linker [101].

lar ONs [104] and to morpholino-caged ONs in zebrafish embryos to effectively control δ-catenin-2 and no tail gene expression [102].

In the reports cited above, the photocaged ONs are light activated. In the subsequent studies, the photocaged ONs are deactivated by light. As a first example, Dmochowski et al. de-

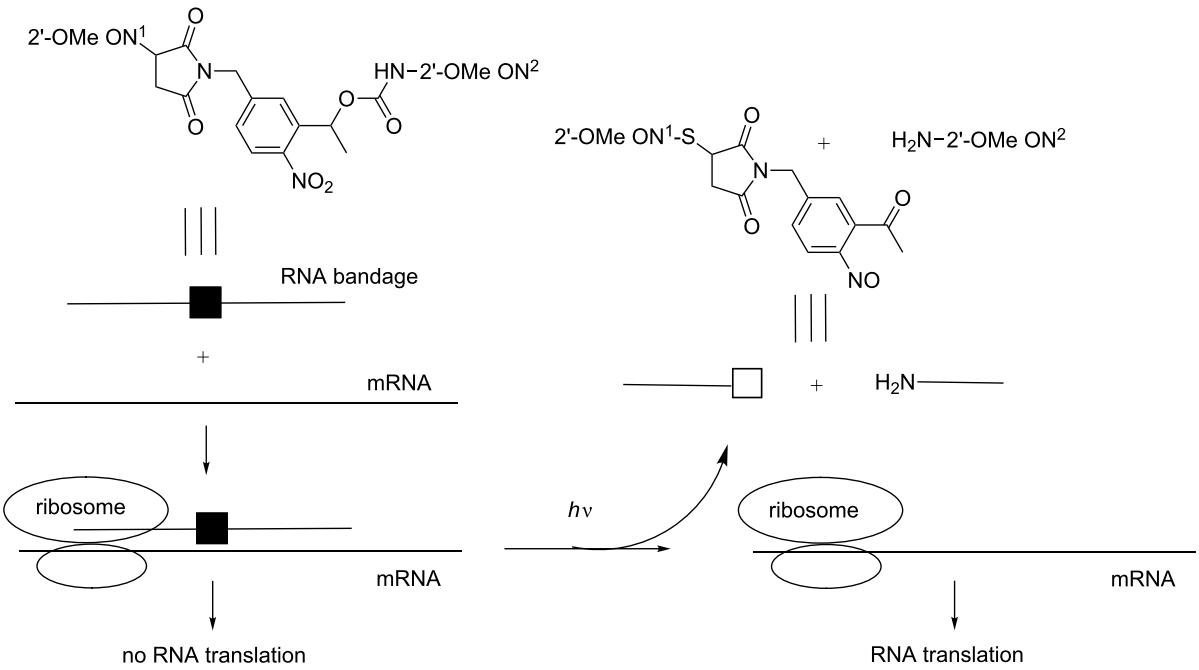
scribed the use of two 6 to 12-mer 2'-OMe RNAs linked together through a photocleavable linker and a 4-base gap in 2008 (Scheme 31). These "RNA bandages" hybridized to an RNA target and blocked its translation. Photoirradiation caused the release of linked short entities that were consequently unable to interact efficiently with the RNA target and obviously blocked its translation [100]. The ability of light to turn off the antisense activity of these "RNA bandages" and to promote gene expression of a GFP transcript was evaluated in rabbit reticulocyte lysates. The most effective photoregulation was obtained using an asymmetric bandage with a short 5' 2'-OMe RNA and a low melting temperature near the start codon linked to a second longer 2'-OMe RNA through the photolabile linker.

Another study relating to light deactivation of a caged ON was reported by Deiters et al., who introduced two photoresponsive *o*-nitrobenzyl linkers into splice-switching ONs (Scheme 32). The use of antisense ONs to correctly aberrant expression during pre-mRNA splicing showed great potential to correct resulting diseases. The photocaged antisense ONs interacted with pre-mRNA and blocked aberrant intron sequences, permitting correct exon splicing and thus correct gene EGFP expression in transfected HeLa cells [86]. Upon UV irradiation, the caged ON fragmented into three shorter pieces, which did not hybridize to pre-mRNA so that the gene was not expressed (on→off effect).

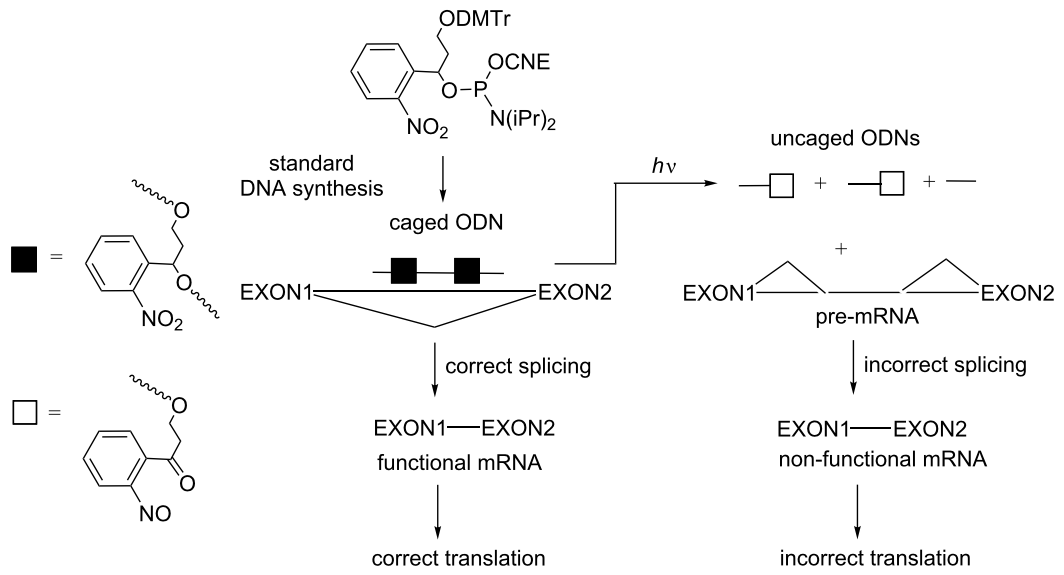
Photocleavable linkers in DNAzymes: Dmochowski et al. have demonstrated that the replacement of thymidine dT8 in the 10–23 DNAzyme with a photocleavable linker introduced as its phosphoramidite in the DNA sequence did not suppress the catalytic effect of the DNAzyme [103]. Unexpectedly, two smaller ONs resulting from cleavage of the linker through photoirradiation also showed a catalytic effect although a reduced one (Scheme 33A). The best difference between the caged DNAzyme and the resulting decaged products was obtained with DNAzyme incorporating two modifications: one in the catalytic site and the other in the recognition site of the DNAzyme. It was argued that in this case, the photolysis produced three ONs, which were too small to hybridize to RNA, and induced its cleavage (on→off effect).

Another approach described in the same article involved a circular DNAzyme incorporating an ON-blocking strand complementary to the recognition site of DNAzyme and joint to the DNAzyme through two linkers at its 5'- and 3'-ends (Scheme 33B). Thus, the DNAzyme was inefficient to hybridize to RNA and, consequently, could not induce its cleavage. Photoirradiation released the free DNAzyme, which then induced the catalytic cleavage of RNA (off→on effect).

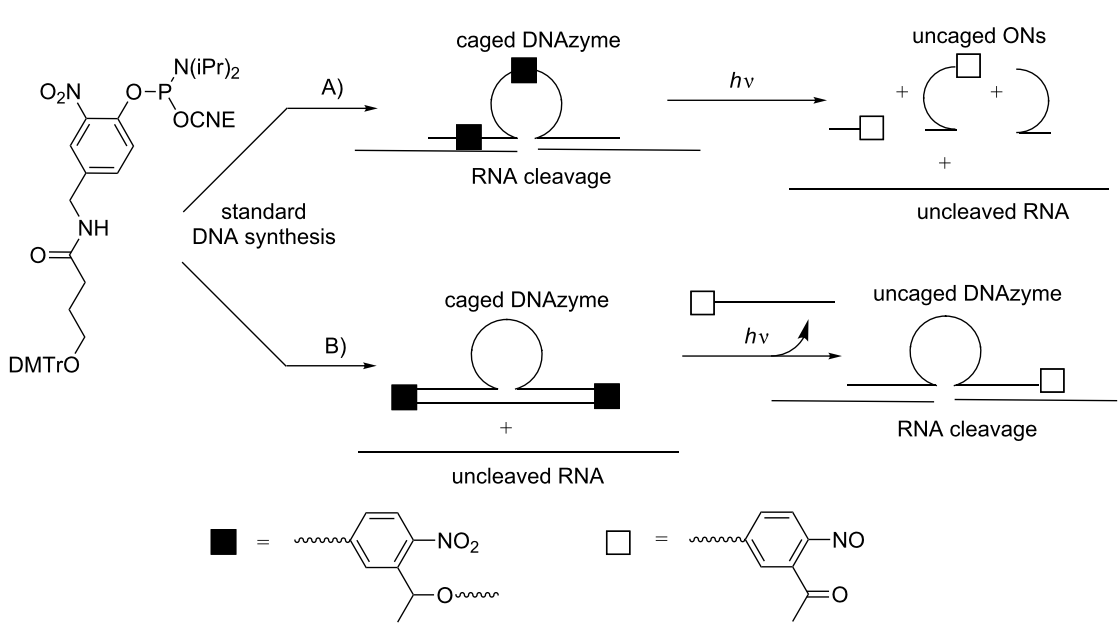
Photocleavable linkers in siRNA conjugates: Tang et al. have described the control of RNAi in HEK293 cells using



Scheme 31: Control of RNA translation using an "RNA bandage" consisting of two short 2'-OMe ONs linked together with a photosensitive linker [100].



Scheme 32: Control of alternative splicing using photocaged ONs resulting from the incorporation of an *o*-nitrobenzyl responsive moiety as its phosphoramidite [86]. Photoirradiation deactivates the ODN, inducing incorrect splicing.

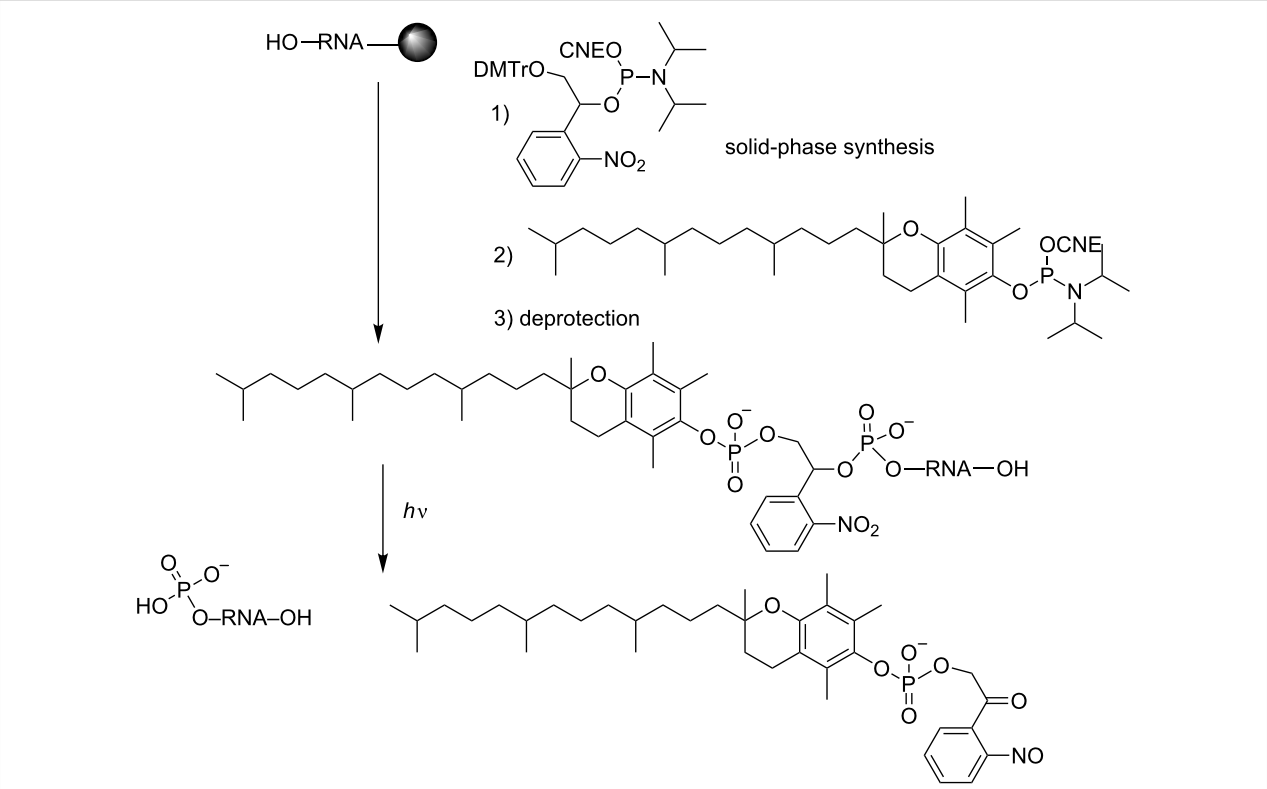


Scheme 33: A) Light deactivation of a photocaged DNAzyme incorporating one photocleavable spacer in its catalytic site and another in the recognition site; B) light activation of a circular photocaged DNAzyme formed through the hybridization and ligation of the DNAzyme with a complementary strand [103].

photocaged siRNAs conjugated with a 5'-terminal vitamin E (vit E) through a photolabile linker and a 4-base gap [106]. Both, the linker and vit E were introduced into siRNAs using their corresponding phosphoramidites (Scheme 34). In this concept, the photoresponsive unit did not directly interfere with the biological activity of the photocaged conjugate. However, vit E, which interacted with the binding protein targets, prevented the association of ON with the RNAi machinery. The photolysis released ON from the vitamin, and siRNA activity was activated.

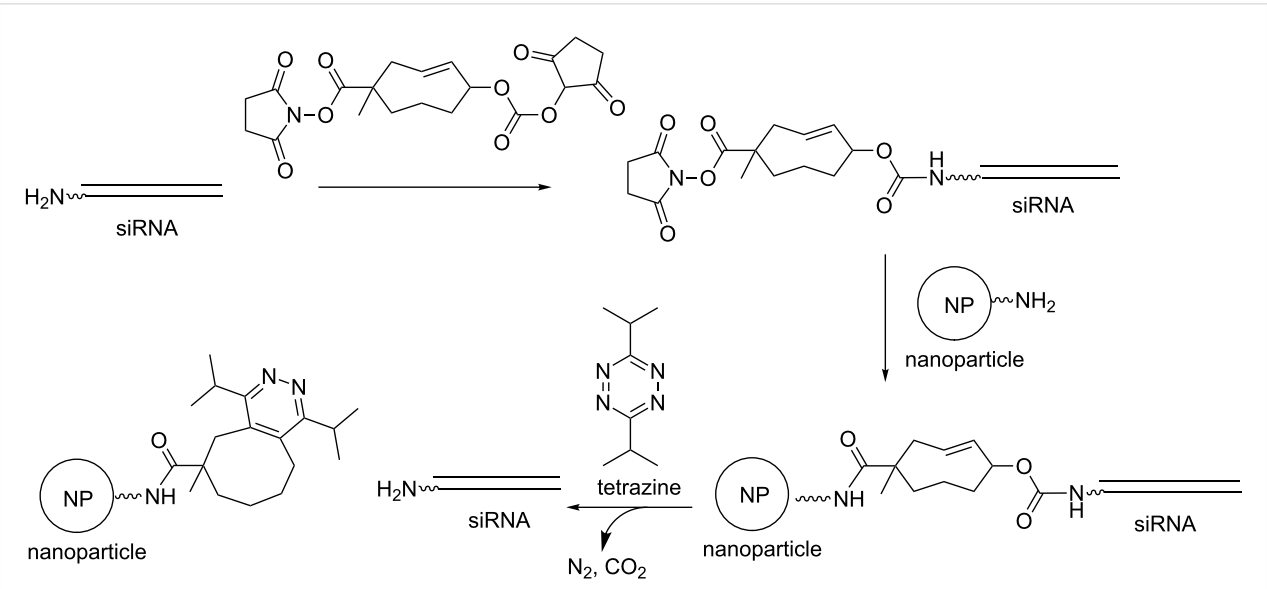
Chemical-responsive ONs

Light or heat is an external physical regulatory element compared to glutathione, for example, which is an internal chemical



Scheme 34: Solid-phase synthesis of a caged vit E-siRNA conjugate and its release upon UV irradiation [106].

regulatory element, or carboxyesterases and reductases, which are internal biochemical regulatory stimuli. The use of an external chemical factor to trigger the activity of ON prodrugs has been rarely reported in the literature. Recently, however, Royzen reported such an approach to control in-cell siRNA activity [107]. To this end, 3'-amino siRNA was linked to amino-functionalized nanoparticles (NP) through a bifunctional trans-cyclooctene heterolinker (Scheme 35). These conjugates cannot



Scheme 35: Synthesis of a siRNA conjugated to a nanoparticle (NP) via a cyclooctene heterolinker from a siRNA-NH_2 and an NP-NH_2 [107]. The conjugate does not induce gene silencing until tetrazine triggers siRNA release.

interfere with RISC and do not allow gene silencing until tetrazine releases the ON from the nanoparticle by an inverse-electron demand Diels–Alder reaction with biocompatible tetrazine. The gene silencing of exogenous GFP and endogenous CDK8 genes in MDA-MB-231 breast cancer cells was demonstrated.

Conclusion

The interest in stimuli-responsive ONs to control gene expression has increased in recent years. This prodrug approach, as most of the permanent ON modifications, aims to overcome the limitations of ONs due to their poor extracellular and intracellular stability, low efficiency of intracellular delivery to target cells or tissues and possible off-target gene silencing, immunostimulation and other side effects. However, for stimuli-responsive ONs, the desired effect is that of "natural" ONs obtained after transformation in response to a stimulus that may be internal or external, biochemical, chemical or physical. Compared to permanent modifications, transient modifications have the great advantage to regulate the activity of ONs as a function of stimuli acting as switches.

Most of the examples of stimuli applicable to ON prodrugs have been gathered in this review. Physical stimuli such as heat and light can be easily controlled by the operator, whereas biochemical stimuli such as enzymes act on a difference between the contents of the intracellular and the extracellular compartment. Creative and ingenious chemistry was used to design all these stimuli-responsive modifications, most of which have been evaluated at least in vitro and some of which seemed promising. Nevertheless, among the stimuli-responsive ONs described in this review, most of them have been tested in cellulo on reporter gene models except for a few studies on specific genes in embryos for some photocaged ONs [85,102]. In addition, it is noteworthy that for the first time, a biological effect was measured in mice with siRNA prodrugs containing charge-neutralizing phosphotriester linkages [43] and these data are promising for ON prodrug-based approaches. The numerous literature references on light-responsive ONs compared to other stimuli-responsive ONs deserve to be highlighted to show how much effort was put on this subject during this last decade. Indeed, this may be explained by the fact that photoirradiation is the major and the simplest method to control the response of caged ONs both, in time and in space.

We hope this review provides insight into the available transient modifications to make efficient ON prodrugs. To date, the successful approach to obtain ON therapeutics based on a prodrug strategy remains unresolved, but the recent report on an example of a chemical external stimulus opens an exciting future in the prodrug field [107]. The abbreviations used in this review are listed in Table 1.

Table 1: List of abbreviations.

abbreviation	full length
A	adenine
Ac	acetyl
AMPrOM	2-amino-2-methylpropionyloxymethyl
AMEBuOM	2-aminomethyl-2-ethyl-butyryloxymethyl
AON	antisense oligonucleotide
Boc	*tert*-butoxycarbonyl
Bn	benzyl
$BuNH_2$	butylamine
C	cytosine
CNE	cyanoethyl
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DMTr	dimethoxytrityl
DMTSF	dimethyl(methylthio)sulfonium tetrafluoroborate
dT	thymidine
DMNPE	dimethoxynitrophenylethyl
CD-DMNPE	cyclododecyl-DMNPE
EDAC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
Et	ethyl

Table 1: List of abbreviations. (continued)

fma	2-(*N*-formyl-*N*-methyl) aminoethyl
Fmoc	fluorenylmethoxycarbonyl
G	guanine
GFP	green fluorescent protein
GMEBuOM	2-guanidinomethyl-2-ethyl-butyryloxymethyl
GSH	glutathione
HIV	human immunodeficiency virus
iPr	isopropyl
iPrPac	isopropylphenoxyacetyl
Lev	levulinyl
Me	methyl
MDTM	methyldithiomethyl
miRNA	micro ribonucleic acid
MMTr	monomethoxytrityl
MOE	methoxyethyl
NADH	nicotinamide adenine dinucleotide
NB	nitrobenzyl
NP	nanoparticule
NPE	4-nitropiperonylethyl
NPOM	6-nitropiperonyloxymethyl
NPP	2-(2-nitrophenyl)propyl
ON	oligonucleotide
ODN	oligodeoxyribonucleotide
Pac	phenoxyacetyl
PiBuOM	phenylisobutyryloxymethyl
PivOM	pivaloyloxymethyl
PNA	peptide nucleic acid
PrOM	propionyloxymethyl
PS	phosphorothioate
psc	phenylsulfonylcarbamoyl
Q-linker	hydroquinone-*O,O'*-diacetic acid
RNAi	RNA interference
RNN	ribonucleic neutral
RSSM	alkyldithiomethyl
A-SATE	aldehyde SATE
Me-SATE	S-acetylthioethyl
t-Bu-SATE	*S*-pivaloylthioethyl
siRNA	small interfering ribonucleic acid
T	thymine
TAR	trans-activation response
TAT	transactivator of transcription
TBA	thrombin-binding DNA aptamer
TBDMS	*tert*-butyldimethylsilyl
Tc-DNA	tricyclo-DNA
tc^{ee}-T	ethyl tricyclo-thymine
tc^{hd}-T	hexadecyl tricyclo-thymine
TEEP	thioether-enol phosphodiester
TPP	triphenylphosphonium
U	uracil
vit E	vitamin E

Acknowledgements

We acknowledge the assistance of Marie-Christine Bergogne who helped with the research bibliography.

ORCID® iDs

Françoise Debart - https://orcid.org/0000-0003-3422-3926
Jean-Jacques Vasseur - https://orcid.org/0000-0002-4379-6139

References

1. Lundin, K. E.; Gissberg, O.; Smith, C. I. E. *Hum. Gene Ther.* **2015,** *26,* 475–485. doi:10.1089/hum.2015.070
2. Sharma, V. K.; Watts, J. K. *Future Med. Chem.* **2015,** *7,* 2221–2242. doi:10.4155/fmc.15.144
3. Stein, C. A.; Castanotto, D. *Mol. Ther.* **2017,** *25,* 1069–1075. doi:10.1016/j.ymthe.2017.03.023
4. Crooke, S. T. *Nucleic Acid Ther.* **2017,** *27,* 70–77. doi:10.1089/nat.2016.0656
5. Müller, S. *Molecules* **2017,** *22,* No. 789. doi:10.3390/molecules22050789
6. Fokina, A. A.; Stetsenko, D. A.; François, J. C. *Expert Opin. Biol. Ther.* **2015,** *15,* 689–711. doi:10.1517/14712598.2015.1025048
7. Chakraborty, C.; Sharma, A. R.; Sharma, G.; Doss, C. G. P.; Lee, S.-S. *Mol. Ther. - Nucleic Acids* **2017,** *8,* 132–143. doi:10.1016/j.omtn.2017.06.005
8. Bernardo, B. C.; Ooi, J. Y.; Lin, R. C.; McMullen, J. R. *Future Med. Chem.* **2015,** *7,* 1771–1792. doi:10.4155/fmc.15.107
9. Deleavey, G. F.; Damha, M. J. *Chem. Biol.* **2012,** *19,* 937–954. doi:10.1016/j.chembiol.2012.07.011
10. Shukla, S.; Sumaria, C. S.; Pradeepkumar, P. I. *ChemMedChem* **2010,** *5,* 328–349. doi:10.1002/cmdc.200900444
11. Selvam, C.; Mutisya, D.; Prakash, S.; Ranganna, K.; Thilagavathi, R. *Chem. Biol. Drug Des.* **2017,** *90,* 665–678. doi:10.1111/cbdd.12993
12. Albert, A. *Nature* **1958,** *182,* 421–423. doi:10.1038/182421a0
13. Saneyoshi, H.; Kondo, K.; Sagawa, N.; Ono, A. *Bioorg. Med. Chem. Lett.* **2016,** *26,* 622–625. doi:10.1016/j.bmcl.2015.11.064
14. Hayashi, J.; Samezawa, Y.; Ochi, Y.; Wada, S.-i.; Urata, H. *Bioorg. Med. Chem. Lett.* **2017,** *27,* 3135–3138. doi:10.1016/j.bmcl.2017.05.031
15. Biscans, A.; Rouanet, S.; Vasseur, J.-J.; Dupouy, C.; Debart, F. *Org. Biomol. Chem.* **2016,** *14,* 7010–7017. doi:10.1039/C6OB01272H
16. Ochi, Y.; Imai, M.; Nakagawa, O.; Hayashi, J.; Wada, S.-i.; Urata, H. *Bioorg. Med. Chem. Lett.* **2016,** *26,* 845–848. doi:10.1016/j.bmcl.2015.12.074
17. Ochi, Y.; Nakagawa, O.; Sakaguchi, K.; Wada, S.-i.; Urata, H. *Chem. Commun.* **2013,** *49,* 7620–7622. doi:10.1039/c3cc43725f
18. Semenyuk, A.; Földesi, A.; Johansson, T.; Estmer-Nilsson, C.; Blomgren, P.; Brännvall, M.; Kirsebom, L. A.; Kwiatkowski, M. *J. Am. Chem. Soc.* **2006,** *128,* 12356–12357. doi:10.1021/ja0636587
19. Chen, Q.; Butler, D.; Querbes, W.; Pandey, R. K.; Ge, P.; Maier, M. A.; Zhang, L.; Rajeev, K. G.; Nechev, L.; Kotelianski, V.; Manoharan, M.; Sah, D. W. Y. *J. Controlled Release* **2010,** *144,* 227–232. doi:10.1016/j.jconrel.2010.02.011
20. Mehiri, M.; Upert, G.; Tripathi, S.; Di Giorgio, A.; Condom, R.; Pandey, V. N.; Patino, N. *Oligonucleotides* **2008,** *18,* 245–256. doi:10.1089/oli.2008.0126
21. Barber, I.; Rayner, B.; Imbach, J.-L. *Bioorg. Med. Chem. Lett.* **1995,** *5,* 563–568. doi:10.1016/0960-894X(95)00074-4
22. Iyer, R. P.; Yu, D.; Agrawal, S. *Bioorg. Chem.* **1995,** *23,* 1–21. doi:10.1006/bioo.1995.1001
23. Iyer, R. P.; Yu, D.; Devlin, T.; Ho, N.-h.; Agrawal, S. *Bioorg. Med. Chem. Lett.* **1996,** *6,* 1917–1922. doi:10.1016/0960-894X(96)00332-0
24. Tosquellas, G.; Barber, I.; Morvan, F.; Rayner, B.; Imbach, J.-L. *Bioorg. Med. Chem. Lett.* **1996,** *6,* 457–462. doi:10.1016/0960-894X(96)00051-0
25. Zhang, N.; Tan, C.; Cai, P.; Zhang, P.; Zhao, Y.; Jiang, Y. *Chem. Commun.* **2009,** 3216–3218. doi:10.1039/b903331a
26. Saneyoshi, H.; Iketani, K.; Kondo, K.; Saneyoshi, T.; Okamoto, I.; Ono, A. *Bioconjugate Chem.* **2016,** *27,* 2149–2156. doi:10.1021/acs.bioconjchem.6b00368
27. Saneyoshi, H.; Hiyoshi, Y.; Iketani, K.; Kondo, K.; Ono, A. *Bioorg. Med. Chem. Lett.* **2015,** *25,* 5632–5635. doi:10.1016/j.bmcl.2015.10.025
28. Ikeda, M.; Kamimura, M.; Hayakawa, Y.; Shibata, A.; Kitade, Y. *ChemBioChem* **2016,** *17,* 1304–1307. doi:10.1002/cbic.201600164
29. Tosquellas, G.; Alvarez, K.; Dell'Aquila, C.; Morvan, F.; Vasseur, J.-J.; Imbach, J.-L.; Rayner, B. *Nucleic Acids Res.* **1998,** *26,* 2069–2074. doi:10.1093/nar/26.9.2069
30. Iyer, R. P.; Ho, N.-h.; Yu, D.; Agrawal, S. *Bioorg. Med. Chem. Lett.* **1997,** *7,* 871–876. doi:10.1016/S0960-894X(97)00119-4
31. Mauritz, R. P.; Schmelz, F. S.; Meier, C. *Nucleosides Nucleotides* **1999,** *18,* 1417–1418. doi:10.1080/07328319908044737
32. Poijarvi-Virta, P.; Lönnberg, H. *Curr. Med. Chem.* **2006,** *13,* 3441–3465. doi:10.2174/092986706779010270
33. Vivès, E.; Dell'Aquila, C.; Bologna, J.-C.; Morvan, F.; Rayner, B.; Imbach, J.-L. *Nucleic Acids Res.* **1999,** *27,* 4071–4076. doi:10.1093/nar/27.20.4071
34. Bologna, J.-C.; Vivès, E.; Imbach, J.-L. *Antisense Nucleic Acid Drug Dev.* **2004,** *12,* 33–41. doi:10.1089/108729002753670247
35. Mignet, N.; Tosquellas, G.; Barber, I.; Morvan, F.; Rayner, B.; Imbach, J. L. *New J. Chem.* **1997,** *21,* 73–79.
36. Lioux, T.; Lefebvre, I.; Vasseur, J. J.; Imbach, J. L. *Rapid Commun. Mass Spectrom.* **1999,** *13,* 1645–1649. doi:10.1002/(SICI)1097-0231(19990830)13:16<1645::AID-RCM692>3.0.CO;2-Y
37. Bres, J.-C.; Morvan, F.; Lefebvre, I.; Vasseur, J.-J.; Pompon, A.; Imbach, J.-L. *J. Chromatogr., B* **2001,** *753,* 123–130. doi:10.1016/S0378-4347(00)00502-8
38. Alvarez, K.; Vasseur, J.-J.; Beltran, T.; Imbach, J.-L. *J. Org. Chem.* **1999,** *64,* 6319–6328. doi:10.1021/jo990479h
39. Spinelli, N.; Meyer, A.; Hayakawa, Y.; Imbach, J.-L.; Vasseur, J.-J. *Eur. J. Org. Chem.* **2002,** 49–56. doi:10.1002/1099-0690(20021)2002:1<49::AID-EJOC49>3.0.CO;2-7
40. Guerlavais-Dagland, T.; Meyer, A.; Imbach, J.-L.; Morvan, F. *Eur. J. Org. Chem.* **2003,** 2327–2335. doi:10.1002/ejoc.200300069
41. Poijarvi, P.; Heinonen, P.; Virta, P.; Lönnberg, H. *Bioconjugate Chem.* **2005,** *16,* 1564–1571. doi:10.1021/bc050143g
42. Leisvuori, A.; Lönnberg, H.; Ora, M. *Eur. J. Org. Chem.* **2014,** 5816–5826. doi:10.1002/ejoc.201402412
43. Meade, B. R.; Gogoi, K.; Hamil, A. S.; Palm-Apergi, C.; van den Berg, A.; Hagopian, J. C.; Springer, A. D.; Eguchi, A.; Kacsinta, A. D.; Dowdy, C. F.; Presente, A.; Lönn, P.; Kaulich, M.; Yoshioka, N.; Gros, E.; Cui, X.-S.; Dowdy, S. F. *Nat. Biotechnol.* **2014,** *32,* 1256–1261. doi:10.1038/nbt.3078

44. Ducho, C. *ChemMedChem* **2015,** *10,* 1625–1627. doi:10.1002/cmdc.201500279
45. Khvorova, A.; Osborn, M. F.; Hassler, M. R. *Nat. Biotechnol.* **2014,** *32,* 1197–1198. doi:10.1038/nbt.3091
46. Parey, N.; Baraguey, C.; Vasseur, J.-J.; Debart, F. *Org. Lett.* **2006,** *8,* 3869–3872. doi:10.1021/ol0616182
47. Martin, A. R.; Lavergne, T.; Vasseur, J.-J.; Debart, F. *Bioorg. Med. Chem. Lett.* **2009,** *19,* 4046–4049. doi:10.1016/j.bmcl.2009.06.015
48. Lavergne, T.; Baraguey, C.; Dupouy, C.; Parey, N.; Wuensche, W.; Sczakiel, G.; Vasseur, J.-J.; Debart, F. *J. Org. Chem.* **2011,** *76,* 5719–5731. doi:10.1021/jo200826h
49. Biscans, A.; Bos, M.; Martin, A. R.; Ader, N.; Sczakiel, G.; Vasseur, J.-J.; Dupouy, C.; Debart, F. *ChemBioChem* **2014,** *15,* 2674–2679. doi:10.1002/cbic.201402382
50. Pon, R. T.; Yu, S. *Nucleic Acids Res.* **1997,** *25,* 3629–3635. doi:10.1093/nar/25.18.3629
51. Kempe, T. Anhydrous amine cleavage of oligonucleotides. US Patent 5750672 A, May 12, 1998.
52. Baraguey, C.; Lescrinier, E.; Lavergne, T.; Debart, F.; Herdewijn, P.; Vasseur, J.-J. *Org. Biomol. Chem.* **2013,** *11,* 2638–2647. doi:10.1039/c3ob27005j
53. Biscans, A.; Bertrand, J.-R.; Dubois, J.; Rüger, J.; Vasseur, J.-J.; Sczakiel, G.; Dupouy, C.; Debart, F. *ChemBioChem* **2016,** *17,* 2054–2062. doi:10.1002/cbic.201600317
54. Biscans, A.; Rouanet, S.; Bertrand, J.-R.; Vasseur, J.-J.; Dupouy, C.; Debart, F. *Bioorg. Med. Chem.* **2015,** *23,* 5360–5368. doi:10.1016/j.bmc.2015.07.054
55. Johnsson, R.; Lackey, J. G.; Bogojeski, J. J.; Damha, M. J. *Bioorg. Med. Chem. Lett.* **2011,** *21,* 3721–3725. doi:10.1016/j.bmcl.2011.04.073
56. Lietard, J.; Leumann, C. J. *J. Org. Chem.* **2012,** *77,* 4566–4577. doi:10.1021/jo300648u
57. Cieślak, J.; Grajkowski, A.; Livengood, V.; Beaucage, S. L. *J. Org. Chem.* **2004,** *69,* 2509–2515. doi:10.1021/jo035861f
58. Grajkowski, A.; Pedras-Vasconcelos, J.; Wang, V.; Ausín, C.; Hess, S.; Verthelyi, D.; Beaucage, S. L. *Nucleic Acids Res.* **2005,** *33,* 3550–3560. doi:10.1093/nar/gki657
59. Grajkowski, A.; Ausín, C.; Kauffman, J. S.; Snyder, J.; Hess, S.; Lloyd, J. R.; Beaucage, S. L. *J. Org. Chem.* **2007,** *72,* 805–815. doi:10.1021/jo062087y
60. Ausín, C.; Kauffman, J. S.; Duff, R. J.; Shivaprasad, S.; Beaucage, S. L. *Tetrahedron* **2010,** *66,* 68–79. doi:10.1016/j.tet.2009.10.096
61. Grajkowski, A.; Cieslak, J.; Gapeev, A.; Beaucage, S. L. *New J. Chem.* **2010,** *34,* 880–887. doi:10.1039/b9nj00692c
62. Jain, H. V.; Takeda, K.; Tami, C.; Verthelyi, D.; Beaucage, S. L. *Bioorg. Med. Chem.* **2013,** *21,* 6224–6232. doi:10.1016/j.bmc.2013.04.071
63. Kiuru, E.; Ahmed, Z.; Lönnberg, H.; Beigelman, L.; Ora, M. *J. Org. Chem.* **2013,** *78,* 950–959. doi:10.1021/jo302421u
64. Kiuru, E.; Lönnberg, H.; Ora, M. *Helv. Chim. Acta* **2013,** *96,* 1997–2008. doi:10.1002/hlca.201300028
65. Ohkubo, A.; Kasuya, R.; Miyata, K.; Tsunoda, H.; Seio, K.; Sekine, M. *Org. Biomol. Chem.* **2009,** *7,* 687–694. doi:10.1039/b816831h
66. Saneyoshi, H.; Shimamura, K.; Sagawa, N.; Ando, Y.; Tomori, T.; Okamoto, I.; Ono, A. *Bioorg. Med. Chem. Lett.* **2015,** *25,* 2129–2132. doi:10.1016/j.bmcl.2015.03.064
67. Deiters, A. *Curr. Opin. Chem. Biol.* **2009,** *13,* 678–686. doi:10.1016/j.cbpa.2009.09.026
68. Tang, X. J.; Dmochowski, I. J. *Mol. BioSyst.* **2007,** *3,* 100–110. doi:10.1039/B614349K
69. Casey, J. P.; Blidner, R. A.; Monroe, W. T. *Mol. Pharmaceutics* **2009,** *6,* 669–685. doi:10.1021/mp900082q
70. Ceo, L. M.; Koh, J. T. *ChemBioChem* **2012,** *13,* 511–513. doi:10.1002/cbic.201100683
71. Ikeda, M.; Kabumoto, M. *Chem. Lett.* **2017,** *46,* 634–640. doi:10.1246/cl.161063
72. Shah, S.; Rangarajan, S.; Friedman, S. H. *Angew. Chem., Int. Ed.* **2005,** *44,* 1328–1332. doi:10.1002/anie.200461458
73. Nguyen, Q. N.; Chavli, R. V.; Marques, J. T.; Conrad, P. G.; Wang, D.; He, W.; Belisle, B. E.; Zhang, A.; Pastor, L. M.; Witney, F. R.; Morris, M.; Heitz, F.; Divita, G.; Williams, B. R. G.; McMaster, G. K. *Biochim. Biophys. Acta, Biomembr.* **2006,** *1758,* 394–403. doi:10.1016/j.bbamem.2006.01.003
74. Shah, S.; Jain, P. K.; Kala, A.; Karunakaran, D.; Friedman, S. H. *Nucleic Acids Res.* **2009,** *37,* 4508–4517. doi:10.1093/nar/gkp415
75. Kala, A.; Friedman, S. H. *Pharm. Res.* **2011,** *28,* 3050–3057. doi:10.1007/s11095-011-0529-z
76. Jain, P. K.; Shah, S.; Friedman, S. H. *J. Am. Chem. Soc.* **2011,** *133,* 440–446. doi:10.1021/ja107226e
77. Wang, X.; Feng, M.; Xiao, L.; Tong, A.; Xiang, Y. *ACS Chem. Biol.* **2016,** *11,* 444–451. doi:10.1021/acschembio.5b00867
78. Wu, L.; Pei, F.; Zhang, J.; Wu, J.; Feng, M.; Wang, Y.; Jin, H.; Zhang, L.; Tang, X. *Chem. – Eur. J.* **2014,** *20,* 12114–12122. doi:10.1002/chem.201403430
79. Buff, M.; Mack, T.; Heckel, A. *Chimia* **2009,** *63,* 261–264. doi:10.2533/chimia.2009.261
80. Liu, Q.; Deiters, A. *Acc. Chem. Res.* **2014,** *47,* 45–55. doi:10.1021/ar400036a
81. Mikat, V.; Heckel, A. *RNA* **2007,** *13,* 2341–2347. doi:10.1261/rna.753407
82. Young, D. D.; Lively, M. O.; Deiters, A. *J. Am. Chem. Soc.* **2010,** *132,* 6183–6193. doi:10.1021/ja100710j
83. Govan, J. M.; Young, D. D.; Lusic, H.; Liu, Q.; Lively, M. O.; Deiters, A. *Nucleic Acids Res.* **2013,** *41,* 10518–10528. doi:10.1093/nar/gkt806
84. Young, D. D.; Lusic, H.; Lively, M. O.; Yoder, J. A.; Deiters, A. *ChemBioChem* **2008,** *9,* 2937–2940. doi:10.1002/cbic.200800627
85. Deiters, A.; Garner, R. A.; Lusic, H.; Govan, J. M.; Dush, M.; Nascone-Yoder, N. M.; Yoder, J. A. *J. Am. Chem. Soc.* **2010,** *132,* 15644–15650. doi:10.1021/ja1053863
86. Hemphill, J.; Liu, Q.; Uprety, R.; Samanta, S.; Tsang, M.; Juliano, R. L.; Deiters, A. *J. Am. Chem. Soc.* **2015,** *137,* 3656–3662. doi:10.1021/jacs.5b00580
87. Lusic, H.; Young, D. D.; Lively, M. O.; Deiters, A. *Org. Lett.* **2007,** *9,* 1903–1906. doi:10.1021/ol070455u
88. Govan, J. M.; Uprety, R.; Hemphill, J.; Lively, M. O.; Deiters, A. *ACS Chem. Biol.* **2012,** *7,* 1247–1256. doi:10.1021/cb300161r
89. Govan, J. M.; Lively, M. O.; Deiters, A. *J. Am. Chem. Soc.* **2011,** *133,* 13176–13182. doi:10.1021/ja204980v
90. Struntz, N. B.; Harki, D. A. *ACS Chem. Biol.* **2016,** *11,* 1631–1638. doi:10.1021/acschembio.6b00130
91. Kröck, L.; Heckel, A. *Angew. Chem., Int. Ed.* **2005,** *44,* 471–473. doi:10.1002/anie.200461779
92. Mayer, G.; Kröck, L.; Mikat, V.; Engeser, M.; Heckel, A. *ChemBioChem* **2005,** *6,* 1966–1970. doi:10.1002/cbic.200500198
93. Chaulk, S. G.; MacMillan, A. M. *Nucleic Acids Res.* **1998,** *26,* 3173–3178. doi:10.1093/nar/26.13.3173

94. Chaulk, S. G.; MacMillan, A. M. *Nat. Protoc.* **2007,** *2,* 1052–1058. doi:10.1038/nprot.2007.154
95. Li, N.-S.; Tuttle, N.; Staley, J. P.; Piccirilli, J. A. *J. Org. Chem.* **2014,** *79,* 3647–3652. doi:10.1021/jo4028374
96. Ruble, B. K.; Yeldell, S. B.; Dmochowski, I. J. *J. Inorg. Biochem.* **2015,** *150,* 182–188. doi:10.1016/j.jinorgbio.2015.03.010
97. Tang, X.; Dmochowski, I. J. *Angew. Chem., Int. Ed.* **2006,** *45,* 3523–3526. doi:10.1002/anie.200600954
98. Shestopalov, I. A.; Sinha, S.; Chen, J. K. *Nat. Chem. Biol.* **2007,** *3,* 650–651. doi:10.1038/nchembio.2007.30
99. Tang, X.; Maegawa, S.; Weinberg, E. S.; Dmochowski, I. J. *J. Am. Chem. Soc.* **2007,** *129,* 11000–11001. doi:10.1021/ja073723s
100. Richards, J. L.; Tang, X. J.; Turetsky, A.; Dmochowski, I. J. *Bioorg. Med. Chem. Lett.* **2008,** *18,* 6255–6258. doi:10.1016/j.bmcl.2008.09.093
101. Tang, X. J.; Su, M.; Yu, L. L.; Lv, C.; Wang, J.; Li, Z. J. *Nucleic Acids Res.* **2010,** *38,* 3848–3855. doi:10.1093/nar/gkq079
102. Wang, Y.; Wu, L.; Wang, P.; Lv, C.; Yang, Z.; Tang, X. *Nucleic Acids Res.* **2012,** *40,* 11155–11162. doi:10.1093/nar/gks840
103. Richards, J. L.; Seward, G. K.; Wang, Y.-H.; Dmochowski, I. J. *ChemBioChem* **2010,** *11,* 320–324. doi:10.1002/cbic.200900702
104. Wu, L.; Wang, Y.; Wu, J.; Lv, C.; Wang, J.; Tang, X. *Nucleic Acids Res.* **2013,** *41,* 677–686. doi:10.1093/nar/gks996
105. Tang, X.; Swaminathan, J.; Gewirtz, A. M.; Dmochowski, I. J. *Nucleic Acids Res.* **2008,** *36,* 559–569. doi:10.1093/nar/gkm1029
106. Ji, Y.; Yang, J.; Wu, L.; Yu, L.; Tang, X. *Angew. Chem., Int. Ed.* **2016,** *55,* 2152–2156. doi:10.1002/anie.201510921
107. Khan, I.; Seebald, L. M.; Robertson, N. M.; Yigit, M. V.; Royzen, M. *Chem. Sci.* **2017,** *8,* 5705–5712. doi:10.1039/C7SC01380A

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Enzyme-free genetic copying of DNA and RNA sequences

Marilyne Sosson and Clemens Richert*§

Review

Open Access

Address:
Institute of Organic Chemistry, University of Stuttgart, 70569 Stuttgart, Germany

Email:
Clemens Richert* - lehrstuhl-2@oc.uni-stuttgart.de

* Corresponding author
§ Tel: +49 (0)711 685-64311, Fax: +49 (0)711/685-64321

Keywords:
base pairing; DNA; enzyme-free primer extension; nucleotides; oligonucleotides; replication; RNA

Beilstein J. Org. Chem. **2018,** *14,* 603–617.
doi:10.3762/bjoc.14.47

Received: 17 December 2017
Accepted: 20 February 2018
Published: 12 March 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

The copying of short DNA or RNA sequences in the absence of enzymes is a fascinating reaction that has been studied in the context of prebiotic chemistry. It involves the incorporation of nucleotides at the terminus of a primer and is directed by base pairing. The reaction occurs in aqueous medium and leads to phosphodiester formation after attack of a nucleophilic group of the primer. Two aspects of this reaction will be discussed in this review. One is the activation of the phosphate that drives what is otherwise an endergonic reaction. The other is the improved mechanistic understanding of enzyme-free primer extension that has led to a quantitative kinetic model predicting the yield of the reaction over the time course of an assay. For a successful modeling of the reaction, the strength of the template effect, the inhibitory effect of spent monomers, and the rate constants of the chemical steps have to be determined experimentally. While challenges remain for the high fidelity copying of long stretches of DNA or RNA, the available data suggest that enzyme-free primer extension is a more powerful reaction than previously thought.

Introduction

Replication of genetic information is critical for all living systems. In the cell, this process is catalyzed by enzymatic machineries that have polymerases at their core [1]. Polymerases catalyze not only the replication of DNA, but are also involved in repair and transcription of genes [2]. Considering that enzymes catalyze processes that lead to protein synthesis, it is reasonable to ask what started replication when life emerged on planet Earth. A solution to the chicken/egg dilemma of replication might be found in RNA, as oligo- and polyribonucleotides can encode genetic information and can catalyze biochemical reactions as ribozymes. More than 30 years ago, it was observed that RNA strands catalyze splicing or ligation of longer oligonucleotides [3,4]. Ancient ribozymes might have acted as polymerases [5], inducing either the oligomerization of activated ribonucleotides or the replication of the first RNA genomes. But ribozymes are usually too long to be likely to emerge from random sequences in one step. Simple forms may have taken advantage of the high ionic strength of the eutectic phase [6], but their evolution must have been preceded by something simples. In a very simple version of RNA-based

replication, genetic copying may have occurred in the absence of both protein enzymes and ribozymes, relying on solely on base pairing for molecular recognition and chemical reactivity to drive the formation of phosphodiester bonds in aqueous media. This is what is usually referred to as "enzyme-free copying" (Figure 1).

Studies on enzyme-free copying of genetic polymers date back more than 50 years [7]. Classical studies were often focused on ligation reactions, including templated ligations of self-complementary sequences [8,9]. Special systems, such as ligation with triplex-forming sequences [10] have produced some impressive results, and the field of ligation-based replication has been reviewed [11]. Ligation reactions will not be discussed further here, as they are limited in their scope, in terms of sequences, whereas monomer-based copying may be used for any given sequence, at least in principle. Rather, we will focus on copying with mononucleotides, for which early examples can also be found in the literature of the 1960s [12]. The early monomer-based work on copying RNA focused on oligomerization of nucleotides on homosequences as templates [13,14]. The best results were observed for poly(C) as template, the 2-methylimidazolide of guanosine as activated monomer (Figure 2), and assay buffers containing high concentrations of Mg^{2+} ions [15]. When advances in automated solid-phase synthesis made oligonucleotides of any given sequence readily available [16], copying reactions involving the extension of a primer bound to a specific sequence of hairpins mimicking this arrangement became the most common way of performing the reaction [17-20].

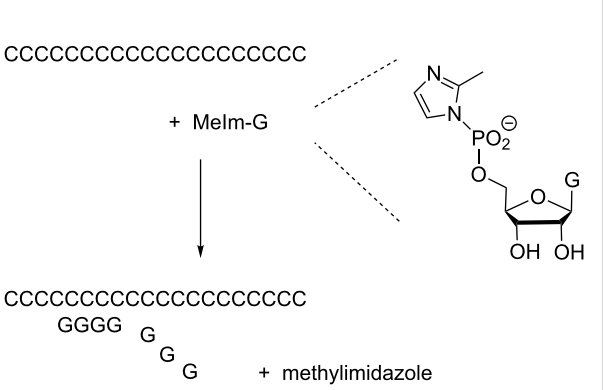


Figure 2: Oligomerization of the 2-methylimidazolide of guanosine-5'-monophosphate on a poly(C) template.

In this brief account we will focus on primer extension reactions on DNA and RNA templates. The copying of DNA sequences is usually performed with primers terminating in a 3'-amino-2',3'-dideoxynucleoside. The amino group is much more nucleophilic than the hydroxy group of natural DNA, so that rapid reactions result. Figure 3 shows the structure of the phosphoramidate formed when 3'-aminoterminal DNA primers are extended, together with the phosphoramidate linkage resulting from reactions with 3'-aminoribonucleotides [21], and the two regioisomeric phosphodiesters that result from the extension of RNA primers that terminate in natural ribonucleosides. We note that the phosphoramidate linkages are isoelectronic and largely isosteric to natural phosphodiesters.

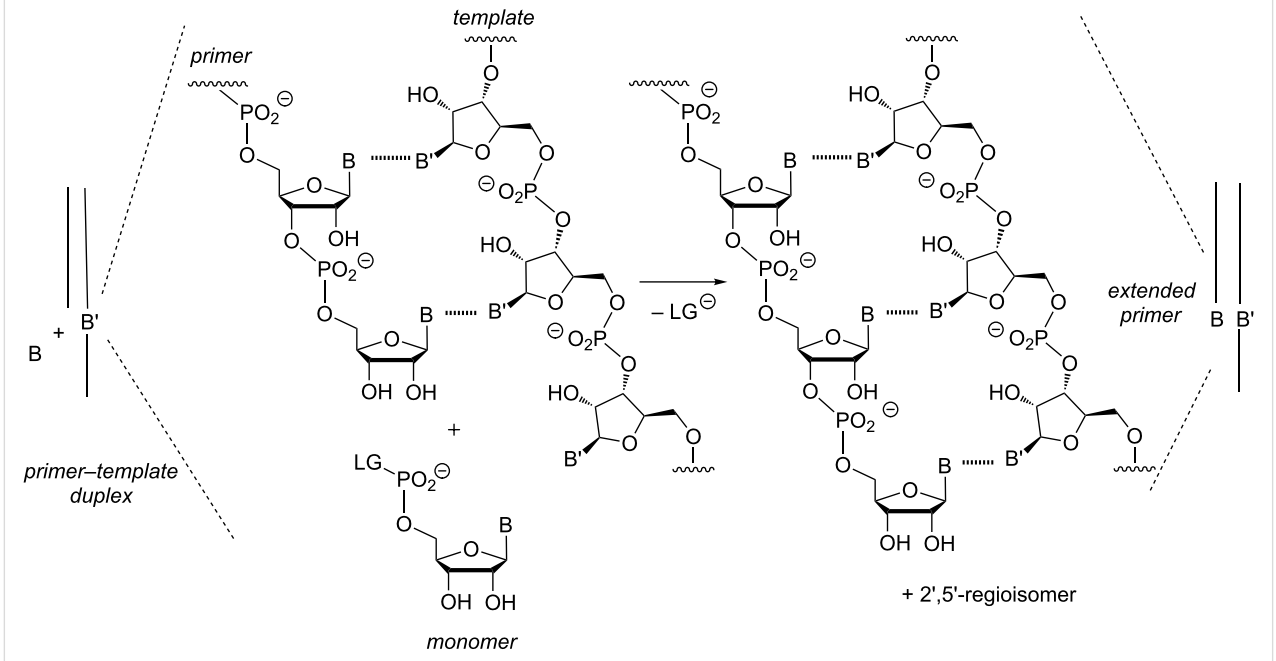


Figure 1: Enzyme-free template-directed extension of an RNA primer by one nucleotide. B = nucleobase, LG = leaving group.

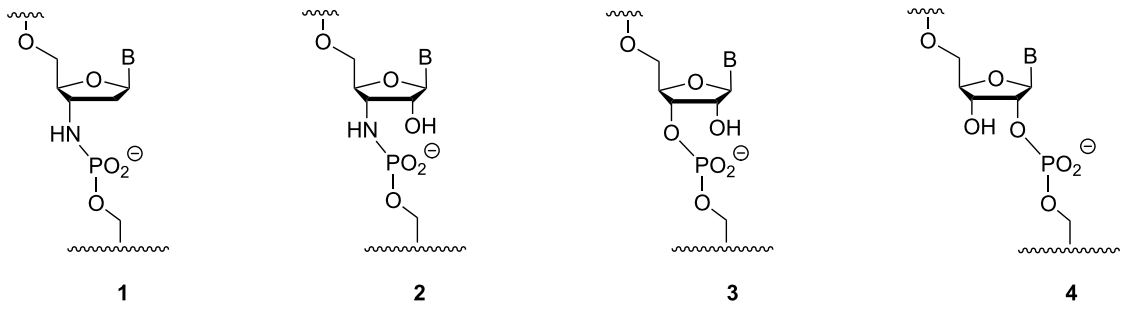


Figure 3: Structures of backbone linkages produced in enzyme-free primer extension reactions: the phosphoramidate of a 3'-amino-2',3'-dideoxynucleoside (**1**), the phosphoramidate of a 3'-amino-3'-deoxynucleoside (**2**), the 3',5'-phosphodiester of a natural ribonucleoside (**3**), and the isomeric 2',5'-phosphodiester of a ribonucleoside (**4**).

In our brief account, we will highlight some of the issues plaguing enzyme-free primer extension. One such issue is incomplete conversion. Many chemical primer extension assays stall long before completion of the reaction, resulting in a mixture of extended and unextended primer. We will then discuss progress in our understanding of the chemical primer extension reaction that was made since our earlier account on the topic [22]. Other reviews that cover enzyme-free copying exist, and the reader is directed to these papers for a more in-depth treatment of issues only touched upon in our account [13,23-25].

Review

Template effect and sequence dependence

One factor that significantly affects whether an enzyme-free primer extension reaction occurs in high yield or not is the strength of the template effect. Unlike the reactions that are catalyzed by polymerases, purely chemical primer extension reactions are not facilitated by the active sites of enzymes. Instead, the base pairing between individual bases of an incoming nucleotide and the templating base must suffice to attract the monomer to the extension site. The stability of different base pairs varies, and so does the templating effect of different stretches of the template sequence. Using random homopolymer templates, Joyce and Orgel concluded that the structure and hybridization status of templates was important for high-yielding copying reactions [26]. In a later series of papers with specific, synthetic sequences, Wu and Orgel reported that primer extension proceeds poorly if too many weakly pairing A or U residues are present in the templating sequence [17-19].

These experiments had been performed with a riboterminal primer/self-priming hairpin that is low in reactivity. Using more nucleophilic 3'-aminoterminal DNA primers and oxyazabenzotriazolides of deoxynucleotides (OAt-dNMPs) as a more reactive combination than that of the traditional methylimidazolides and RNA primers, we screened all 64 possible base triplets at the elongation site [27] (Figure 4). Both, the base at the center position of the triplet that acts as templating base and either of the flanking bases were varied systematically, and downstream-binding oligonucleotides were tested for their effect. Under these conditions, 90% of the primer was extended successfully in each of the 64 different sequence contexts.

This suggested that the template effect is strong enough to support successful copying, at least when sufficient reactivity exists. When we determined the rates for each of the different

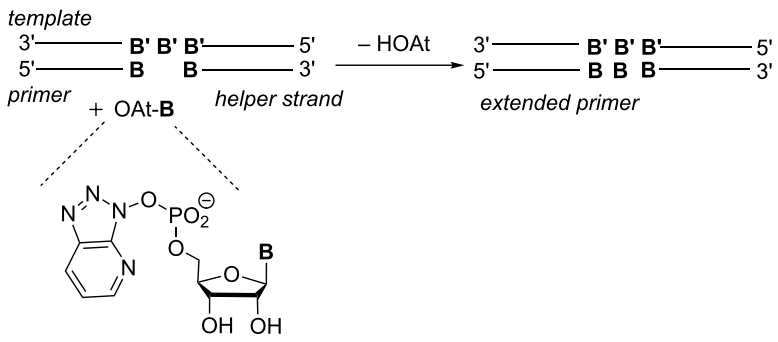


Figure 4: System used for studying the template effect with all 64 possible triplets at the extension site (B/B' = nucleobase).

templating triplets, we found that the rate constant for extension on the poorest templating sequence (CAG) and on the best templating sequence (TCT) differed by less than two orders of magnitude, with rate constants k'_{CAG} = 100 h^{-1} M^{-1} and k'_{TCT} = 8 310 h^{-1} M^{-1} [27]. This was encouraging. As expected, the incorporation of G was most favorable, as this base strongly pairs via three hydrogen bonds and has a large surface area for stacking. Numerically, the $t_{1/2}$ values for the incorporation of G ranged from 1 min to 15 min, whereas those for T were between 13 min and approx. 2 h under the experimental conditions chosen. Further, a primer terminating in an A residue was found to be favorable. This, the most lipophilic of the bases, probably offers the stickiest stacking surface for incoming bases. When a downstream-binding oligonucleotide is present, stacking with the base of its 5'-terminal nucleoside further adds to the attractive forces experienced by incoming monomers. This is shown schematically in Figure 5. Because downstream-binding strands favorably affect the rate and selectivity of primer extension, we have dubbed them "helper oligonucleotides" [28,29]. Kinetics measured without downstream-binding element were two- to seven-fold slower, depending on the sequence context [27].

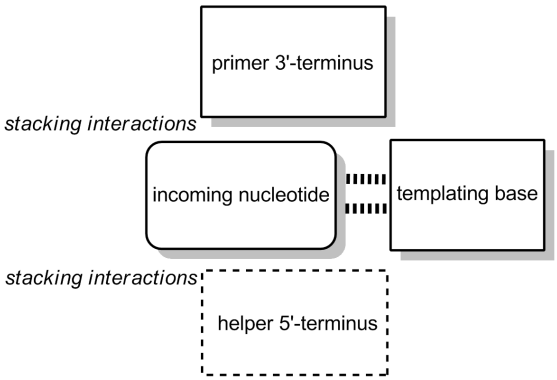


Figure 5: Interactions attracting the incoming nucleotide to the extension site. Besides base pairing via hydrogen bonding to the templating base, stacking interactions with neighboring bases of primer and a possible downstream-binding helper oligonucleotide, as well as help of solvophobic effects, influence the strength of the template effect.

Overall, the data mentioned above suggest that there is indeed a strong dependence of the templating base and sequence context on the rate of enzyme-free primer extension assays. Whether the available template effect suffices to induce successful extension in aqueous buffer depends on the reactivity of the nucleophilic group at the primer's 3'-terminus and that of the activated phosphate of the monomer.

Quantitative model

To gain a better understanding of the factors responsible for successful or unsuccessful primer extension assays, we embarked on a project aimed at gaining a quantitative understanding of enzyme-free copying. What are the rate- and yield-limiting steps of the reaction? What concentration of the monomer is needed to achieve near-quantitative conversion? Are there other factors that need to be considered to be able to predict the yield of primer extension reactions? These were just some of the questions that motivated this work. We wished to know what the fate of the many nucleotides was that were employed in the assay (usually in large excess over the primer). Figure 6 shows three of the more obvious reaction pathways that came to mind.

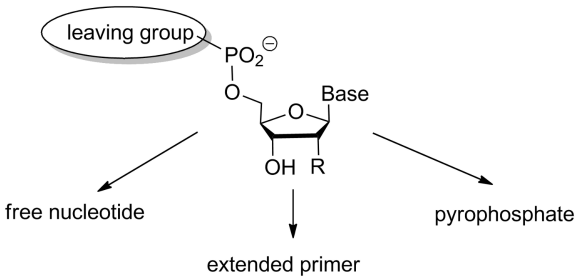


Figure 6: Three possible fates of activated nucleotides in aqueous buffer that result from hydrolysis, primer extension, and reaction with a free nucleotide, respectively. Other possible pathways, such as cyclization to the 3',5'-cyclic diester or oligomerization are not shown graphically; R is OH for ribonucleotides and is H for deoxynucleotides.

Our experimental work used nucleotides pre-activated as oxyazabenzotriazolides (OAt esters, compare Figure 4) [28,29] or as 2-methylimidazolides (MeIm amides, compare Figure 2) [13,22,30]. In aqueous media, hydrolysis of activated nucleotides is all but unavoidable, and hydrolytic deactivation becomes more likely when significant concentrations of magnesium ions are present [31]. High initial concentrations of monomers are usually used to compensate for this problem (0.1 M solutions are not uncommon), but still there is incomplete conversion for extensions that involve incorporation of A or U [19].

Further, it was clear that monomers have to bind to the primer-template duplex prior to experience the template effect and to be incorporated sequence specifically. So, a quantitative understanding of the binding equilibrium was called for. Bimolecular binding equilibria are usually described mathematically via the binding constant or dissociation constant. The latter is more intuitive, as it gives the concentration at which half of the binding partners are in the bound state and the other half is in the free state in an equimolar mixture of the two.

Next, it had become clear from our study on RNA-based copying that the hydrolysis of activated monomers not only

reduces the amount of available starting material, but actively lowers reactivity because the hydrolyzed monomer can inhibit primer extension. The hydrolyzed, free nucleotide can still bind to the extension site on the template, and in doing so, prevent the activated form from entering the site, acting as a competitive inhibitor [32]. So, both the rate of hydrolysis and the strength of the inhibitory effect were important factors to be considered in a quantitative model.

Figure 7 shows our model for an RNA-based primer extension system. For primer extension to occur, binding between activated nucleotide and primer–template duplex takes place. So, the dissociation constant (K_d) has to be determined experimentally [33]. Once bound, the terminal hydroxy group of the primer has to attack the activated 5'-phosphate of the primer, most likely producing a pentavalent intermediate. Unless the leaving group finds itself in the proper apical position of the intermediate, this is followed by pseudorotation and then the release of the leaving group. Either of these steps can be rate-limiting, and we have encountered two-step kinetics with a lag-phase in some reactions involving aminoterminal primers [27]. More often, though, and in all cases involving ribonucleosides at the 3'-terminus of the primer, kinetics characteristic of a single rate-limiting step are found, so that the modeling requires no more than a single rate constant for the covalent step (k_{cov}). Determining the rate constant experimentally requires knowledge of K_d, so that a defined concentration of the kinetically relevant species can be entered in the rate equation for what is now a pseudo-first order reaction [33,34]. To properly model the inhibition, both the rate of hydrolysis (k_h) and the dissociation constant of the inhibitor–primer/template complex have to be known. The latter (K_{dh}) is often similar to the K_d value for the complex with the activated monomer, so that an approximation assuming this, produces results that are not far off from what modeling with all four constants (K_d, K_{dh}, k_h and k_{cov}) gives [34].

Hydrolysis of activated nucleotides

In order to gain any insights from the model presented in the preceding paragraph, binding constants and rate constants had to be determined. Among the rate constants was that for the hydrolysis of activated monomers. Hydrolysis was expected to be fast for highly reactive monomers, and the reactivity toward water was expected to be similar to the reactivity toward the ter-

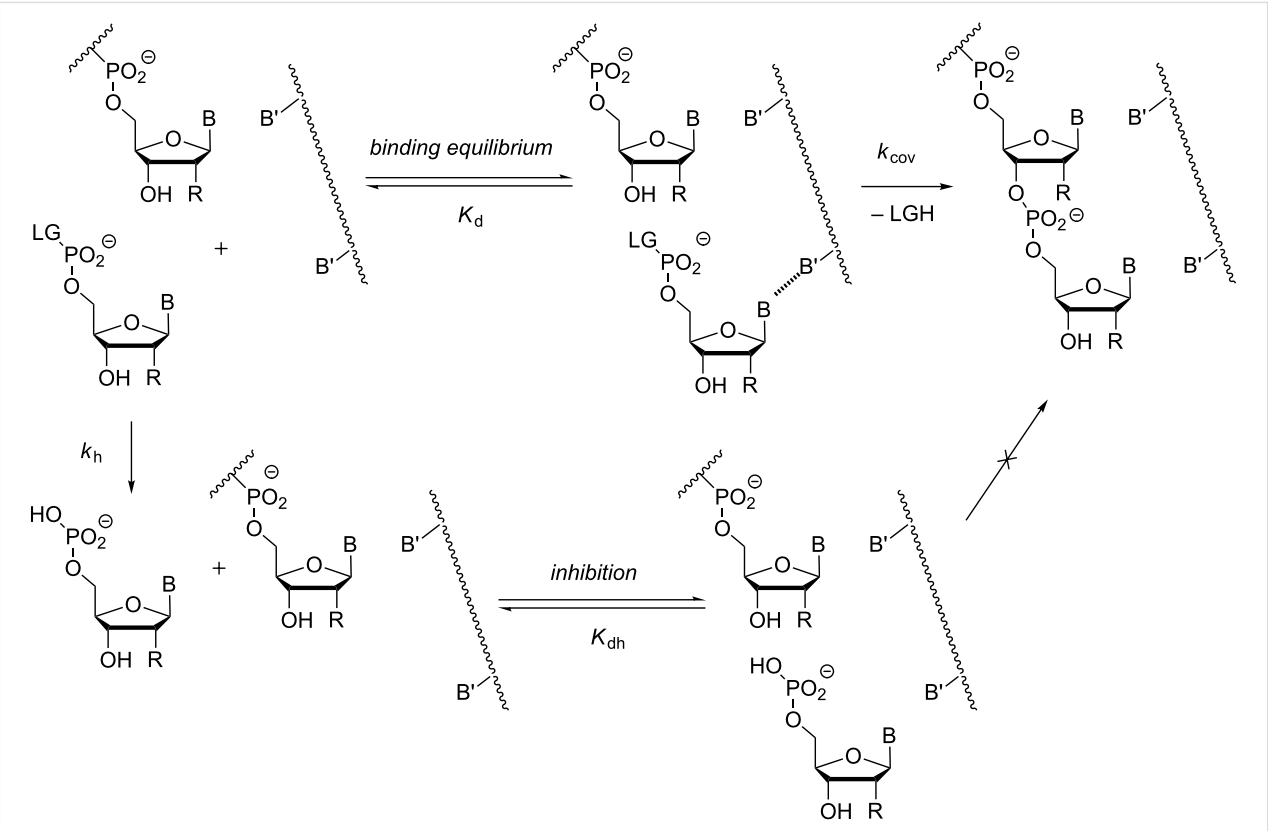


Figure 7: Steps and equilibria considered in our quantitative model of chemical primer extension [34]. The model considers the binding of the activated monomer with its leaving group (LG) to the primer–template complex in the form of the dissociation constant (K_d). It takes into account the rate of hydrolysis with the corresponding rate constant (k_h), the binding equilibrium for the hydrolyzed monomer that acts as inhibitor (K_{dh}), and it assumes a single rate-limiting chemical step (k_{cov}); B, B' = nucleobase = OH for RNA.

minal diol of an RNA primer, so hydrolysis was considered a very relevant parameter. We focused on the two classes of activated monomers mentioned above: 2-methylimidazolides and oxyazabenzotriazolides. Synthetic methods for producing such monomers from free nucleotides were briefly reviewed in our earlier account [22]. The methylimidazolides were chosen because a large body of literature exists on their reactions including studies by Orgel [35], Kanavarioti [31,36], Szostak [37,38], and Göbel [39]. The oxyazabenzotriazolides are our preferred monomers because they gave us the fastest primer extensions, both on RNA and on DNA templates [28,29,34,40].

Hydrolysis follows pseudo-first order kinetics and is readily measured by ^{31}P NMR spectroscopy [28,32-34]. Oxyazabenzotriazolides were indeed found to hydrolyze faster than methylimidazolides, and half-live times of hydrolysis at room temperature for the different nucleotides, in extension buffer containing 80 mM $MgCl_2$, were typically found to be in the range of 5–8 h at a pH of 8.9, both for ribonucleotides and for deoxynucleotides. Only OAt-dTMP was slower to hydrolyze, with a $t_{1/2}$ of 16 h [28]. Molecular modeling suggested that this may be due to the steric effect of the methyl group at the 5-position of the pyrimidine ring, shielded the leaving group-bearing phosphate from incoming water from some angles of attack. For OAt esters of ribonucleotides, we also measured the rates of hydrolysis at 0 °C and −20 °C, and the detailed data can be found in Supplementary Table S1 of reference [32]. At the lowest of the temperatures assayed, the half-live times increased to values between 51 h for OAt-UMP and 86 h for OAt-CMP.

Methylimidazolides were slower to hydrolyze. The half-lives of hydrolysis for deoxynucleotides were ranging between 19 h and 29 h whereas $t_{1/2}$ varied from 53 h to 63 h for ribonucleotides [34]. Our results were thus comparable to the ones obtained by Ruzicka and Frey who studied the hydrolysis of 5'-phosphorimidazolates of uridine at different pH values [41] and found a half-life toward hydrolysis of about 60 h in the absence of Mg^{2+} and at neutral pH, i.e., conditions favoring longevity for this type of activated monomer, which requires protonation of the imidazole ring to be turned it into a good leaving group.

Binding equilibria

As mentioned above, primer extension involves the binding of the incoming nucleotide to the primer–template complex, being directed by base pairing and stacking interactions. Therefore, it was important to determine the binding constants for activated and unactivated nucleotides experimentally. Theoretical predictions for triphosphates had suggested very tight binding [42], but the strong base dependence of the yield and selectivity of primer extension reactions suggested to us that not all nucleotides occupied the extension site to the same extent.

Initially, we wished to better understand how strong the inhibitory effect of spent monomers was, and we set up experiments to determine the K_{dh} value for complexes between free nucleotides and primer–template duplexes. This required methodology adjusted to measuring weak binding, i.e., much weaker than the strand-to-strand hybridization of oligonucleotides leading to duplexes, which is usually monitored by UV-melting analysis [43]. We chose NMR spectroscopy, partly because it is performed at much higher concentrations (millimolar, rather than micromolar analytes), and partly because it provides site-specific information without labeling. Labeling of an analyte as small as a mononucleotide with something other than isotopes was considered problematic, as it would strongly change the structural characteristics, and simple techniques, such as gel shift, do not work for complexes with a fast off-rate because the complex dissociates during the time it takes to perform the electrophoresis.

While NMR spectroscopy is sensitive when performed with a modern high field-spectrometer, it does have the disadvantage of producing complex spectra that require detailed analysis to assign at least the most critical resonances unambiguously. This is why we chose small hairpins with a non-nucleosidic hexaethyleneglycol loop [44] for our NMR-monitored titrations (Figure 8). The Szostak group later measured binding constants for complexes of three of the four unactivated ribonucleotides (A, C, and G) by NMR using longer constructs [45]. The hairpins are stable at room temperature and consist of only seven nucleotides, facilitating the interpretation of spectra. Resonances of the nucleobases at the terminus with the templating base were readily identified. Dissociation constants were determined by fitting the chemical shifts of terminal nucleotides in the ^{1}H NMR, measured at different nucleotide concentrations.

With the DNA hairpins and unactivated deoxynucleotides, depending on the sequence and experimental conditions, dissociation constants ranging from 10 mM (dGMP) and 280 mM (TMP) were measured [33,34]. In the RNA systems, the values measured were between 14 mM (GMP) and >500 mM (UMP). These results are similar to those obtained by Szostak and co-workers [45], who found that CMP binds most strongly, however, when studying a different sequence context.

We also measured K_d values for activated nucleotides, either with 2-methylimidazole or with oxyazabenzotriazole as leaving group using rapid NMR titrations to avoid hydrolysis. In the case of the DNA system, a largely unreactive natural deoxyri-

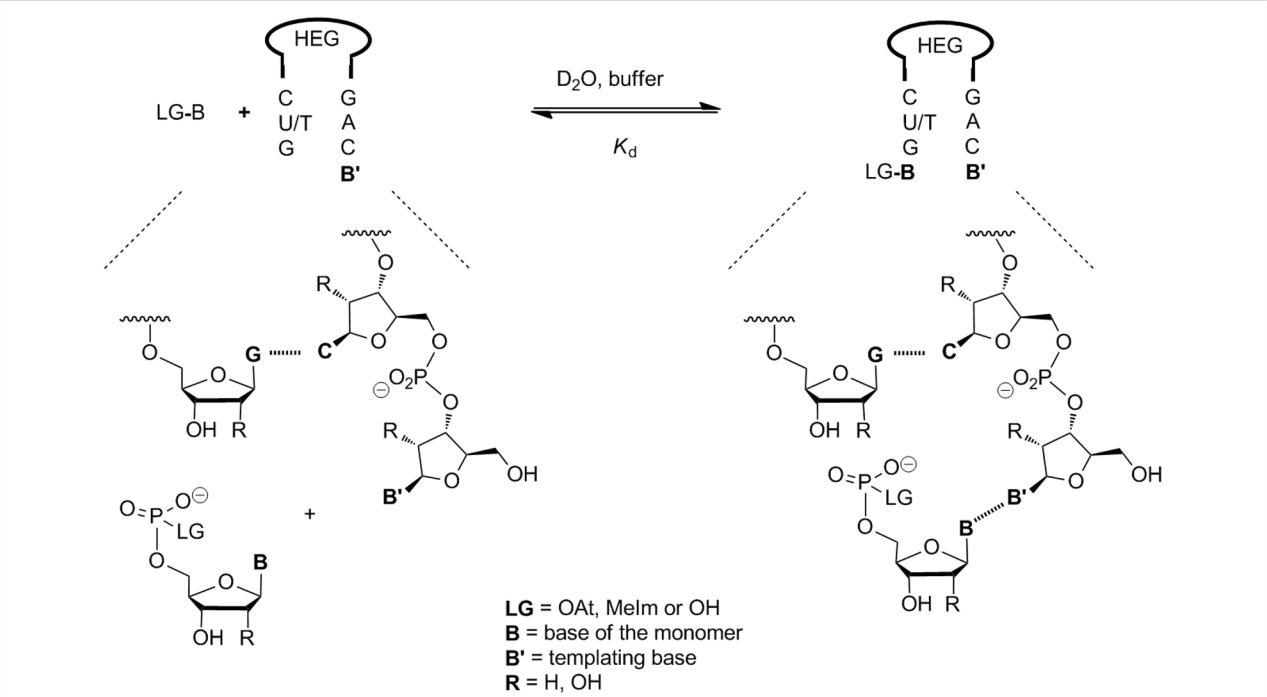


Figure 8: Binding equilibrium between mononucleotides and hairpins representing primer–template duplexes, as chosen for measuring dissociation constants by NMR titration.

bonucleotide at the 3'-terminus was used, not an aminodideoxynucleotide to prevent complications from reactions taking place during the titration. For RNA systems, there were also no significant signs of conversion on the time scale of our one-dimensional NMR experiments. The methylimidazole leaving group did not show a significant effect on the affinity of the nucleotides for the hairpins [34]. In contrast, the OAt group did lead to slightly stronger binding, and the effect was greatest for OAt-TMP and OAt-UMP, with a decrease in K_d by a factor of four to ten. Nevertheless, for strongly pairing bases, the difference between dissociation constants of free and activated monomers was minimal.

Besides NMR titrations with hairpins, we also used a different experimental approach to determine binding or dissociation constants. The complementary technique utilized the inhibitory effect of free nucleotides on primer extension. By adding increasing concentrations of the free nucleotide to the assays solutions and measuring the kinetics of extension, we were able to quantify binding independently, using global fits to the data sets. Here, longer templates were used, as well as downstream-binding oligonucleotides. Thus, 25 different dissociation constants were measured for different sequence contexts, ranging from 2 mM for dGMP and 200 mM for TMP [33]. This confirmed the positive effect of downstream-binding strands that we first reported in 2005 [28,29]. In the RNA case, we had shown that the presence of a 'helper strand' that is only three bases long can increase the yield of the extension by a factor of three at room temperature and by a factor of six in the cold. In their recent work, Szostak and co-workers measured dissociation constants for complexes of CMP via isothermal titration calorimetry [46]. When a downstream-binding strand was present, binding of the monomer was up to two orders of magnitude tighter than in its absence.

Simulating primer extension

With dissociation constants in hand, we were now in a position to determine rate constants for the covalent step of primer extension. For each case, the concentration of the kinetically relevant species (the monomer–primer–template complex), i.e., the occupancy of the extension site by the monomer, was now known, and measuring the initial rates led to the k_{cov} value via fitting. For OAt esters and an aminoterminal primer on a DNA template, values of 2–10 h^{-1} were found, whereas methylimidazolides gave rate constants between 0.3 and 1.4 h^{-1} [34]. For TMP, the reactivity with an OAt leaving group is four-fold higher than with a MeIm leaving group. The largest increase in reactivity was found for dCMP whose reaction with the amino-primer in the kinetically relevant complex is 25-fold faster as oxyazabenzotriazolide than as 2-methylimidazolide. For RNA primers on an RNA template, the values were between 0.01 h^{-1} (MeIm-AMP) and 0.1 h^{-1} (OAt-GMP). Overall, depending on the backbone, primer terminus, base, and leaving group, the rates of the chemical step vary by two orders of magnitude.

Using the set of four constants, one may then calculate the time-dependent yield of primer extension using the mathematical form of the model shown in Figure 7 [33]. The data predicted by our model agreed quite well with experimental data for either of the four nucleobases and the two different backbone chemistries. Figure 9a and 9b show representative plots of theoretical yields and data points from RNA-based assays at different monomer concentrations. It can be discerned that 7.2 mM monomer concentration does not suffice to induce more than approx. 30% conversion of the primer. The theoretical data on the left also shows the calculated time–yield curve for a hypothetical assay that does not suffer from inhibition by spent monomers. For such a scenario, full conversion is expected to occur. This is in agreement with the experimental observation that periodic removal of spent monomers prevents the stalling of primer extension that otherwise plagues these assays [32].

As explained in more detail in reference [33], there are three extreme cases. In the first case, both primer and tightly binding monomer are so reactive that full conversion is achieved before inhibition can become significant. This is the scenario found for OAt esters and the aminoterminal primer. The second scenario involves reactive monomer and primer, but the monomer is binding poorly (e.g., TMP), with just a few percent occupation of the extension site at the beginning of the assay. Here, hydrolysis does catch up with the desired reaction eventually, but it is inconsequently, because the low occupancy does not produce a significant level of competitive inhibition. In other words, if there is not much of a template effect to begin with, the spent monomers will not outcompete the monomer over time, and the reaction will largely proceed as expected for a second-order reaction with a competing reaction that just drains active monomer (hydrolysis). In the third case the primer is fairly unreactive, being equipped with just the terminal diol of natural RNA. Further, the monomer is a strongly binding one (OAt-GMP). In this case, inhibition becomes significant over time, and removal or re-activation of the monomer is required to prevent the extension from ceasing before near-quantitative conversion is achieved. This is what was done in the successful copying assays with immobilized primer–template duplex (Figure 10).

The insights gained from the quantitative analysis of primer extension leaves several options to push assays to completion. The first is to employ highly reactive and well binding mono-

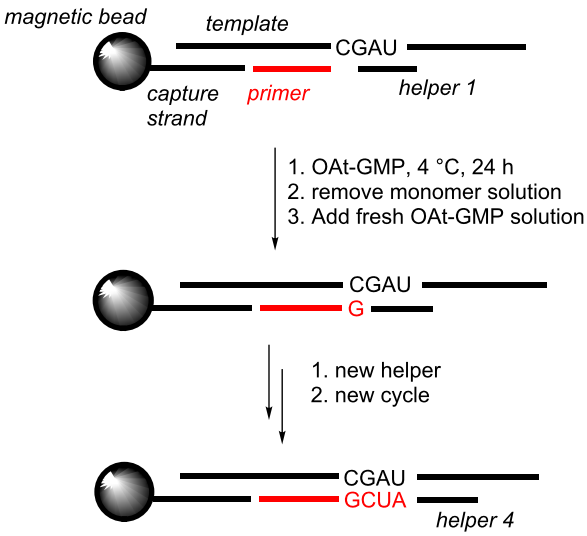


Figure 10: Copying of four nucleotides on an immobilized RNA duplex, as reported by Deck et al. [32].

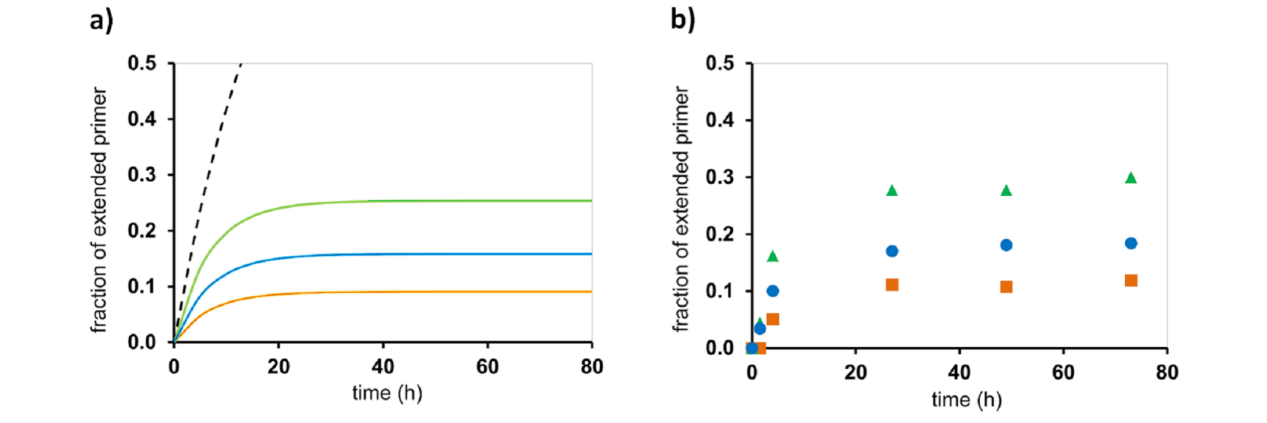


Figure 9: Template-directed primer extension on an RNA template performed with OAt-GMP at 1.8 mM (orange), 3.6 mM (blue), or 7.2 mM (green) initial concentration. a) Conversion over time, as simulated with our quantitative model, using the dissociation constants of both activated and free nucleotide, and rate constants for hydrolysis and chemical step. The broken black line is the hypothetical conversion of the primer without hydrolysis of monomer and the resulting inhibition; b) Corresponding experimental data, acquired in primer extension assays at 20 °C in buffer (200 mM HEPES, 400 mM NaCl, 80 mM $MgCl_2$, at pH 8.9) at 36 μM primer–template (5'-UAUGCUGG-3' – 3'-CACCCACCACAUACGACCCAAGCACAC-5'); see reference [34] for further details.

mers only. For RNA, this approach does not appear realistic, if one wants to work with any given sequence context and all four bases (A/C/G/U). The second option is to remove the spent monomers when their concentration reaches a critical threshold. This requires immobilization of the primer–template complex and washing [32,47] or removal of hydrolyzed monomers by dialysis [38]. The third option is finding conditions for in situ activation, so that spent monomers can be re-activated during the time course of the assay. The fourth option is searching for better leaving groups that give a more favorable ratio of rates for primer extension and hydrolysis (k_{cov}/k_{hydr}), when reacting with an RNA primer. It will not be trivial to find such a leaving group, as the nucleophilicity of alcohols is quite similar to that of water, so that it is difficult to utilize the chemoselectivity toward reaction partners with different softness, pK_a, or other structural features.

Copying on solid support

As mentioned above, one option to avoid stalling of primer extension reactions is to perform them on solid support. For RNA, the immobilization of the primer–template duplex was achieved by employing a biotinylated capture strand that was bound to streptavidin-coated magnetic beads (Figure 10) [32]. The assays allowed for near-quantitative incorporation of any of the four nucleobases opposite their complementary base in the template, but the reactions on the RNA-based system are quite slow.

For aminoterminal DNA, a methodology was developed by us that allows repeated incorporation of reactive 3'-amino-2',3'-dideoxynucleotide building blocks, activated as OAt esters [47]. This methodology can, in principle, be automated, and was established with a view towards sequencing, using fluorophore-labeled nucleotides [48,49]. To avoid cyclization or oligomerization of the monomer, the 3'-amine was protected with an azidomethyloxycarbonyl (Azoc) protecting group. This protecting group can be rapidly removed under non-denaturing conditions after incorporation by the complementary nucleotide using the Staudinger reaction with a water-soluble phosphine (Figure 11). This protocol, with what in the sequencing commu-

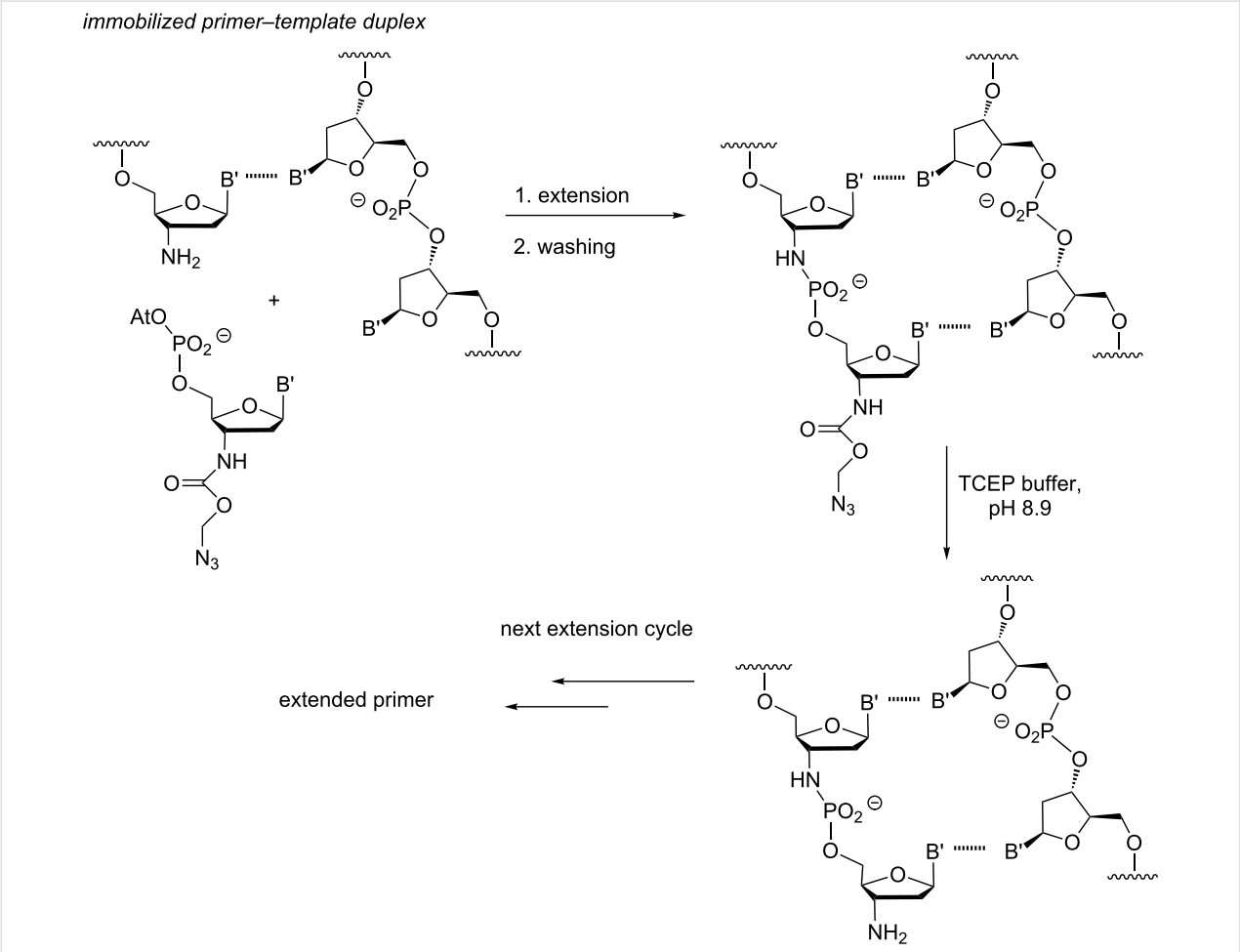


Figure 11: Extension cycle of aminoterminal primer with N-protected nucleotides on solid support, as described by Kaiser et al. [47].

nity is called "reversible termination", allowed efficient copying with any of the four nucleobases (A/C/G/T) in less than 12 h for each incorporation at room temperature. It was also used to demonstrate enzyme-free, template-directed primer extension in the non-natural direction (P3'→N5'), using 3'-phosphates of 3'-amino-2',3'-dideoxynucleosides [47].

Activation chemistry

Imidazolium bisphosphates

During our work on the effect of leaving groups on the yield of primer extension reactions, we noticed a burst phase in the kinetics of methylimidazolides that was only observed with monomers that were not carefully purified. The high reactivity was traced to a species with a chemical shift of −10.8 ppm in the ^{31}P NMR spectrum that was identified as the imidazolium bisphosphate (Figure 12 and Figure 13) [34]. We calculated a second order rate constant for the reaction of the imidazolium bisphosphate with the primer of 2.9×10^4 M^{-1} h^{-1}, which is approx. 600-fold larger than that of the pure methylimidazolide. The kinetics and analytical data were presented in the Supporting Information of ref. [34]. Shortly afterwards, Szostak and co-workers published a series of papers on the role of imidazolium bisphosphates in primer extension [50-53], including NMR data for ^{13}C-labeled 2-methylimidazolides that showed bonding to two phosphates. We did not pursue the imidazolium bisphosphate further because we did not observe full conversion of the primer at reasonable concentrations of this labile species. Other imidazolium phosphates, such as those formed upon in situ activation appeared more promising (vide infra).

The extensive work on imidazolium bisphosphates by the Szostak group was prompted by an observation made during assays with a trimer downstream of the primer extension site, pre-activated as methylimidazolide. The presence of the leaving

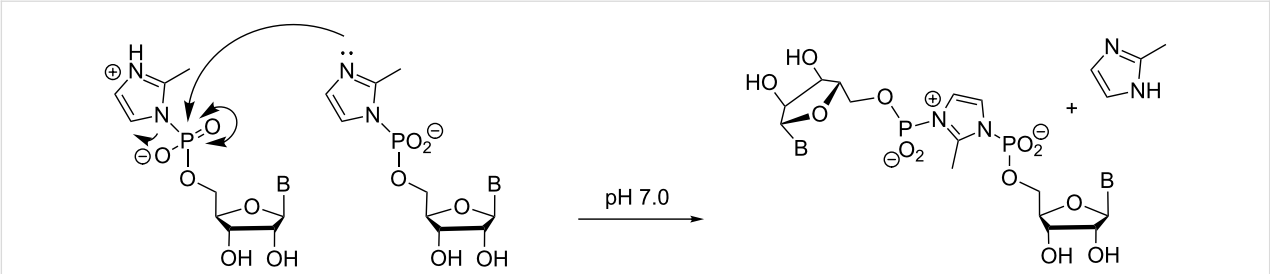


Figure 12: Formation of a highly reactive methylimidazolium bisphosphate from methylimidazolides of nucleotides.

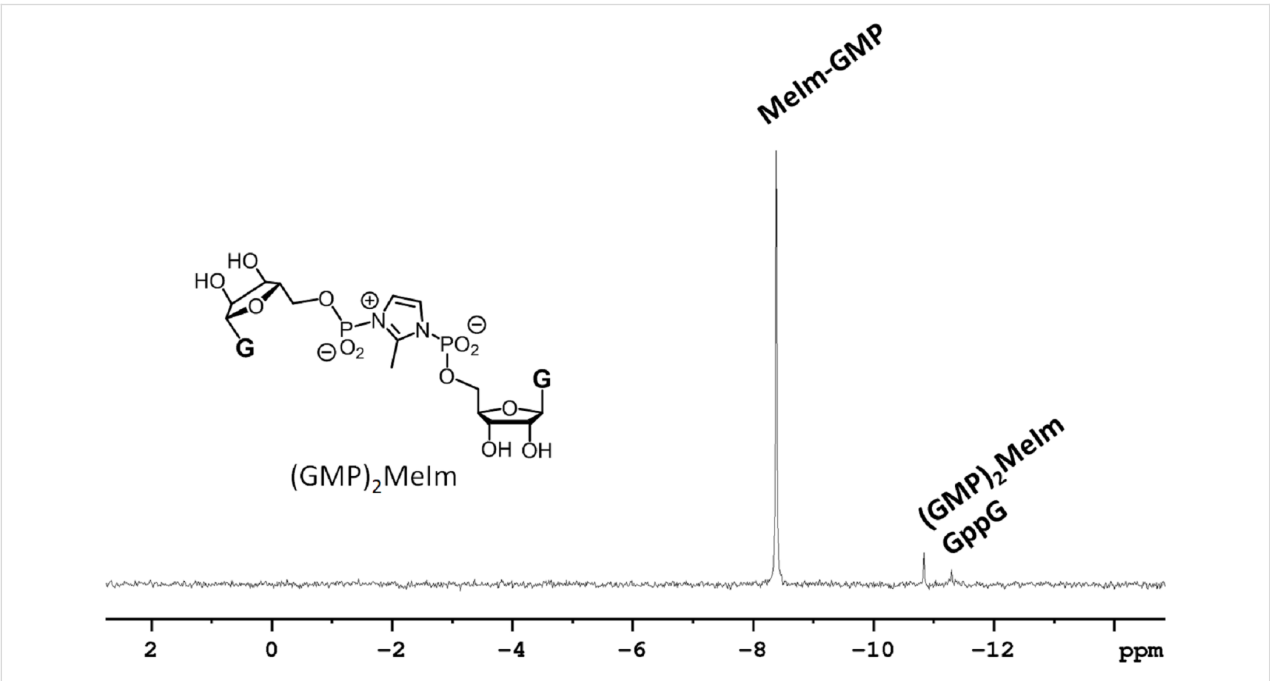


Figure 13: ^{31}P NMR spectrum (161.9 MHz) of crude MeIm-GMP in D_2O. The resonance of the imidazolium bisphosphate appears at −10.8 ppm, and that of the pyrophosphate GppG at −11.3 ppm.

group was found to accelerate the incorporation reaction [51]. A subsequent optimization identified 2-aminoimidazolides as monomers with superior properties [50]. Figure 14a shows the proposed intermediate forming when two neighboring monomers have reacted, and Figure 14b shows the binding mode of a GpppG dimer that was found to bind in a fashion described as structurally similar to the proposed intermediate shown on the left-hand side [53]. In the latter case, LNA residues were used in the template strand to facilitate crystallization.

We note that the neighboring group participation proposed should be limited to leaving groups with a second nucleophilic site at the appropriate position. Other leaving groups than methylimidazole should not be able to react via the same dominant reaction pathway. In our hands, compounds with a different structure such as OAt esters give rapid and high-yielding reactions. This is also true for the intermediates of extension with in situ activation, which lack the second nucleophilic group entirely (vide infra). Further, we have consistently found that helper oligonucleotides without a phosphate group at the 5'-terminus accelerate primer extension reactions, both for aminoterminal primers [28] and for RNA-based systems [29]. If formation of an imidazolium bisphosphate was the dominant reaction pathway, this should not be the case. Without a phosphate group, the bisphosphate cannot be formed, and the helper should block the reaction pathway that requires this species bound to the template. Full conversion was found with OAt esters in the presence of an unphosphorylated helper, even for UMP [40]. The aminoimidazolium phosphates are interesting and well-binding species. Time will tell whether they provide the most favorable pathway for primer extension. Perhaps, the successful copying of long stretches of RNA templates will be the ultimate test for their ability to support enzyme-free copying.

In situ activation

Re-activation of hydrolyzed monomers during the course of the extension assay is another approach to avoid stalling due to inhibition. As mentioned in the Introduction, ligation reactions had been achieved by Naylor and Gilham in aqueous media in presence of a water-soluble carbodiimide as condensing reagent [7]. Likewise, Sulston et al. had used EDC to oligomerize AMP in the presence of poly(U) as template [12]. Polymerization of nucleotides with in situ activation had also been attempted with the aid of montmorillonite, a clay mineral, but had led mostly to dimers and pyrophosphate [54]. For DNA, ligations starting from unactivated starting materials were known [10,55,56], but not always high-yielding, unless an aminoterminal strand was reacted with the phosphate-terminated counterpart [57-59], to form a phosphoramidate-linked product. Efficient versions of extension of an RNA primer with in situ activation were not known to us.

One difficulty in inducing the extension of RNA primers with ribonucleotides without preactivation lies in the different pH optima of the two reactions. The activation, which now has to occur in the same solution as the extension, is most easily performed under slightly acidic conditions, whereas the extension reaction is favored under basic conditions, particularly when good leaving groups, such as oxyazabenzotriazolides are involved. Further, the activating agent (condensing agent) is an electrophile, and there is significant potential for side reactions

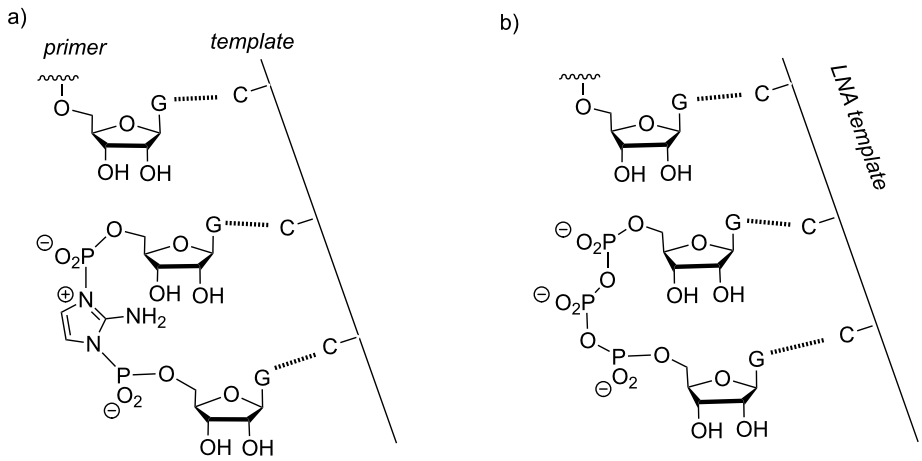


Figure 14: Imidazolium bisphosphate as intermediate in the primer extension reaction, as described by Szostak and colleagues. a) Intermediate of an extension with aminoimidazolides as monomers [52]; b) one of the structural arrangements found in a recent crystallography study that used oligophosphates as model compounds [53].

of the reagent with other nucleophilic groups than the 5'-phosphate of the desired nucleotide. As a consequence, assays involving in situ activation were slow and low yielding.

We assumed that the inefficient primer extension with in situ activation could be improved via organocatalysis. We had previously found, when studying extension of aminoterminal primers, that small heterocycles, such as pyridine, can increase the rate of the reaction [60]. Most probably, this effect was organocatalytic in nature, being caused by a pyridinium species that forms in the reaction medium, an effect known from the acceleration of DCC-induced acylation reactions with dimethylaminopyridine [61]. With aminoterminal primers, in situ activation and organocatalysis with 1-methylimidazole in a magnesium-free buffer had led to encouraging results, even at submillimolar nucleotide concentration [62]. So, starting from a primer extension reaction with an RNA-system that gave less than 1% conversion after 24 h, a number of heterocycles were screened. The best results were obtained for 1-methyladenine and 1-ethylimidazole (1-EtIm) [63]. Optimization of the reaction conditions then led to a method that gave 90% conversion in 48 h, successful incorporation of more than one nucleotide in a row, and high yielding extension even with poorly binding UMP, all under the same conditions. We dubbed these conditions "general condensation buffer". The optimized buffer contains 500 mM HEPES, 800 mM EDC, 80 mM $MgCl_2$, and 150 mM 1-EtIm. Assays are usually performed at 0 °C to shift the binding equilibrium for the incoming nucleotide to the bound side, and thus strengthen the template effect.

The proposed mechanism for the reaction is shown in Figure 15. In order to start the activation, the carbodiimide has to react with the phosphate group, leading to what is sometimes called a "covalent adduct". This first step may either occur in solution, or while the nucleotide is already bound to the primer–template duplex. We have not been able to observe a signal for the "EDC adduct" in the NMR spectra, and the binding equilibrium establishes itself quickly. We assume that the on- and off-rate are much faster than the NMR time scale. The EDC adduct is then expected to react with the organocatalyst, yielding the alkylimidazolium nucleotide that acts as the kinetically most relevant monomer in the extension reaction. The ethylimidazolium species can be observed as a small peak in ^{31}P NMR spectra. The extension occurs as expected, most

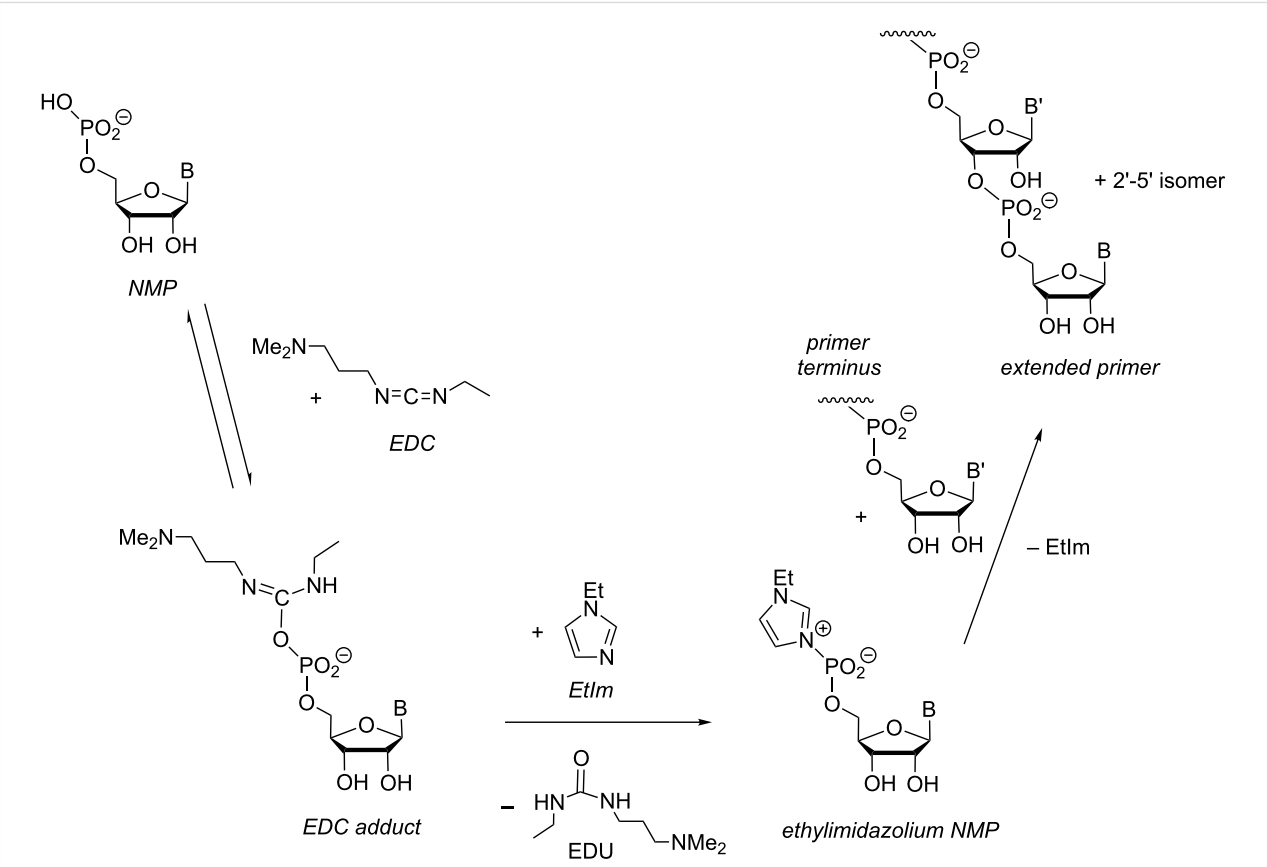


Figure 15: Proposed steps of enzyme-free primer extension with in situ activation, using the "general condensation buffer" containing EDC and 1-ethylimidazole as organocatalyst.

probably via addition/elimination, including a pentavalent intermediate and possibly by a pseudorotation to place the ethylimidazole leaving group in an apical position. Since either of the two alcohols of the terminal diol of the primer can attack, a mixture of 3',5'- and 2',5'-isomers is expected for this reaction. Work on quantifying the ratio of the diastereomers formed is under way in our laboratories.

Unlike assays performed with pre-activated monomers, primer extension with in situ activation of ribonucleotides led to significant levels of oligomers that form via untemplated polymerization of the nucleotides [63]. This was a welcome side reaction, as it helps to explain how RNA may have been formed and copied under prebiotic conditions. In fact, when oligomerization assays were performed with any of the four natural ribonucleotides, oligomers of mixed sequence long enough to hybridize stably to complementary strands were formed. Such strands may then be the templates or primers required to start enzyme-free copying.

Further, the general condensation buffer noted above gave rise to the spontaneous formation of ribonucleotide- or RNA-linked peptides [64]. These peptides are linked via their N-terminus to the ribonucleotide portion as phosphoramidates, which is why we refer to them as "peptido RNAs". Peptide chain growth on the 5'-phosphate is much faster than the background reaction [65], and will thus predominate over background oligomerization of amino acids alone. Further, the rate of formation of peptido RNA depends on the structure of the amino acid, and, to a lesser degree, on that of the ribonucleotide [66], so that a very primitive, not yet encoded form of RNA-induced peptide synthesis can occur under conditions that support the formation and copying of genetic information. We felt that this was significant for theories on the emergence of life from inanimate materials, even more so as the same reaction conditions also support the formation of pivotal cofactors of primary metabolism from nucleotide precursors [64]. The reactions mentioned above occur spontaneously in cold aqueous solution, without the need for mineral surfaces or enzymes.

Condensation producing peptido nucleotides also occurs with other activating agents, such as cyanamide or carbonyl diimidazole (CDI). Cyanamide, a tautomer of unsubstituted carbodiimide, has long been considered a prebiotically relevant activating agent [67], and it has previously been used in experiments aimed at generating peptides in the absence of enzymes or a ribosomal machinery [68]. Reactions with cyanamide are much less efficient than with EDC, so that successful primer extension has not yet been observed in our assays. But, while reactions that take weeks or months are not problematic for prebiotic evolution, which probably occurred over many millions of years, they are difficult to study in detail in an academic setting that requires results on the time scale of Ph.D. theses.

Conclusion

Enzyme-free primer extension is a fascinating reaction that has been linked to the origin of the first self-replicating systems. The reaction does produce extended primers with nucleotides complementary to the template sequence being appended at the 3'-terminus, but it is slow and low-yielding, particularly when performed with natural RNA/ribonucleotides. Because it relies on weak Watson–Crick base pairing between a single nucleotide and a templating base, the reaction cannot be driven to completion by heating or harsh conditions. Instead, a subtle interplay of binding equilibria and chemical steps either leads to successful incorporation of the nucleotide monomer or to the more likely path of hydrolysis, which in turn can prevent further extension via competitive inhibition [69]. Detailed quantitative work has led to a better understanding of the processes underlying incomplete conversion and thus to approaches that reduce inhibition or slow conversion. Among them is the removal of hydrolyzed monomer, improved activation chemistries, or in situ (re)activation with the support of an organocatalyst. Despite progress in the field, the ultimate goal of demonstrating enzyme-free replication of RNA strands long enough to code for an oligo- or polypeptide is not yet in sight. This is particularly true, if one considers that the issue of low sequence fidelity was not even discussed in this short account. Much remains to be done for chemists and biochemists alike.

Acknowledgements

We would like to thank H.-P. Mattelaer and E. Kervio for critical comments on the manuscript. The work of the authors on enzyme-free copying is supported by DFG grant No. RI 1063/16-1 to C.R.

References

1. Kornberg, A.; Baker, T. A. *DNA Replication,* 2nd ed.; University Science Books: Mill Valley, 2005.
2. Crick, F. *Nature* **1970,** *227,* 561–563. doi:10.1038/227561a0
3. Cech, T. R. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** *83,* 4360–4363. doi:10.1073/pnas.83.12.4360
4. Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982,** *31,* 147–157. doi:10.1016/0092-8674(82)90414-7
5. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. *Science* **2011,** *332,* 209–212. doi:10.1126/science.1200752
6. Attwater, J.; Wochner, A.; Holliger, P. *Nat. Chem.* **2013,** *5,* 1011–1018. doi:10.1038/nchem.1781
7. Naylor, R.; Gilham, P. T. *Biochemistry* **1966,** *5,* 2722–2728. doi:10.1021/bi00872a032
8. von Kiedrowski, G. *Angew. Chem., Int. Ed. Engl.* **1986,** *25,* 932–935. doi:10.1002/anie.198609322

9. Zielinski, W. S.; Orgel, L. E. *Nature* **1987,** *327,* 346–347. doi:10.1038/327346a0
10. Li, T.; Nicolaou, K. C. *Nature* **1994,** *369,* 218–221. doi:10.1038/369218a0
11. Patzke, V.; von Kiedrowski,, G. *ARKIVOC* **2007,** *(v),* 293–310. doi:10.3998/ark.5550190.0008.522
12. Sulston, J.; Lohrmann, R.; Orgel, L. E.; Miles, H. T. *Proc. Natl. Acad. Sci. U. S. A.* **1968,** *59,* 726–733. doi:10.1073/pnas.59.3.726
13. Orgel, L. E.; Lohrmann, R. *Acc. Chem. Res.* **1974,** *7,* 368–377. doi:10.1021/ar50083a002
14. Joyce, G. F. *Cold Spring Harbor Symp. Quant. Biol.* **1987,** *52,* 41–51. doi:10.1101/SQB.1987.052.01.008
15. Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1982,** *162,* 201–217. doi:10.1016/0022-2836(82)90169-3
16. Matteucci, M. D.; Caruthers, M. H. *J. Am. Chem. Soc.* **1981,** *103,* 3185–3191. doi:10.1021/ja00401a041
17. Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992,** *114,* 317–322. doi:10.1021/ja00027a040
18. Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992,** *114,* 5496–5501. doi:10.1021/ja00040a002
19. Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992,** *114,* 7963–7969. doi:10.1021/ja00047a001
20. Kurz, M.; Göbel, K.; Hartel, C.; Göbel, M. W. *Angew. Chem.* **1997,** *109,* 873–876. doi:10.1002/ange.19971090812
21. Zielinsky, W. S.; Orgel, L. E. *Nucleic Acids Res.* **1985,** *13,* 2469–2484. doi:10.1093/nar/13.7.2469
22. Rojas Stütz, J. A.; Kervio, E.; Deck, C.; Richert, C. *Chem. Biodiversity* **2007,** *4,* 784–802. doi:10.1002/cbdv.200790064
23. Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.* **2004,** *39,* 99–123. doi:10.1080/10409230490460765
24. Dörr, M.; Löffler, P. M. G.; Monnard, P.-A. *Curr. Org. Synth.* **2012,** *9,* 735–763. doi:10.2174/157017912803901691
25. Szostak, J. W. *J. Syst. Chem.* **2012,** *3,* No. 2. doi:10.1186/1759-2208-3-2
26. Joyce, G. F.; Orgel, L. E. *J. Mol. Biol.* **1986,** *188,* 433–441. doi:10.1016/0022-2836(86)90166-X
27. Kervio, E.; Hochgesand, A.; Steiner, U. E.; Richert, C. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107,* 12074–12079. doi:10.1073/pnas.0914872107
28. Hagenbuch, P.; Kervio, E.; Hochgesand, A.; Plutowski, U.; Richert, C. *Angew. Chem., Int. Ed.* **2005,** *44,* 6588–6592. doi:10.1002/anie.200501794
29. Vogel, S. R.; Deck, C.; Richert, C. *Chem. Commun.* **2005,** 4922–4924. doi:10.1039/b510775j
30. Lohrmann, R.; Orgel, L. E. *Tetrahedron* **1978,** *34,* 853–855. doi:10.1016/0040-4020(78)88129-0
31. Kanavarioti, A.; Bernasconi, C. F.; Doodokyan, D. L.; Alberas, D. J. *J. Am. Chem. Soc.* **1989,** *111,* 7247–7257. doi:10.1021/ja00200a053
32. Deck, C.; Jauker, M.; Richert, C. *Nat. Chem.* **2011,** *3,* 603–608. doi:10.1038/nchem.1086
33. Kervio, E.; Claasen, B.; Steiner, U. E.; Richert, C. *Nucleic Acids Res.* **2014,** *42,* 7409–7420. doi:10.1093/nar/gku314
34. Kervio, E.; Sosson, M.; Richert, C. *Nucleic Acids Res.* **2016,** *44,* 5504–5514. doi:10.1093/nar/gkw476
35. Kozlov, I. A.; Orgel, L. E. *Mol. Biol.* **2000,** *34,* 781–789. doi:10.1023/A:1026663422976
36. Kanavarioti, A.; Bernasconi, C. F.; Alberas, D. J.; Baird, E. E. *J. Am. Chem. Soc.* **1993,** *115,* 8537–8546. doi:10.1021/ja00072a003
37. Zhang, S.; Zhang, N.; Craig Blain, J.; Szostak, J. W. *J. Am. Chem. Soc.* **2013,** *135,* 924–932. doi:10.1021/ja311164j
38. Adamala, K.; Szostak, J. W. *Science* **2013,** *342,* 1098–1100. doi:10.1126/science.1241888
39. Hey, M.; Hartel, C.; Göbel, M. W. *Helv. Chim. Acta* **2003,** *86,* 844–854. doi:10.1002/hlca.200390084
40. Vogel, S. R.; Richert, C. *Chem. Commun.* **2007,** 1896–1898. doi:10.1039/b702768k
41. Ruzicka, F. J.; Frey, P. A. *Bioorg. Chem.* **1993,** *21,* 238–248. doi:10.1006/bioo.1993.1020
42. Poater, J.; Sart, M.; Fonseca Guerra, C.; Bickelhaupt, F. M. *Chem. Commun.* **2011,** *47,* 7326–7328. doi:10.1039/c0cc04707d
43. Breslauer, K. J. *Methods Enzymol.* **1995,** *259,* 221–242. doi:10.1016/0076-6879(95)59046-3
44. Altmann, S.; Labhardt, A. M.; Bur, D.; Lehmann, C.; Bannwarth, W.; Billeter, M.; Wüthrich, K.; Leupin, W. *Nucleic Acids Res.* **1995,** *23,* 4827–4835. doi:10.1093/nar/23.23.4827
45. Izgu, E. C.; Fahrenbach, A. C.; Zhang, N.; Li, L.; Zhang, W.; Larsen, A. T.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2015,** *137,* 6373–6382. doi:10.1021/jacs.5b02707
46. Tam, C. P.; Fahrenbach, A. C.; Björkbom, A.; Prywes, N.; Cagri Izgu, E.; Szostak, J. W. *J. Am. Chem. Soc.* **2017,** *139,* 571–574. doi:10.1021/jacs.6b09760
47. Kaiser, A.; Spies, S.; Lommel, T.; Richert, C. *Angew. Chem., Int. Ed.* **2012,** *51,* 8299–8303. doi:10.1002/anie.201203859
48. Griesang, N.; Gießler, K.; Lommel, T.; Richert, C. *Angew. Chem., Int. Ed.* **2006,** *45,* 6144–6148. doi:10.1002/anie.200600804
49. Gießler, K.; Griesser, H.; Göhringer, D.; Sabirov, T.; Richert, C. *Eur. J. Org. Chem.* **2010,** 3611–3620. doi:10.1002/ejoc.201000210
50. Walton, T.; Szostak, J. W. *J. Am. Chem. Soc.* **2016,** *138,* 11996–12002. doi:10.1021/jacs.6b07977
51. Prywes, N.; Blain, J. C.; Del Frate, F.; Szostak, J. W. *eLife* **2016,** *5,* e17756. doi:10.7554/eLife.17756
52. Walton, T.; Szostak, J. W. *Biochemistry* **2017,** *56,* 5739–5747. doi:10.1021/acs.biochem.7b00792
53. Zhang, W.; Tam, C. P.; Walton, T.; Fahrenbach, A. C.; Birrane, G.; Szostak, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114,* 7659–7664. doi:10.1073/pnas.1704006114
54. Ferris, J. P.; Ertem, G.; Agarval, V. *Origins Life Evol. Biospheres* **1989,** *19,* 165–178. doi:10.1007/BF01808150
55. Shabarova, Z. A.; Merenkova, I. N.; Oretskaya, T. S.; Sokolova, N. I.; Skripkin, E. A.; Alexeyeva, E. V.; Balakin, A. G.; Bogdanov, A. A. *Nucleic Acids Res.* **1991,** *19,* 4247–4251. doi:10.1093/nar/19.15.4247
56. Narayanan, S.; Dalpke, A. H.; Siegmund, K.; Heeg, K.; Richert, C. *J. Med. Chem.* **2003,** *46,* 5031–5044. doi:10.1021/jm030902l
57. Achilles, T.; von Kiedrowski, G. *Angew. Chem., Int. Ed. Engl.* **1993,** *32,* 1198–1201. doi:10.1002/anie.199311981
58. Vogel, H.; Richert, C. *ChemBioChem* **2012,** *13,* 1474–1482. doi:10.1002/cbic.201200214
59. Kalinowski, M.; Haug, R.; Said, H.; Piasecka, S.; Kramer, M.; Richert, C. *ChemBioChem* **2016,** *17,* 1150–1155. doi:10.1002/cbic.201600061
60. Röthlingshofer, M.; Kervio, E.; Lommel, T.; Plutowski, U.; Hochgesand, A.; Richert, C. *Angew. Chem., Int. Ed.* **2008,** *47,* 6065–6068. doi:10.1002/anie.200801260
61. Neises, B.; Steglich, W. *Angew. Chem., Int. Ed. Engl.* **1978,** *17,* 522–524. doi:10.1002/anie.197805221
62. Röthlingshöfer, M.; Richert, C. *J. Org. Chem.* **2010,** *75,* 3945–3952. doi:10.1021/jo1002467

63. Jauker, M.; Griesser, H.; Richert, C. *Angew. Chem., Int. Ed.* **2015,** *54,* 14559–14563. doi:10.1002/anie.201506592
64. Jauker, M.; Griesser, H.; Richert, C. *Angew. Chem., Int. Ed.* **2015,** *54,* 14564–14569. doi:10.1002/anie.201506593
65. Griesser, H.; Tremmel, P.; Kervio, E.; Pfeffer, C.; Steiner, U. E.; Richert, C. *Angew. Chem., Int. Ed.* **2017,** *56,* 1219–1223. doi:10.1002/anie.201610650
66. Griesser, H.; Bechthold, M.; Tremmel, P.; Kervio, E.; Richert, C. *Angew. Chem., Int. Ed.* **2017,** *56,* 1224–1228. doi:10.1002/anie.201610651
67. Schimpl, A.; Lemmon, R. M.; Calvin, M. *Science* **1965,** *147,* 149–150. doi:10.1126/science.147.3654.149
68. Hawker, J. R., Jr.; Oró, J. *J. Mol. Evol.* **1981,** *17,* 285–294. doi:10.1007/BF01795750
69. Kaiser, A.; Richert, C. *J. Org. Chem.* **2013,** *78,* 793–799. doi:10.1021/jo3025779

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Recent advances in synthetic approaches for medicinal chemistry of C-nucleosides

Kartik Temburnikar[1] and Katherine L. Seley-Radtke[*2]

Review

Open Access

Address:
[1]Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, 725 N. Wolfe St. Baltimore, MD 21205, United States and [2]Department of Chemistry and Biochemistry, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, United States

Email:
Katherine L. Seley-Radtke[*] - kseley@umbc.edu

* Corresponding author

Keywords:
C-nucleosides; convergent synthesis; modular synthesis

Beilstein J. Org. Chem. **2018,** *14,* 772–785.
doi:10.3762/bjoc.14.65

Received: 17 December 2017
Accepted: 06 March 2018
Published: 05 April 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

C-nucleosides have intrigued biologists and medicinal chemists since their discovery in 1950's. In that regard, C-nucleosides and their synthetic analogues have resulted in promising leads in drug design. Concurrently, advances in chemical syntheses have contributed to structural diversity and drug discovery efforts. Convergent and modular approaches to synthesis have garnered much attention in this regard. Among them nucleophilic substitution at C1' has seen wide applications providing flexibility in synthesis, good yields, the ability to maneuver stereochemistry as well as to incorporate structural modifications. In this review, we describe recent reports on the modular synthesis of C-nucleosides with a focus on D-ribonolactone and sugar modifications that have resulted in potent lead molecules.

Introduction

Nucleic acids form the genetic blueprint for all living organisms and are involved with a wide range of cellular functions [1-9]. Modifications to their chemical structure can have profound effects on structure and function of enzymes, cells and supramolecular complexes [10-22]. Nucleic acids are composed of a monomeric nucleoside unit that features an aromatic nitrogenous moiety (a nucleobase) connected to a pentose sugar, which in turn is attached to a phosphate group (Figure 1) [7]. The pentose sugar and the nucleobase are connected by a carbon–nitrogen bond that is adjacent to the sugar oxygen resulting in an hemiaminal ether bond, also known as the glycosidic bond.

Because of their key role in many biological processes, modifications to the nucleoside structure have been widely employed in the design of drugs, most notably in the fields of virology and cancer research [13-15]. Variations in the nucleoside scaffold are typically accomplished by the insertion, deletion or transposition of functional groups or atoms [23-29]. The varied properties of such modified nucleosides arise from changes in hydro-

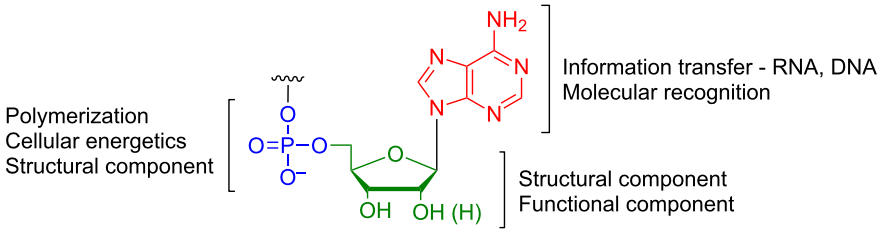


Figure 1: Structural components of nucleic acids. Shown is the monomeric building block of nucleic acids. Changes to the nucleotide structure can affect molecular recognition, as well as structure and function.

gen bonding motifs, electronic effects, hydrophobic interactions, acid-base properties and chemical reactivity [25-37]. One such modification is the change in the nature of the glycosidic bond [29,37].

Although the glycosidic bond is stable under physiological conditions, cleavage of the bond is common and is highly dependent on the nature of the nucleobase and local pH. In addition, the rate of glycosidic bond cleavage is higher for purines than pyrimdines [38-44]. Moreover, the glycosidic bond in 2'-deoxy ribonucleosides has a higher susceptibility to cleavage than in the corresponding ribonucleosides [38-41,43]. The rate of glycosidic (C–N) bond cleavage is enhanced by decreasing pH and enzymes, which modify the localized acid–base environment [31,35,36]. The C–N bond cleavage proceeds either by activation of a nucleophile that attacks C1' or by stabilization of the leaving group, which could either be the nucleobase or an oxocarbenium ion [31,36]. As such, the oxocarbenium ion is a species formed during the glycosidic bond cleavage, which may be present as an intermediate or a transition state depending upon the accumulation of the positive charge on the sugar ring (Figure 2). As a result, any change in the nucleobase–sugar connectivity (C–N) affects the formation of the oxocarbenium ion and thus influences the stability (or instability) of the nucleoside analogues.

Replacing the hemiaminal (O–C–N) connectivity of the canonical nucleosides with an O–C–C bond (Figure 3) results in a class of compounds called “C-nucleosides” [45-51]. Further modification to a C–C–C connectivity results in “carbocyclic C-nucleosides” (Figure 3) [52,53]. C-nucleosides feature (hetero)aryl aromatic groups such as 9-deazapurines, pyrimidines, pyridines and phenyl groups connected by a C–C bond to a sugar (or sugar mimic) as shown in Figure 4 [30,45-

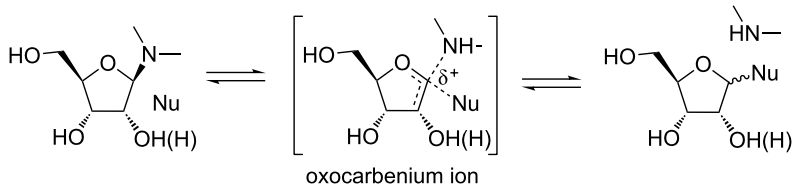


Figure 2: Formation of oxocarbenium ion during glycosidic bond cleavage in nucleosides [31]. The extent of leaving group stabilization and approach of the nucleophile determine charge accumulation on the sugar ring. A concerted process leads to a transition state-like species shown in the figure, while a greater accumulation of positive charge leads to an oxocarbenium ion intermediate.

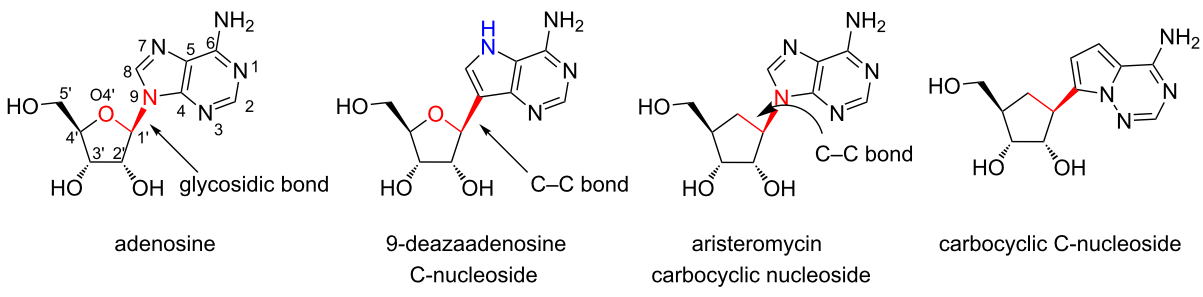


Figure 3: Structural modifications to nucleobase-sugar connectivity. The O–C–N bond between nucleobase and sugar defines the glycosidic bond. Replacement of nucleobase nitrogen by carbon results in C-nucleosides. A carbon (CH_2) replacing the sugar oxygen results in carbocyclic nucleosides. When both the heteroatoms in the glycosidic bond are replaced by carbon, the resultant compounds are called carbocyclic C-nucleosides.

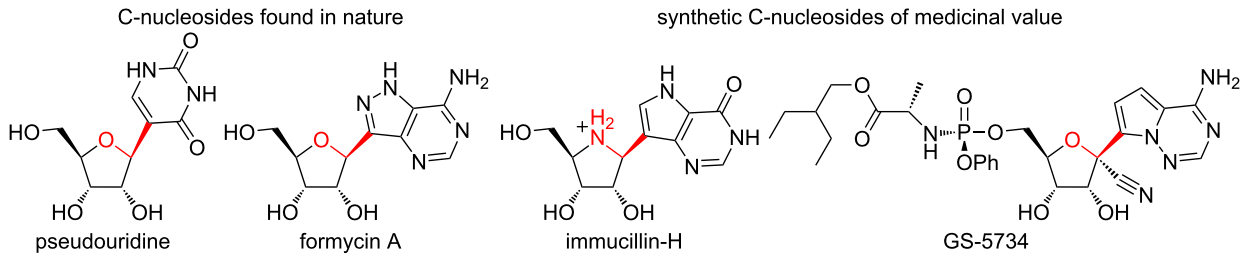


Figure 4: Examples of natural and synthetic C-nucleosides. Pseudouridine and formcycin are among several naturally occurring C-nucleosides that are being studied for their role in RNA biology and antibiotic properties respectively. In the recent past, synthetic C-nucleosides, such as immucillin-H and GS-5734, have shown potent activity against purine nucleoside phosphorylases (PNP) and broad spectrum antiviral activities.

47,50,54-57]. The change in the nature of the glycosidic bond is accompanied by i) increased hydrolytic stability, ii) altered hydrogen bonding motifs, and iii) altered molecular recognition properties [25,29,37,58]. Because of these changes, C-nucleosides have been useful in the study of RNA and DNA processing enzymes, as well as drug design efforts and novel supramolecular structures [12,29,59].

Pseudouridine is a naturally occurring C-nucleoside that was first discovered in the 1950s [45-47,50]. Subsequently, many more C-nucleosides were discovered and their medicinal properties evaluated (Figure 4) [18,29,37,45-50,60-62]. Due to advances in synthetic methodologies over the years, the repertoire of C-nucleosides has since expanded and has enabled the discovery of clinically useful molecules. Some of the more prominent biologically active analogues that have advanced to clinical evaluations include the immucillins developed by Schramm et al, and Gilead's antiviral pyrrolo[2,1-*f*]triazine C-nucleosides (GS-5734 and GS-6620) [32,63-65]. Thus, this review attempts to capture the progress in the synthesis efforts and subsequent drug discovery of the C-nucleosides over the past few years. In the first section, the structural and stereochemical underpinnings of nucleophilic substitutions to D-ribonolactone are discussed, a method that has seen wide applications. Next, we describe reports of different applications and structural variants that have expanded the diversity of the C-nucleosides. Finally, we discuss a modular synthetic approach to carbocyclic C-nucleosides that is also based on the nucleophilic substitution of ribonolactone.

Review

Nucleophilic addition to D-ribonolactone and its stereochemistry

Two prominent methods of C-nucleoside syntheses involve either i) the linear construction of a (hetero)aryl moiety on a C1'-functionalized ribose or ii) coupling of a pre-synthesized (hetero)aryl with a ribosyl moiety (Figure 5A) [48,49,62]. The C–C bond formation usually involves a functional group at C1' of the ribosyl moiety that is amenable to additional functionalization (Figure 5B). Like other nucleoside coupling approaches (other than the well-known Vorbrüggen coupling reaction [66], the synthesis of C-nucleosides typically gives a mixture of stereoisomers (α and β) at the anomeric carbon [48,49,54,62,67,68]. Since the naturally occurring nucleosides (and most biologically active nucleosides) are β-anomers, achieving 100% stereospecificity in C–C bond formation is an important goal, but often difficult to attain [62].

Among the aforementioned approaches for C-nucleoside syntheses, the coupling of (hetero)aryls to the ribosyl moiety has seen the widest application [52,53,58,62-65,69-77]. This can be ascribed to the modular nature of syntheses that allows for simultaneous alterations in the sugar and the nucleobase to generate diverse analogues in a facile manner. Within this approach, nucleophilic substitution of ribonolactone (Figure 6A) has garnered much attention [61-63,69-75,78]. Ribonolactone typically with its hydroxy groups protected, is amenable to nucleophilic substitutions [78,79]. Use of C-nucleophiles such as lithiated (hetero)aryls leads to C-C bond via a lactol intermediate (Figure 6A). Subsequent deoxygenation of the C1'–OH by Lewis acids (e.g., $BF_3{\cdot}OEt_2$) results in an oxocarbenium ion [62,70,80-83]. Reduction of this intermediate by various silanes gives C-nucleosides resembling the canonical nucleosides [82,83]. The stereochemical fate of oxocarbenium ion reduction is dictated by the conformation and stability of the oxocarbenium ion, which in turn, is affected by the nature of the C2', C3' and C5' substituents [80,81].

Codée and coworkers elaborated on the mechanism and stereochemistry of this reaction by calculating the energies of different oxocarbenium conformers using a free energy surface (FES) mapping method [80,81]. These studies were based on the Woerpel's model comprising of two stable conformers, namely ^3E and E_3, in equilibrium (Figure 6B) [84,85]. The nucleophile approaches from the side presenting the least number of eclipsing interactions with the C2' substituent (Figure 6B) [80]. Examining the energies of the various conformers of the permethylated furanosyl oxocarbenium intermediate revealed

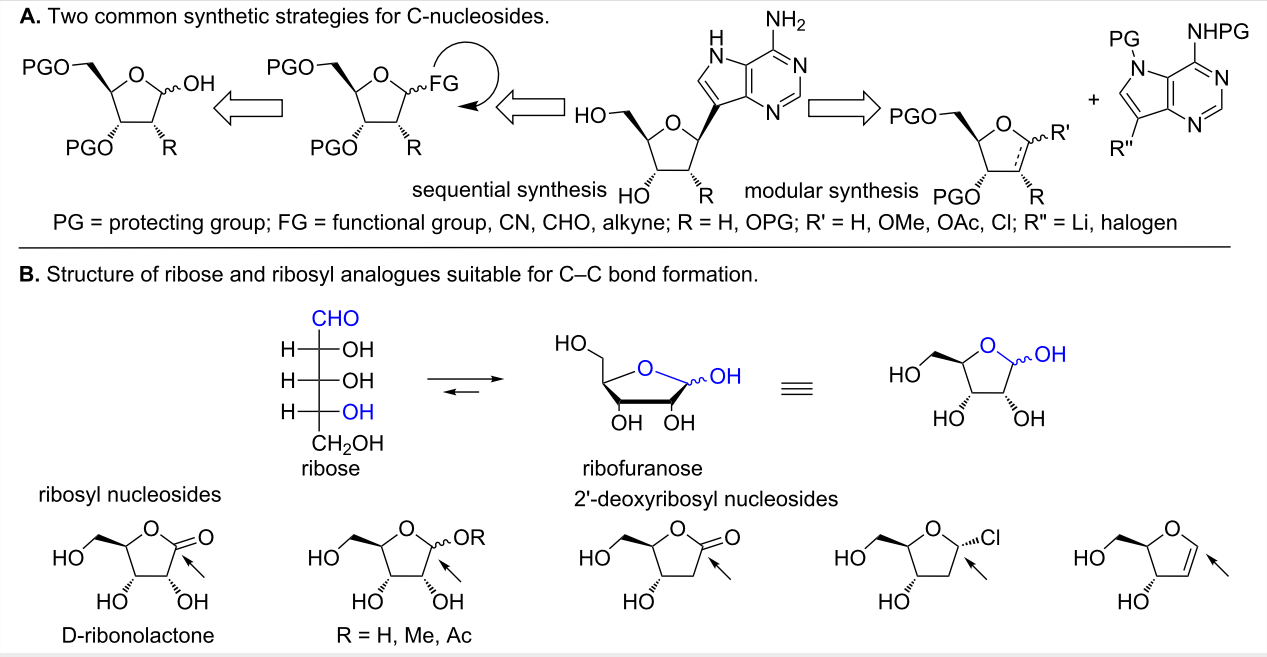


Figure 5: Synthetic approaches to C-nucleosides. A. Two common strategies for C-nucleoside synthesis involve functionalization at C1' and coupling of preformed sugar and heterocylic compounds. B. Structure of ribose and C1' functional groups that enable coupling reactions and synthesis of C-nucleosides.

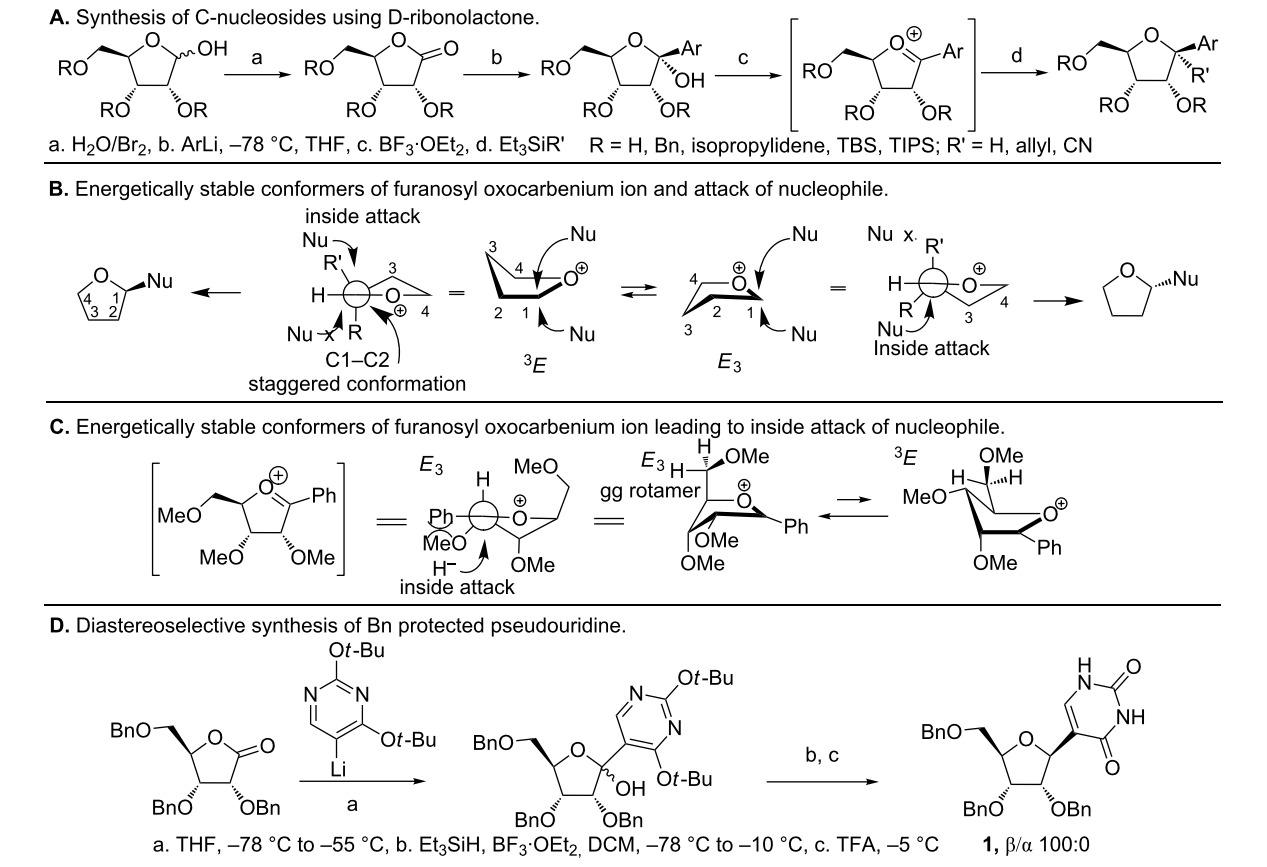


Figure 6: Steroselective C-nucleoside synthesis using D-ribonolactone. A. Nucleophilic substitution of D-ribonolactone results in an oxocarbenium ion intermediate. B and C. Functional groups at C2', C3' and C5' stabilize the charged sugar ring and direct the approach of nucleophile to affect the stereochemical outcome of oxocarbenium ion reduction. D. Stereoselective synthesis of protected pseudouridine [80,81].

that the E_3 conformer with the C5'–OMe oriented over the positively charged furanosyl ring (Figure 6C) has a large stabilizing effect due to C5'–O5 dipole interactions. In addition, the C2' pseudoequatorial methoxy and C3' pseudoaxial methoxy groups further stabilize the intermediate in E_3 conformer, thereby favoring the E_3 confomer over the ^3E. In the case of an anomeric phenyl group (Ph, Figure 6C), stabilization of the positive charge (C=O$^+$) through conjugation, via parallel alignment, helps to overcome the unfavorable steric interactions between the C2'–OMe and the Ph group [81]. Because E_3 is the favored conformer, an inside attack of the nucleophile (H$^-$) results in an α orientation in the final product, which is evident from the the synthesis of **1** (OBn-substituted Pseudouridine, Figure 6D). Despite the greater stability of the E_3 conformer, it is the faster reacting conformer (E_3 or ^3E) that ultimately affects the ratio of diastereomers in the final product [80]. This difference in reactivity results in the differences in various α/β mixtures obtained during the synthesis of C-nucleosides using the D-ribonolactone approach.

Antiviral C-nucleosides

The formation of the lactol and oxocarbenium ion illustrated in Figure 6 also presents the possibility of C1' di-substitution, which was exploited by researchers at Gilead in the discovery of the potent antiviral 4-aza-7,9-dideazaadenine (pyrrolo[2,1-*f*][1,2,4]triazine) C-nucleosides (Figure 7) [63-65,69,70,77].

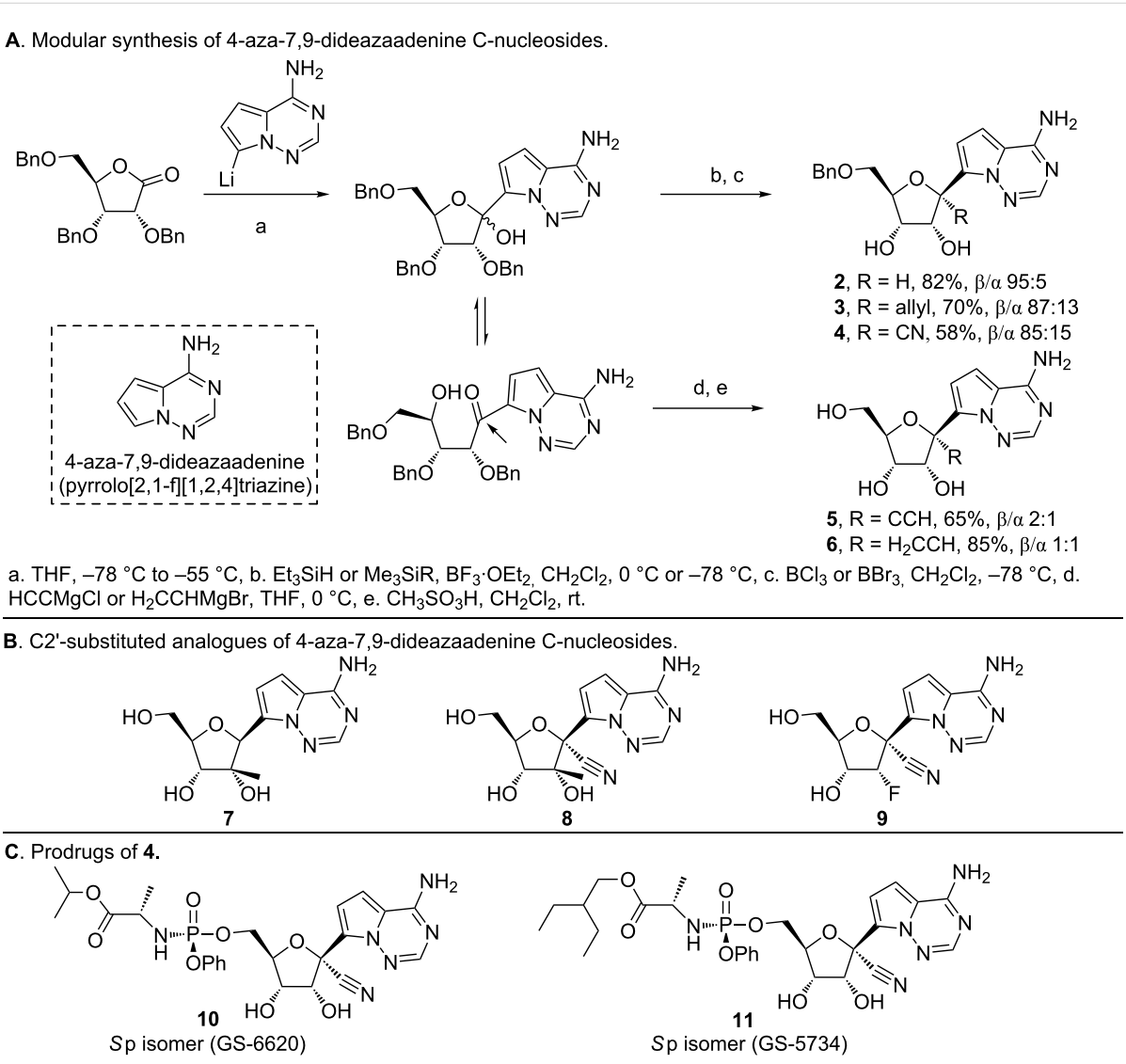


Figure 7: Synthesis of C1'-substituted 4-aza-7,9-dideazaadenine C-nucleosides [63-65,69,70]. A. Reaction of D-ribonolactone and lithiated heterocycle leading to C1'-substituted C-nucleosides. Steroechemical scrambling is observed when ring opened form of ribose is employed during nucleophilic substitution. B. Synthesis of C2' C-nucleosides using analogous D-ribonolactones. C. Masked C-nucleoside monophosphates exhibiting potent antiviral activity and clinical utility.

The synthesis of 4-aza-7,9-dideazaadenine C-nucleoside **2** (Figure 7A) was first reported by Patil et al. [86] using a sequential approach. In contrast, scientists at Gilead treated perbenzylated ribonolactone with lithiated 4-aza-7,9-dideazaadenine to obtain the lactol intermediate (Figure 7A) [63,65,69,70]. Deoxygenation of the lactol intermediate by $BF_3{\cdot}OEt_2$ resulted in the oxocarbenium ion, which was then reduced using triethylsilane to obtain **2**. Replacing triethylsilane with allyl trimethylsilane and trimethylsilyl cyanide gave C1'-allyl (**3**) and C1'-cyano (**4**) substitutions respectively. A β/α ratio of 95:5, 87:13 and 89:11 was observed for **2, 3** and **4**, respectively, which was sensitive to the reaction temperature and the reagents used [69,70]. A marked difference in diastereomeric purity was observed when the open form of the ribofuranose ring (which exists in equilibrium with the ring closed form), was exploited for C1' substitution using Grignard reagents [69]. Acid-catalyzed dehydration resulted in a diastereomeric mixture of C1'-disubstituted products **5** and **6** with an observed β/α ratio of 2:1 and 1:1, respectively. Similarily, C2'-substituted ribonolactones were employed in the synthesis of 2'-β-Me analogues **7** and **8** and the 2'-deoxy-2'-fluoro **9** (Figure 7B) [65].

The 1'-α-H analogue **2** was reported to exhibit inhibitory activity against neoplastic cell lines [86]. This scaffold was later elaborated by Gilead to discover broad spectrum activity of the related C-nucleosides (**3**–**11**) against viruses from the *Flaviviridae*, *Orthomyxoviridae*, *Paramyxoviridae* and *Coronaviridae* families [63-65,69,70]. Cell-based assays revealed potent activity for compounds **2**–**9** against various viruses including Ebola (EBOV, *Filoviridae*), respiratory syncytial virus (RSV, *Pneumoviridae*) and the hepatitis-C virus (HCV, *Flaviviridae*) family [63,65,69]. Through structure activity relationship studies, the 1'-CN compound **4** emerged as a compound with activity against EBOV, HCV and RSV [65,69]. It is active against EBOV in human microvascular endothelial cells and RSV with low cytotoxicity towards Huh-7, HEp-2 and MT4 cells. Moreover, the triphosphate of **4** selectively inhibits, HCV RdRp and RSV RdRp over human RNA Pol II and DNA polymerases (α, β, γ) [65]. The 2'-Me compound **7** as its triphosphate (TP) shows anti-HCV activity in replicon assays [77]. However, **7**-TP serves as a substrate for mitochondrial RNA polymerase, thereby causing toxicity in rats [63]. The 2'-F and 2'-β-Me compounds **8** and **9** are active against the HCV, but lack activity against EBOV and RSV in cell-based assays [65]. The pharmacokinetic properties of **4** were improved by converting it to the masked monophosphates (**10** and **11**, Figure 7C), which serves to facilitate transport into the infected cells, and conversion to the active triphosphate form, thereby leading to high and persistent levels [63-65]. The 2-ethylbutyl L-alanine phosphoramidate prodrug (Sp isomer, GS-5734, **11**) increases the loading of macrophages derived from human monocytes over its unmasked analogue [64]. It was also observed that intravenous administration of the prodrug leads to increased liver loading (as the triphosphate) in hamsters compared to oral dosing [63].

Draffan et al. synthesized a series of 2'-β-Me analogues of pyrrolo- and imidazo[2,1-*f*][1,2,4]triazine C-nucleosides using a 2'-β-Me lactone that mimic adenosine and guanosine (**12**–**19**, Figure 8) [71,72]. The adenine analogues of pyrrolo- and

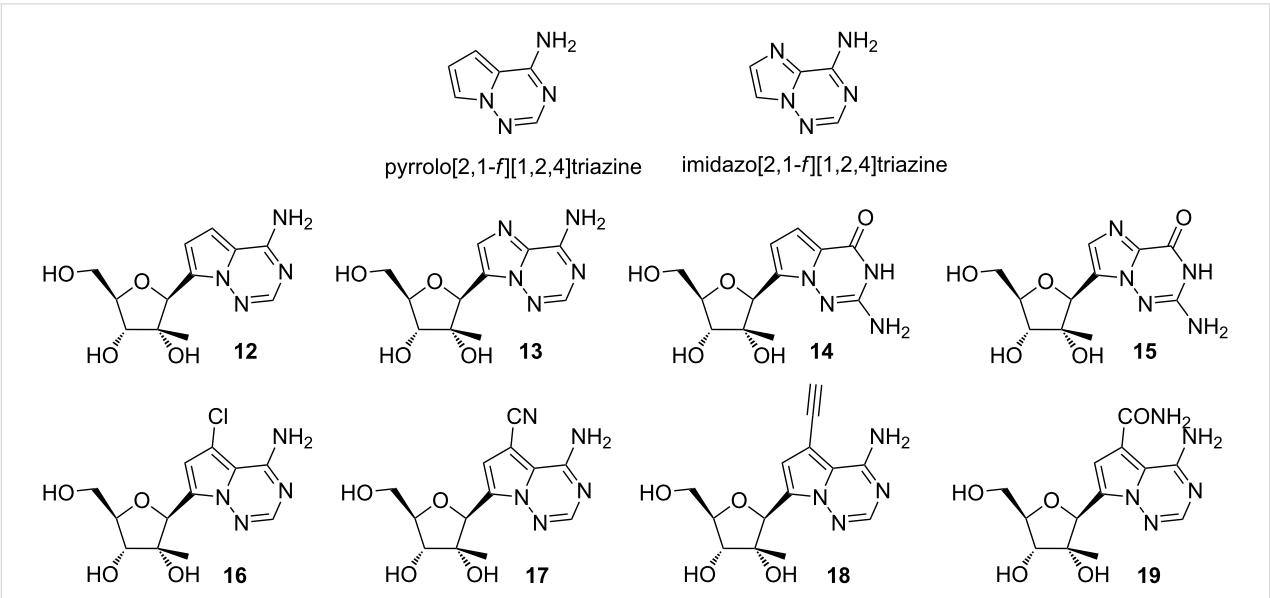


Figure 8: Pyrrolo- and imidazo[2,1-*f*][1,2,4]triazine C-nucleosides. A series of sugar- and nucleobase-substituted C-nucleosides were synthesized via nucleophilic substitution of D-ribonolactone for structure–activity relationship [71,72].

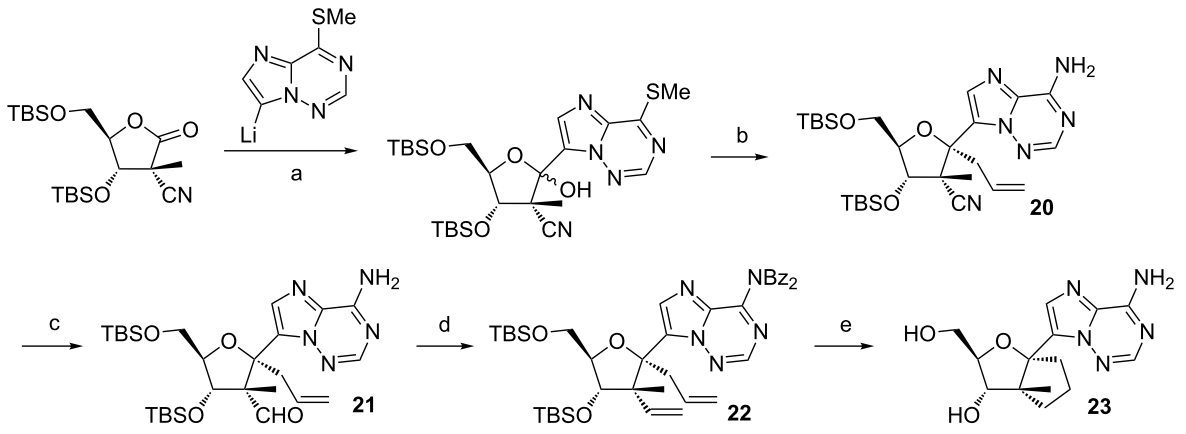
imidazo[2,1-*f*][1,2,4]triazine were active as nucleosides in HCV1b RNA replication assays, and as triphosphates they inhibit the NS5B polymerase as did the triphosphates of the guanosine analogues [71]. The library of adenosine analogues was further expanded by introducing functional groups at C7 (**16**–**19**), which exhibit potent activity in RNA replication assays, with the carboxamide group in particular imparting high potency but also high cytotoxicity [72].

A further modification to the imidazo[2,1-*f*][1,2,4]triazine C-nucleoside scaffold was reported by Dang et al., wherein they synthesized a series of 2'-β-Me analogues possessing a 1',2' cyclopentyl ring (Figure 9) [73]. A representative synthesis (compound **23**) is shown in Figure 9, which involves installing an allyl group at C1' (**20**) and converting the C2'–CN to an aldehyde (**21**) followed by a Wittig reaction to install a second allyl group at C2' (**22**). Second generation Grubb's catalyst was used for the ring formation, followed by hydrogenation to give the desired cyclopentane ring (**23**) [73]. The biological data of these compounds has yet to be reported.

Wang et al. synthesized a series of pyridine and pyrimidine C-nucleosides (**24**–**26)** that mimic the riboside of favipiravir in their effort to develop novel anti-influenza compounds (Figure 10A) [74]. Protected D-ribonolactone **27** was treated with lithiated pyridine to obtain lactol **28** (Figure 10B). Deoxygenation and reduction gave **29**, wherein the isopropylidene group was also removed. Conversion of the cyano to an amide group, followed by removal of the silyl protecting group gave **24**, which proved to be the most promising compound. The fluorine on **29** was replaced with a methoxy group after re-installing the isopropylidene protecting group. The cyano group was then converted to an amide and the methoxy converted to a hydroxy group. Removal of the protecting groups on the sugar gave **25**, which exhibited potent activity against the H1N1 influenza strain (A/WSN/33) in cell based assays [74]. The pyrimidine compound **26** was synthesized using an identical approach and is not shown here. The activity of **24** and **25** as nucleosides was comparable to favipiravir and its riboside. Furthermore, they found that the triphosphate of **24** (**24**-TP) was incorporated opposite U and C of an RNA template by the influenza polymerase [74]. These experiments indicate that the H-bonding motifs of **24** allow it to mimic both A and G (Figure 10A) [74]. Despite the mis-incorporation, an unmodified sugar moiety may not result in obligate chain termination. While **24**-TP is incorporated opposite U and forms more of the full length product than terminated product, its incorporation opposite C results in greater truncated product. Thus, the putative mechanism of action of **24** is through mutagenesis of viral genomic RNA and inhibition of viral polymerase [74].

Synthesis of C2'-substituted furanolactone

In view of sugar scaffolds possessing C2' substitutions and their value to drug design, a report by Peifer et al. on the synthesis of C2'-substituted ribonolactones is notable (Figure 11A) [75]. Their finding appends known methods of C2' substitution that involve conversion of the C2'-OH to a ketone followed by Me or F substitution [87-94]. Using the Mukaiyama aldol reaction, Peifer obtained a C2'-substituted ribonolactone, which can then be employed in C-nucleoside synthesis [75]. This involves condensation of alkyl-substituted silyl ketene acetals (**32**) with enantioenriched α-2,2,6,6-tetramethylpiperidinyl-β-benzy-



a. THF, –78 C to –20 °C, b. i) Ac_2O, DMAP, pyr, ii) Ally-TMS, $BF_3{\cdot}OEt_2$, DCM, –78 °C, iii) NH_3, iPrOH, c. DIBAL-H, THF, 25 °C, d. Ph_3PMeBr, LiHMDS, 25 °C, e. i) BzCl, DMAP, pyr, 50 °C, ii) 2nd gen. Grubbs catalyst, DCE, 40 °C iii) TBAF, THF, 25 °C, iv) MeONa, MeOH, 25 °C, v) Pd/C, MeOH, 25 °C.

Figure 9: Synthesis of 1',2'-cyclopentyl C-nucleoside [73]. Functional groups at C1' and C2' were installed and employed for ring cyclization.

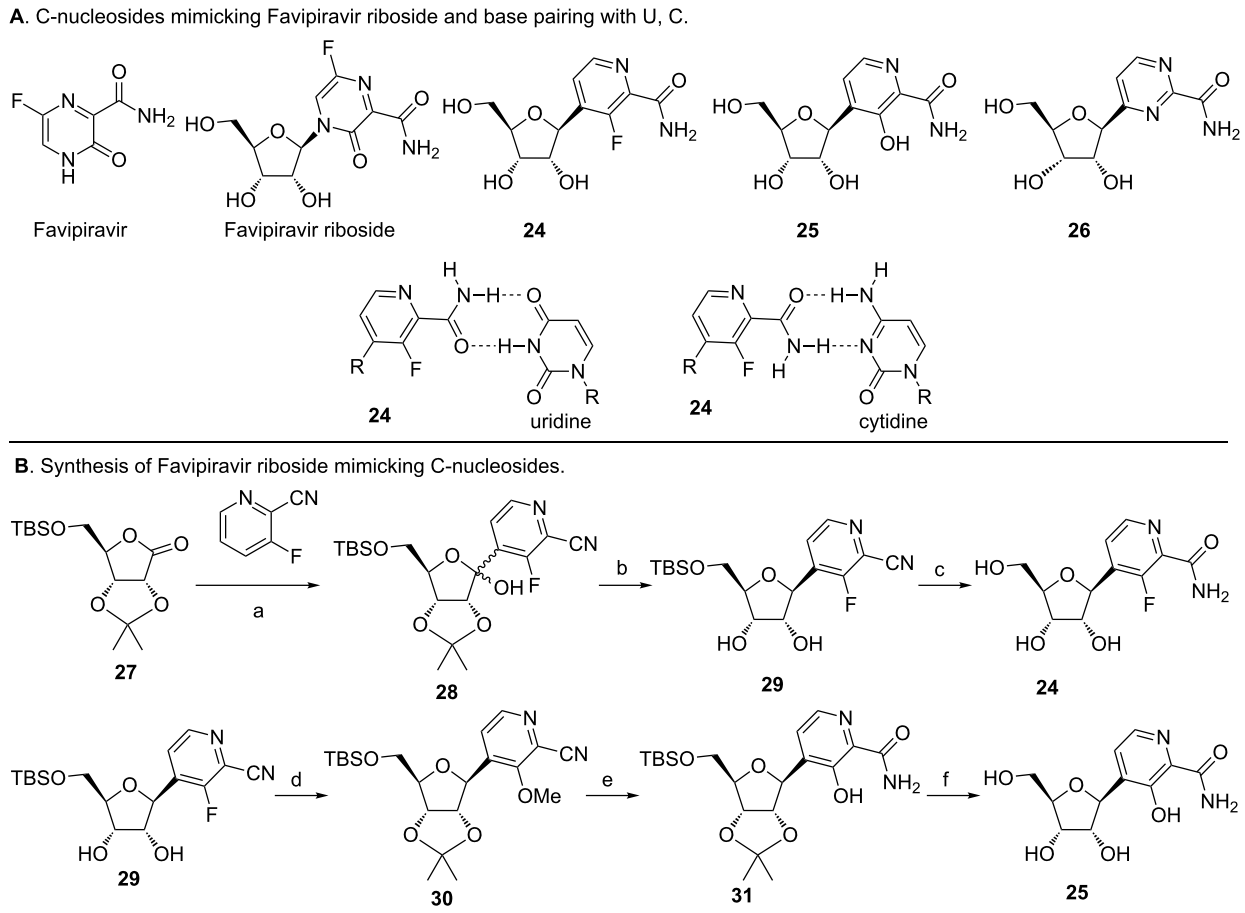


Figure 10: Anti-influenza C-nucleosides mimicking favipiravir riboside [74]. A. Structure of favipiravir and its riboside, which exhibits anti-influenza activity. C-nucleoside variants of favipiravir riboside and their base pairing with uridine and cytosine. B. Synthesis of C-nucleoside variants of favirpiravir starting from D-ribonolactone.

loxypropionaldehyde (**33**) in presence of $TiCl_2(OiPr)_2$ to give the β-hydroxyester **34** that is diastereomerically enriched [75,95]. Reductive cleavage of the 2,2,6,6-tetramethylpiperidinyl (TMP) group by Zn and trifloroacetic acid results in cyclization and formation of the C2'- substituted ribonolactone (**35**). $TiCl_2(OiPr)_2$ has been identified as the optimal Lewis acid for the synthesis of most ribonolactones with the exception of unsubstituted silyl ketene acetals (R = R' = H) that leads to stereochemical inversion at C3' [75]. The desired stereoselectivity for 2'-deoxy analogues was obtained when $BF_3{\cdot}OEt_2$ was used. Furthermore, another route to the synthesis of C-nucleosides was demonstrated by direct addition of aryl lithium reagents to the 2'-OMe ribonolactone (Figure 11B). While the expected lactol was formed, deoxygenation by $BF_3{\cdot}OEt_2$ and reduction in presence of the Hantzsch ester afforded the desired β-anomer, while the use of Et_3SiH gave the α-anomer [75].

Carbocyclic C-nucleosides

In an attempt to synthesize carbocyclic C-nucleosides, Maier et al. found that reaction of aryl lithiums with pentanone **37** results in carbocyclic C-nucleosides with a C1'-hydroxy group (**38** and **39**, respectively, Figure 12A) [52,53]. They synthesized cyclopentanone **37** in 7 steps starting from norbornadiene (**40**, Figure 12B). Furthermore, silyl protection (TIPS) of the C2' and C3' was observed to be critical for the stability of **37** and to facilitate functional group interconversions as shown in Figure 12A [53].

In order to obtain carbocyclic C-nucleosides that resemble canonical nucleosides, Maier and coworkers synthesized a stable enol triflate (**46,** Figure 13A), which then enables Suzuki coupling and a modular synthesis of carbocyclic C-nucleosides [53]. The boronic acids/boronates (inset, Figure 13) of several (hetero)aryls were conducive to Suzuki coupling with the best

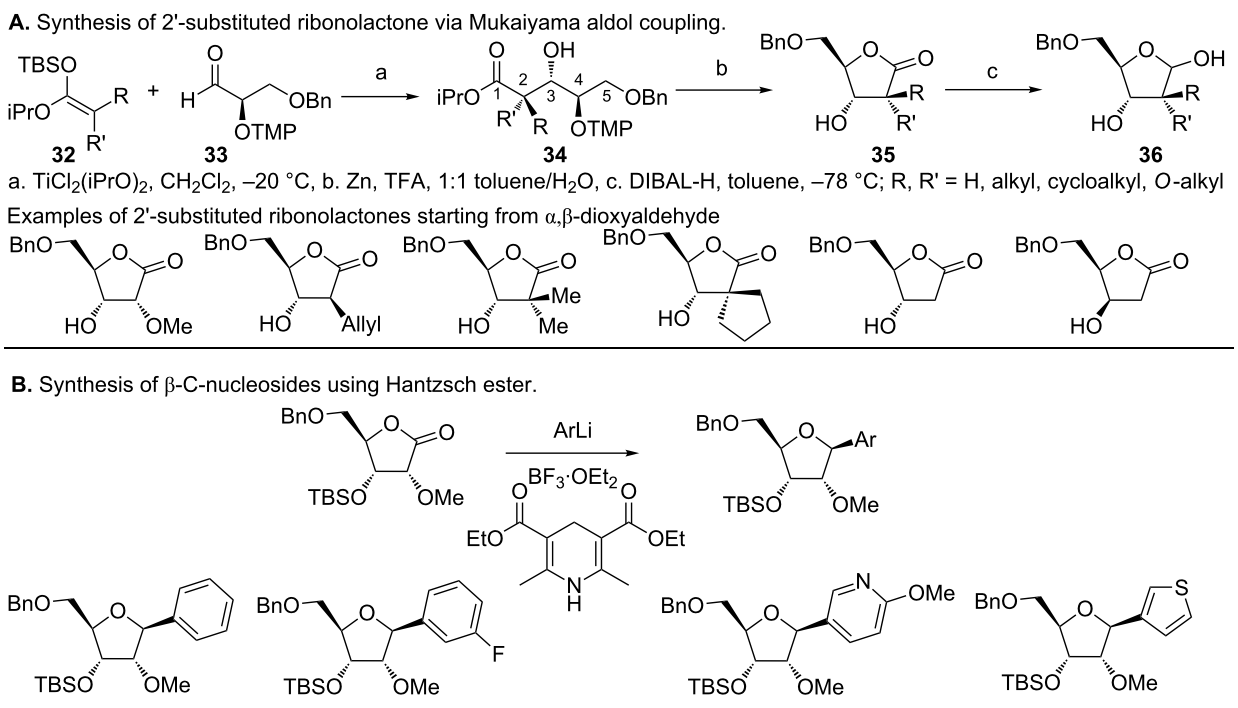


Figure 11: Alternative method for synthesis of 2'-substituted C-nucleosides [75]. A. Synthesis of C2'-substituted D-ribonolactone via Mukaiyama aldol reaction. A series of 2'-O-alkyl, alkyl, cycloalkyl and deoxy D-ribonolactone were synthesized using this method. B. Use of Hantzsch ester to obtain the β-anomer of C-nucleosides.

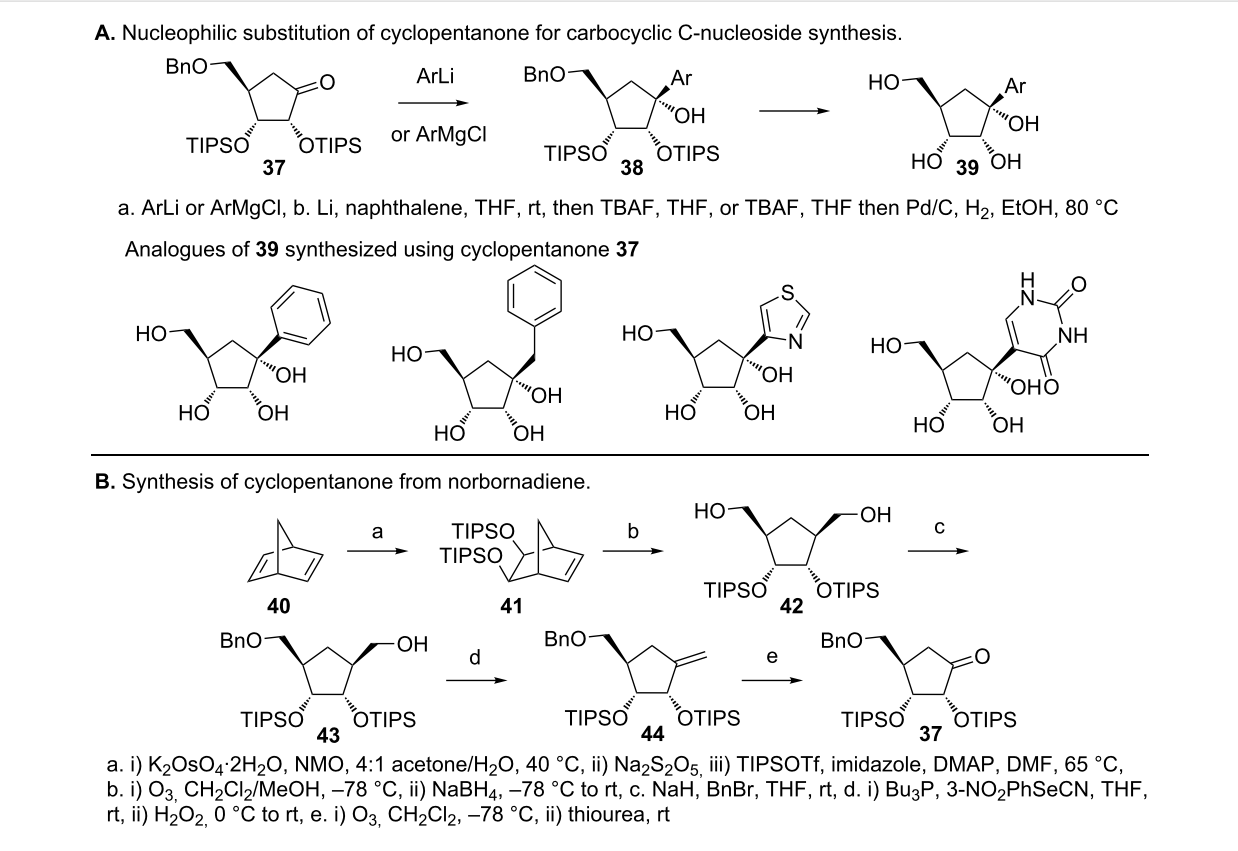


Figure 12: Synthesis of carbocyclic C-nucleosides using cyclopentanone [53]. A. Nucleophlic substitution on cyclopentanone gives C1'–OH carbocyclic C-nucleosides. B. Synthesis of cyclopentanone from norbornadiene and substituents that facilitate carbocyclic C-nucleoside syntheses.

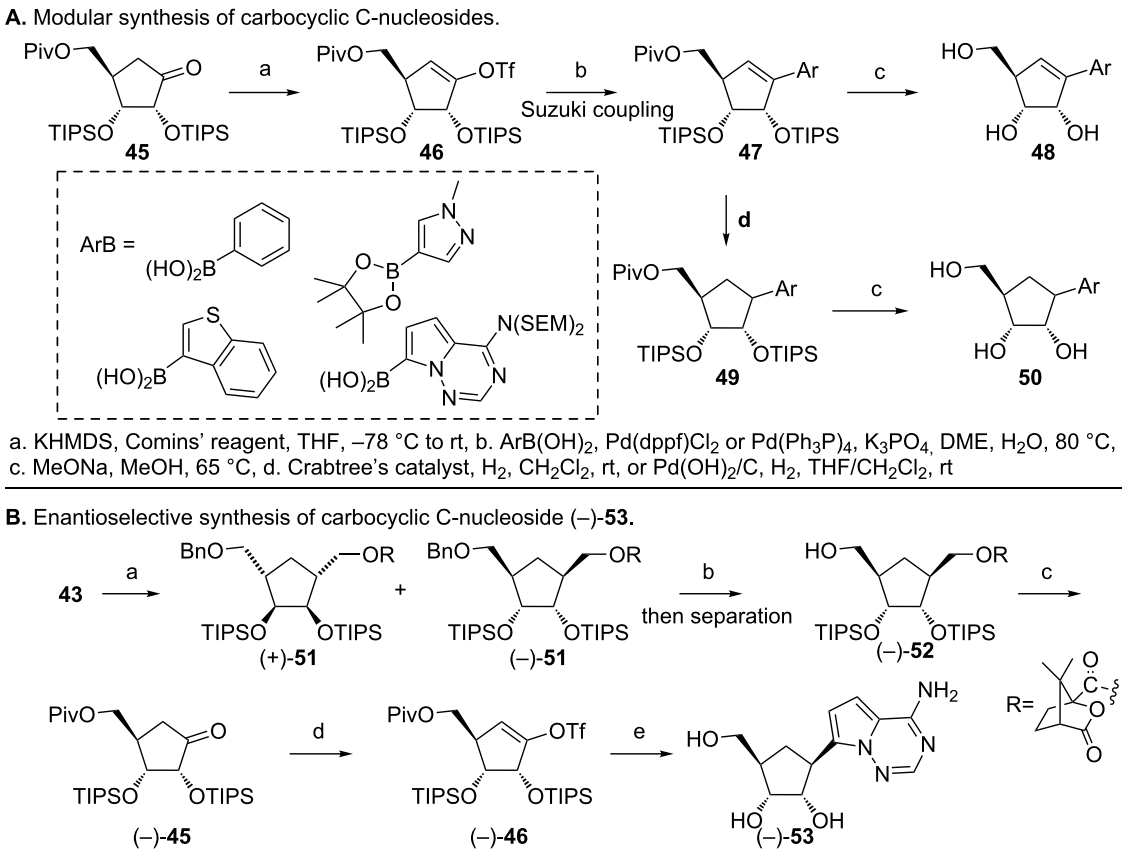


Figure 13: Synthesis of carbocyclic C-nucleosides via Suzuki coupling [53]. A. Synthesis of OTf-cyclopentene that enable Suzuki coupling and modular synthesis of carbocyclic C-nucleosides. B. Synthesis of enantiomerically pure 4-aza-7,9-dideazaadenine carbocyclic C-nucleoside.

result obtained when the C2', C3' and C5'–OHs were protected with TIPS and pivaloyl groups, respectively [53]. The cross-coupling reaction gave the unsaturated compounds (**47** and **48**), which, upon hydrogenation in presence of Crabtree's catalyst, gave the saturated compounds with the desired diasteroselectivity (**49** and **50**). In the case of nitrogen containing heterocycles, $Pd(OH)_2$ was found to be a suitable catalyst that gave a separable mixture of diastereomers (2:1). In addition, optically pure cyclopentanone (−)-**45** was obtained by converting the cyclopentane **43** to camphanates (**51,** Figure 13B) followed by separation of the diastereomers [53]. Subsequent synthesis of the enol triflate (−)-**46** (Figure 13B), Suzuki coupling and hydrogenation afforded the optically pure carbocyclic tubercidine analogue (−)-**53**. This compound has shown potent activity against breast cancer cell lines and human foreskin fibroblasts [53].

Conclusion

With increasing reports of emerging and reemerging infectious diseases globally, there is a need to develop more effective and safer drugs. In that regard, C-nucleosides have recently shown great potential, which in turn, has resurrected interest in this class of molecules [29]. Several antiviral C-nucleosides have been discovered in the past five years and are now in advanced stages of clinical applications. The overarching features of these compounds with regards to changes in the nucleobase and sugars allow optimal interactions with enzymes resulting in potent and often times, selective, inhibitory activities [18,65,74,96]. As continuing efforts to design greater diversity in C-nucleosides, methods of their synthesis have become critical to more effective drug discovery. For example, the pyrrolo[2,1-*f*][1,2,4]triazine scaffold has been key to the discovery of several highly active molecules [53,69,71-73,86].

Modular and convergent synthetic routes have proved valuable in this regard both in terms of increasing diversity and reducing the time and length of the syntheses [70-73,76]. Efforts have been aided by advances in the synthesis of modified sugars and sugar mimics, particularly D-ribonolactone analogues [53,73,75,97]. Furthermore, chemical and theoretical studies have elucidated the mechanism and stereochemical preferences of reactions involving D-ribonolactone [80,81,84,85]. Therefore, the chemist has better control over the reactions with more predictable outcomes. In the coming years, new applications may be reported. Moreover, with the biological potential of C-nucleosides now being revisited, studies of naturally occurring C-nucleosides and their biosynthetic pathways have garnered renewed interest, as has the pursuit of new biosynthetic C-nucleosides [98-104]. Previously reported C-nucleosides are also being revisited and may be repurposed with increased knowledge of new biological targets [29,65,86,96]. In summary, these efforts, in concert with improved synthetic advances, provide strong impetus for the next wave of C-nucleoside design and the discovery of nucleoside therapeutics.

ORCID® iDs

Katherine L. Seley-Radtke - https://orcid.org/0000-0002-0154-3459

References

1. Miescher, F. *Med. Chem. Unters.* **1871,** 441–460.
2. Avery, O. T.; Macleod, C. M.; McCarty, M. *J. Exp. Med.* **1944,** *79,* 137–158. doi:10.1084/jem.79.2.137
3. Chargaff, E.; Lipshitz, R. *J. Am. Chem. Soc.* **1953,** *75,* 3658–3661. doi:10.1021/ja01111a016
4. Franklin, R. E.; Gosling, R. G. *Nature* **1953,** *172,* 156–157. doi:10.1038/172156a0
5. Watson, J. D.; Crick, F. H. C. *Nature* **1953,** *171,* 737–738. doi:10.1038/171737a0
6. Watson, J. D.; Crick, F. H. C. *Nature* **1953,** *171,* 964–967. doi:10.1038/171964b0
7. Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry,* 4th ed.; Freemand and Company, 2005.
8. Matera, A. G.; Terns, R. M.; Terns, M. P. *Nat. Rev. Mol. Cell Biol.* **2007,** *8,* 209–220. doi:10.1038/nrm2124
9. Rhodes, D.; Lipps, H. J. *Nucleic Acids Res.* **2015,** *43,* 8627–8637. doi:10.1093/nar/gkv862
10. Hamma, T.; Ferre-D'Amare, A. R. *Chem. Biol.* **2006,** *13,* 1125–1135. doi:10.1016/j.chembiol.2006.09.009
11. McDonald, M. K.; Miracco, E. J.; Chen, J.; Xie, Y.; Mueller, E. G. *Biochemistry* **2011,** *50,* 426–436. doi:10.1021/bi101737z
12. Boettcher, T.; Sieber, S. A. *J. Am. Chem. Soc.* **2010,** *132,* 6964–6972. doi:10.1021/ja909150y
13. Jordheim, L. P.; Durantel, D.; Zoulim, F.; Dumontet, C. *Nat. Rev. Drug Discovery* **2013,** *12,* 447–464. doi:10.1038/nrd4010
14. Coats, S. J.; Garnier-Amblard, E. C.; Amblard, F.; Ehteshami, M.; Amiralaei, S.; Zhang, H.; Zhou, L.; Boucle, S. R. L.; Lu, X.; Bondada, L.; Shelton, J. R.; Li, H.; Liu, P.; Li, C.; Cho, J. H.; Chavre, S. N.; Zhou, S.; Mathew, J.; Schinazi, R. F. *Antiviral Res.* **2014,** *102,* 119–147. doi:10.1016/j.antiviral.2013.11.008
15. Pradere, U.; Garnier-Amblard, E. C.; Coats, S. J.; Amblard, F.; Schinazi, R. F. *Chem. Rev.* **2014,** *114,* 9154–9218. doi:10.1021/cr5002035
16. Wiegmann, D.; Koppermann, S.; Wirth, M.; Niro, G.; Leyerer, K.; Ducho, C. *Beilstein J. Org. Chem.* **2016,** *12,* 769–795. doi:10.3762/bjoc.12.77
17. Serpi, M.; Ferrari, V.; Pertusati, F. *J. Med. Chem.* **2016,** *59,* 10343–10382. doi:10.1021/acs.jmedchem.6b00325
18. Maffioli, S. I.; Zhang, Y.; Degen, D.; Carzaniga, T.; Del Gatto, G.; Serina, S.; Monciardini, P.; Mazzetti, C.; Guglierame, P.; Candiani, G.; Chiriac, A. I.; Facchetti, G.; Kaltofen, P.; Sahl, H.-G.; Dehò, G.; Donadio, S.; Ebright, R. H. *Cell* **2017,** *169,* 1240–1248. doi:10.1016/j.cell.2017.05.042
19. Davis, J. T.; Spada, G. P. *Chem. Soc. Rev.* **2007,** *36,* 296–313. doi:10.1039/B600282J
20. Jones, M. R.; Seeman, N. C.; Mirkin, C. A. *Science* **2015,** *347,* 1260901. doi:10.1126/science.1260901
21. Krishnan, Y.; Simmel, F. C. *Angew. Chem., Int. Ed.* **2011,** *50,* 3124–3156. doi:10.1002/anie.200907223
22. Fenniri, H.; Temburnikar, K. W.; Johnson, R. S. In *Comprehensive Supramolecular Chemistry II;* Atwood, J. L., Ed.; Elsevier: Oxford, 2017; pp 83–113. doi:10.1016/B978-0-12-409547-2.12583-1
23. Cihak, A. *Oncology* **1974,** *30,* 405–422. doi:10.1159/000224981
24. Lan, T.; McLaughlin, L. W. *Biochemistry* **2001,** *40,* 968–976. doi:10.1021/bi0015212
25. Geyer, C. R.; Battersby, T. R.; Benner, S. A. *Structure* **2003,** *11,* 1485–1498. doi:10.1016/j.str.2003.11.008
26. Benner, S. A. *Acc. Chem. Res.* **2004,** *37,* 784–797. doi:10.1021/ar040004z
27. Benner, S. A.; Sismour, A. M. *Nat. Rev. Genet.* **2005,** *6,* 533–543. doi:10.1038/nrg1637
28. Romeo, G.; Chiacchio, U.; Corsaro, A.; Merino, P. *Chem. Rev.* **2010,** *110,* 3337–3370. doi:10.1021/cr800464r
29. De Clercq, E. *J. Med. Chem.* **2016,** *59,* 2301–2311. doi:10.1021/acs.jmedchem.5b01157
30. Kool, E. T. *Annu. Rev. Biophys. Biomol. Struct.* **2001,** *30,* 1–22. doi:10.1146/annurev.biophys.30.1.1
31. Berti, P. J.; McCann, J. A. B. *Chem. Rev.* **2006,** *106,* 506–555. doi:10.1021/cr040461t
32. Ho, M.-C.; Shi, W.; Rinaldo-Matthis, A.; Tyler, P. C.; Evans, G. B.; Clinch, K.; Almo, S. C.; Schramm, V. L. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107,* 4805–4812. doi:10.1073/pnas.0913439107
33. Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Correa, I. R.; Romesberg, F. E. *Nature* **2014,** *509,* 385–388. doi:10.1038/nature13314
34. Malyshev, D. A.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2015,** *54,* 11930–11944. doi:10.1002/anie.201502890
35. Rios, A. C.; Yu, H. T.; Tor, Y. *J. Phys. Org. Chem.* **2015,** *28,* 173–180. doi:10.1002/poc.3318
36. Lenz, S. A. P.; Kohout, J. D.; Wetmore, S. D. *J. Phys. Chem. B* **2016,** *120,* 12795–12806. doi:10.1021/acs.jpcb.6b09620
37. Wellington, K. W.; Benner, S. A. *Nucleosides, Nucleotides Nucleic Acids* **2006,** *25,* 1309–1333. doi:10.1080/15257770600917013

38. Lindahl, T.; Karlstrom, O. *Biochemistry* **1973,** *12,* 5151–5154. doi:10.1021/bi00749a020
39. Lindahl, T.; Nyberg, B. *Biochemistry* **1974,** *13,* 3405–3410. doi:10.1021/bi00713a035
40. Lindahl, T. *Nature* **1993,** *362,* 709–715. doi:10.1038/362709a0
41. Levy, M.; Miller, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95,* 7933–7938.
42. Schroeder, G. K.; Wolfenden, R. *Biochemistry* **2007,** *46,* 13638–13647. doi:10.1021/bi701480f
43. Stockbridge, R. B.; Schroeder, G. K.; Wolfenden, R. *Bioorg. Chem.* **2010,** *38,* 224–228. doi:10.1016/j.bioorg.2010.05.003
44. Slater, A. G.; Hu, Y.; Yang, L.; Argent, S. P.; Lewis, W.; Blunt, M. O.; Champness, N. R. *Chem. Sci.* **2015,** *6,* 1562–1569. doi:10.1039/C4SC03531C
45. Davis, F. F.; Allen, F. W. *J. Biol. Chem.* **1957,** *227,* 807–815.
46. Michelson, A. M.; Cohn, W. E. *Biochemistry* **1962,** *1,* 490–495. doi:10.1021/bi00909a020
47. Cortese, R.; Kammen, H. O.; Spengler, S. J.; Ames, B. N. *J. Biol. Chem.* **1974,** *249,* 1103–1108.
48. Daves, G. D., Jr.; Cheng, C. C. *Prog. Med. Chem.* **1976,** *13,* 303–349. doi:10.1016/S0079-6468(08)70141-3
49. Hacksell, U.; Daves, G. D., Jr. *Prog. Med. Chem.* **1985,** *22,* 1–65. doi:10.1016/S0079-6468(08)70228-5
50. Samuelsson, T.; Boren, T.; Johansen, T. I.; Lustig, F. *J. Biol. Chem.* **1988,** *263,* 13692–13699.
51. Townsend, L. B., Ed. *Chemistry of Nucleosides and Nucleotides;* 1994; Vol. 3.
52. Maier, L.; Hylse, O.; Necas, M.; Trbusek, M.; Ytre-Arne, M.; Dalhus, B.; Bjoras, M.; Paruch, K. *Tetrahedron Lett.* **2014,** *55,* 3713–3716. doi:10.1016/j.tetlet.2014.05.030
53. Maier, L.; Khirsariya, P.; Hylse, O.; Adla, S. K.; Černová, L.; Poljak, M.; Krajčovičová, S.; Weis, E.; Drapelá, S.; Souček, K.; Paruch, K. *J. Org. Chem.* **2017,** *82,* 3382–3402. doi:10.1021/acs.joc.6b02594
54. Lim, M. I.; Ren, W. Y.; Otter, B. A.; Klein, R. S. *J. Org. Chem.* **1983,** *48,* 780–788. doi:10.1021/jo00154a005
55. Chu, M. Y.; Zuckerman, L. B.; Sato, S.; Crabtree, G. W.; Bogden, A. E.; Lim, M. I.; Klein, R. S. *Biochem. Pharmacol.* **1984,** *33,* 1229–1234. doi:10.1016/0006-2952(84)90174-6
56. Joubert, N.; Pohl, R.; Klepetárova, B.; Hocek, M. *J. Org. Chem.* **2007,** *72,* 6797–6805. doi:10.1021/jo0709504
57. Berger, M.; Ogawa, A. K.; McMinn, D. L.; Wu, Y.; Schultz, P. G.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2000,** *39,* 2940–2942. doi:10.1002/1521-3773(20000818)39:16<2940::AID-ANIE2940>3.0.CO;2-#
58. Kim, H.-J.; Leal, N. A.; Hoshika, S.; Benner, S. A. *J. Org. Chem.* **2014,** *79,* 3194–3199. doi:10.1021/jo402665d
59. Teo, Y. N.; Kool, E. T. *Chem. Rev.* **2012,** *112,* 4221–4245. doi:10.1021/cr100351g
60. Daves, G. D., Jr. *Acc. Chem. Res.* **1990,** *23,* 201–206. doi:10.1021/ar00174a006
61. Krohn, K.; Heins, H.; Wielckens, K. *J. Med. Chem.* **1992,** *35,* 511–517. doi:10.1021/jm00081a012
62. Štambaský, J.; Hocek, M.; Kočovský, P. *Chem. Rev.* **2009,** *109,* 6729–6764. doi:10.1021/cr9002165
63. Cho, A.; Zhang, L.; Xu, J.; Lee, R.; Butler, T.; Metobo, S.; Aktoudianakis, V.; Lew, W.; Ye, H.; Clarke, M.; Doerffler, E.; Byun, D.; Wang, T.; Babusis, D.; Carey, A. C.; German, P.; Sauer, D.; Zhong, W.; Rossi, S.; Fenaux, M.; McHutchison, J. G.; Perry, J.; Feng, J.; Ray, A. S.; Kim, C. U. *J. Med. Chem.* **2014,** *57,* 1812–1825. doi:10.1021/jm400201a
64. Warren, T. K.; Jordan, R.; Lo, M. K.; Ray, A. S.; Mackman, R. L.; Soloveva, V.; Siegel, D.; Perron, M.; Bannister, R.; Hui, H. C.; Larson, N.; Strickley, R.; Wells, J.; Stuthman, K. S.; Van Tongeren, S. A.; Garza, N. L.; Donnelly, G.; Shurtleff, A. C.; Retterer, C. J.; Gharaibeh, D.; Zamani, R.; Kenny, T.; Eaton, B. P.; Grimes, E.; Welch, L. S.; Gomba, L.; Wilhelmsen, C. L.; Nichols, D. K.; Nuss, J. E.; Nagle, E. R.; Kugelman, J. R.; Palacios, G.; Doerffler, E.; Neville, S.; Carra, E.; Clarke, M. O.; Zhang, L.; Lew, W.; Ross, B.; Wang, Q.; Chun, K.; Wolfe, L.; Babusis, D.; Park, Y.; Stray, K. M.; Trancheva, I.; Feng, J. Y.; Barauskas, O.; Xu, Y.; Wong, P.; Braun, M. R.; Flint, M.; McMullan, L. K.; Chen, S.-S.; Fearns, R.; Swaminathan, S.; Mayers, D. L.; Spiropoulou, C. F.; Lee, W. A.; Nichol, S. T.; Cihlar, T.; Bavari, S. *Nature* **2016,** *531,* 381–385. doi:10.1038/nature17180
65. Siegel, D.; Hui, H. C.; Doerffler, E.; Clarke, M. O.; Chun, K.; Zhang, L.; Neville, S.; Carra, E.; Lew, W.; Ross, B.; Wang, Q.; Wolfe, L.; Jordan, R.; Soloveva, V.; Knox, J.; Perry, J.; Perron, M.; Stray, K. M.; Barauskas, O.; Feng, J. Y.; Xu, Y.; Lee, G.; Rheingold, A. L.; Ray, A. S.; Bannister, R.; Strickley, R.; Swaminathan, S.; Lee, W. A.; Bavari, S.; Cihlar, T.; Lo, M. K.; Warren, T. K.; Mackman, R. L. *J. Med. Chem.* **2017,** *60,* 1648–1661. doi:10.1021/acs.jmedchem.6b01594
66. Townsend, L. B., Ed. *Chemistry of Nucleosides and Nucleotides;* 1988; Vol. 1.
67. Lim, M.-I.; Klein, R. S. *Tetrahedron Lett.* **1981,** *22,* 25–28. doi:10.1016/0040-4039(81)80031-7
68. Ren, W. Y.; Lim, M. I.; Otter, B. A.; Klein, R. S. *J. Org. Chem.* **1982,** *47,* 4633–4637. doi:10.1021/jo00145a005
69. Cho, A.; Saunders, O. L.; Butler, T.; Zhang, L.; Xu, J.; Vela, J. E.; Feng, J. Y.; Ray, A. S.; Kim, C. U. *Bioorg. Med. Chem. Lett.* **2012,** *22,* 2705–2707. doi:10.1016/j.bmcl.2012.02.105
70. Metobo, S. E.; Xu, J.; Saunders, O. L.; Butler, T.; Aktoudianakis, E.; Cho, A.; Kim, C. U. *Tetrahedron Lett.* **2012,** *53,* 484–486. doi:10.1016/j.tetlet.2011.11.055
71. Draffan, A. G.; Frey, B.; Pool, B.; Gannon, C.; Tyndall, E. M.; Lilly, M.; Francom, P.; Hufton, R.; Halim, R.; Jahangiri, S.; Bond, S.; Nguyen, V. T. T.; Jeynes, T. P.; Wirth, V.; Luttick, A.; Tilmanis, D.; Thomas, J. D.; Pryor, M.; Porter, K.; Morton, C. J.; Lin, B.; Duan, J.; Kukolj, G.; Simoneau, B.; McKercher, G.; Lagace, L.; Amad, M. a.; Bethell, R. C.; Tucker, S. P. *ACS Med. Chem. Lett.* **2014,** *5,* 679–684. doi:10.1021/ml500077j
72. Draffan, A. G.; Frey, B.; Fraser, B. H.; Pool, B.; Gannon, C.; Tyndall, E. M.; Cianci, J.; Harding, M.; Lilly, M.; Hufton, R.; Halim, R.; Jahangiri, S.; Bond, S.; Jeynes, T. P.; Nguyen, V. T. T.; Wirth, V.; Luttick, A.; Tilmanis, D.; Pryor, M.; Porter, K.; Morton, C. J.; Lin, B.; Duan, J.; Bethell, R. C.; Kukolj, G.; Simoneau, B.; Tucker, S. P. *Bioorg. Med. Chem. Lett.* **2014,** *24,* 4984–4988. doi:10.1016/j.bmcl.2014.09.030
73. Dang, Q.; Zhang, Z.; Chen, T.; Tang, B.; He, X.; He, S.; Song, Y.; Bogen, S.; Girijavallabhan, V.; Olsen, D. B.; Meinke, P. T. *Tetrahedron Lett.* **2014,** *55,* 5092–5095. doi:10.1016/j.tetlet.2014.07.037

74. Wang, G.; Wan, J.; Hu, Y.; Wu, X.; Prhavc, M.; Dyatkina, N.; Rajwanshi, V. K.; Smith, D. B.; Jekle, A.; Kinkade, A.; Symons, J. A.; Jin, Z.; Deval, J.; Zhang, Q.; Tam, Y.; Chanda, S.; Blatt, L.; Beigelman, L. *J. Med. Chem.* **2016,** *59,* 4611–4624. doi:10.1021/acs.jmedchem.5b01933
75. Peifer, M.; Berger, R.; Shurtleff, V. W.; Conrad, J. C.; MacMillan, D. W. C. *J. Am. Chem. Soc.* **2014,** *136,* 5900–5903. doi:10.1021/ja502205q
76. Evans, G. B.; Schramm, V. L.; Tyler, P. C. *Curr. Med. Chem.* **2015,** *22,* 3897–3909. doi:10.2174/0929867322666150821100851
77. Cho, A.; Zhang, L.; Xu, J.; Babusis, D.; Butler, T.; Lee, R.; Saunders, O. L.; Wang, T.; Parrish, J.; Perry, J.; Feng, J. Y.; Ray, A. S.; Kim, C. U. *Bioorg. Med. Chem. Lett.* **2012,** *22,* 4127–4132. doi:10.1016/j.bmcl.2012.04.065
78. Silveira, G. P.; Cardozo, H. M.; Rossa, T. A.; Sá, M. M. *Curr. Org. Synth.* **2015,** *12,* 584–602. doi:10.2174/1570179412051508211301 47
79. Batra, H.; Moriarty, R. M.; Penmasta, R.; Sharma, V.; Stanciuc, G.; Staszewski, J. P.; Tuladhar, S. M.; Walsh, D. A.; Datla, S.; Krishnaswamy, S. *Org. Process Res. Dev.* **2006,** *10,* 484–486. doi:10.1021/op050222n
80. van Rijssel, E. R.; van Delft, P.; Lodder, G.; Overkleeft, H. S.; van der Marel, G. A.; Filippov, D. V.; Codée, J. D. C. *Angew. Chem., Int. Ed.* **2014,** *53,* 10381–10385. doi:10.1002/anie.201405477
81. van Rijssel, E. R.; van Delft, P.; van Marle, D. V.; Bijvoets, S. M.; Lodder, G.; Overkleeft, H. S.; van der Marel, G. A.; Filippov, D. V.; Codée, J. D. C. *J. Org. Chem.* **2015,** *80,* 4553–4565. doi:10.1021/acs.joc.5b00419
82. Matulic-Adamic, J.; Beigelman, L.; Portmann, S.; Egli, M.; Usman, N. *J. Org. Chem.* **1996,** *61,* 3909–3911. doi:10.1021/jo960091b
83. Czernecki, S.; Ville, G. *J. Org. Chem.* **1989,** *54,* 610–612. doi:10.1021/jo00264a020
84. Larsen, C. H.; Ridgway, B. H.; Shaw, J. T.; Woerpel, K. A. *J. Am. Chem. Soc.* **1999,** *121,* 12208–12209. doi:10.1021/ja993349z
85. Larsen, C. H.; Ridgway, B. H.; Shaw, J. T.; Smith, D. M.; Woerpel, K. A. *J. Am. Chem. Soc.* **2005,** *127,* 10879–10884. doi:10.1021/ja0524043
86. Patil, S. A.; Otter, B. A.; Klein, R. S. *Tetrahedron Lett.* **1994,** *35,* 5339–5342. doi:10.1016/S0040-4039(00)73494-0
87. Eldrup, A. B.; Allerson, C. R.; Bennett, C. F.; Bera, S.; Bhat, B.; Bhat, N.; Bosserman, M. R.; Brooks, J.; Burlein, C.; Carroll, S. S.; Cook, P. D.; Getty, K. L.; MacCoss, M.; McMasters, D. R.; Olsen, D. B.; Prakash, T. P.; Prhavc, M.; Song, Q.; Tomassini, J. E.; Xia, J. *J. Med. Chem.* **2004,** *47,* 2283–2295. doi:10.1021/jm030424e
88. Shi, J.; Du, J.; Ma, T.; Pankiewicz, K. W.; Patterson, S. E.; Tharnish, P. M.; McBrayer, T. R.; Stuyver, L. J.; Otto, M. J.; Chu, C. K.; Schinazi, R. F.; Watanabe, K. A. *Bioorg. Med. Chem.* **2005,** *13,* 1641–1652. doi:10.1016/j.bmc.2004.12.011
89. Wauchope, O. R.; Tomney, M. J.; Pepper, J. L.; Korba, B. E.; Seley-Radtke, K. L. *Org. Lett.* **2010,** *12,* 4466–4469. doi:10.1021/ol101482h
90. Carroll, S. S.; Tomassini, J. E.; Bosserman, M.; Getty, K.; Stahlhut, M. W.; Eldrup, A. B.; Bhat, B.; Hall, D.; Simcoe, A. L.; LaFemina, R.; Rutkowski, C. A.; Wolanski, B.; Yang, Z.; Migliaccio, G.; De Francesco, R.; Kuo, L. C.; MacCoss, M.; Olsen, D. B. *J. Biol. Chem.* **2003,** *278,* 11979–11984. doi:10.1074/jbc.M210914200
91. Migliaccio, G.; Tomassini, J. E.; Carroll, S. S.; Tomei, L.; Altamura, S.; Bhat, B.; Bartholomew, L.; Bosserman, M. R.; Ceccacci, A.; Colwell, L. F.; Cortese, R.; De Francesco, R.; Eldrup, A. B.; Getty, K. L.; Hou, X. S.; LaFemina, R. L.; Ludmerer, S. W.; MacCoss, M.; McMasters, D. R.; Stahlhut, M. W.; Olsen, D. B.; Hazuda, D. J.; Flores, O. A. *J. Biol. Chem.* **2003,** *278,* 49164–49170. doi:10.1074/jbc.M305041200
92. Clark, J. L.; Hollecker, L.; Mason, J. C.; Stuyver, L. J.; Tharnish, P. M.; Lostia, S.; McBrayer, T. R.; Schinazi, R. F.; Watanabe, K. A.; Otto, M. J.; Furman, P. A.; Stec, W. J.; Patterson, S. E.; Pankiewicz, K. W. *J. Med. Chem.* **2005,** *48,* 5504–5508. doi:10.1021/jm0502788
93. Sofia, M. J.; Bao, D.; Chang, W.; Du, J.; Nagarathnam, D.; Rachakonda, S.; Reddy, P. G.; Ross, B. S.; Wang, P.; Zhang, H.-R.; Bansal, S.; Espiritu, C.; Keilman, M.; Lam, A. M.; Steuer, H. M. M.; Niu, C.; Otto, M. J.; Furman, P. A. *J. Med. Chem.* **2010,** *53,* 7202–7218. doi:10.1021/jm100863x
94. Chang, W.; Bao, D.; Chun, B.-K.; Naduthambi, D.; Nagarathnam, D.; Rachakonda, S.; Reddy, P. G.; Ross, B. S.; Zhang, H.-R.; Bansal, S.; Espiritu, C. L.; Keilman, M.; Lam, A. M.; Niu, C.; Steuer, H. M.; Furman, P. A.; Otto, M. J.; Sofia, M. J. *ACS Med. Chem. Lett.* **2011,** *2,* 130–135. doi:10.1021/ml100209f
95. Simonovich, S. P.; Van Humbeck, J. F.; MacMillan, D. W. C. *Chem. Sci.* **2012,** *3,* 58–61. doi:10.1039/C1SC00556A
96. Warren, T. K.; Wells, J.; Panchal, R. G.; Stuthman, K. S.; Garza, N. L.; Van Tongeren, S. A.; Dong, L.; Retterer, C. J.; Eaton, B. P.; Pegoraro, G.; Honnold, S.; Bantia, S.; Kotian, P.; Chen, X.; Taubenheim, B. R.; Welch, L. S.; Minning, D. M.; Babu, Y. S.; Sheridan, W. P.; Bavari, S. *Nature* **2014,** *508,* 402–405. doi:10.1038/nature13027
97. Bergeron-Brlek, M.; Meanwell, M.; Britton, R. *Nat. Commun.* **2015,** *6,* 6903. doi:10.1038/ncomms7903
98. Blauenburg, B.; Oja, T.; Klika, K. D.; Metsä-Ketelä, M. *ACS Chem. Biol.* **2013,** *8,* 2377–2382. doi:10.1021/cb400384c
99. Thapa, K.; Oja, T.; Metsae-Ketelae, M. *FEBS J.* **2014,** *281,* 4439–4449. doi:10.1111/febs.12950
100. Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V. *Nature* **2014,** *515,* 143–146. doi:10.1038/nature13802
101. Friedt, J.; Leavens, F. M. V.; Mercier, E.; Wieden, H.-J.; Kothe, U. *Nucleic Acids Res.* **2014,** *42,* 3857–3870. doi:10.1093/nar/gkt1331
102. Veerareddygari, G. R.; Singh, S. K.; Mueller, E. G. *J. Am. Chem. Soc.* **2016,** *138,* 7852–7855. doi:10.1021/jacs.6b04491
103. Palmu, K.; Rosenqvist, P.; Thapa, K.; Ilina, Y.; Siitonen, V.; Baral, B.; Makinen, J.; Belogurov, G.; Virta, P.; Niemi, J.; Metsä-Ketelä, M. *ACS Chem. Biol.* **2017,** *12,* 1472–1477. doi:10.1021/acschembio.7b00078
104. Ko, Y.; Wang, S.-A.; Ogasawara, Y.; Ruszczycky, M. W.; Liu, H.-w. *Org. Lett.* **2017,** *19,* 1426–1429. doi:10.1021/acs.orglett.7b00355

License and Terms

The definitive version of this article is the electronic one which can be found at:
doi:10.3762/bjoc.14.65

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Phosphodiester models for cleavage of nucleic acids

Satu Mikkola*, Tuomas Lönnberg* and Harri Lönnberg*

Review

Open Access

Address:
Department of Chemistry, University of Turku, FIN-20014 Turku, Finland

Email:
Satu Mikkola* - satkuu@utu.fi; Tuomas Lönnberg* - tuanlo@utu.fi; Harri Lönnberg* - harlon@utu.fi

* Corresponding author

Keywords:
Cleavage; DNA; kinetics; mechanism; RNA

Beilstein J. Org. Chem. **2018,** *14,* 803–837.
doi:10.3762/bjoc.14.68

Received: 09 November 2017
Accepted: 12 March 2018
Published: 10 April 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

Nucleic acids that store and transfer biological information are polymeric diesters of phosphoric acid. Cleavage of the phosphodiester linkages by protein enzymes, nucleases, is one of the underlying biological processes. The remarkable catalytic efficiency of nucleases, together with the ability of ribonucleic acids to serve sometimes as nucleases, has made the cleavage of phosphodiesters a subject of intensive mechanistic studies. In addition to studies of nucleases by pH-rate dependency, X-ray crystallography, amino acid/nucleotide substitution and computational approaches, experimental and theoretical studies with small molecular model compounds still play a role. With small molecules, the importance of various elementary processes, such as proton transfer and metal ion binding, for stabilization of transition states may be elucidated and systematic variation of the basicity of the entering or departing nucleophile enables determination of the position of the transition state on the reaction coordinate. Such data is important on analyzing enzyme mechanisms based on synergistic participation of several catalytic entities. Many nucleases are metalloenzymes and small molecular models offer an excellent tool to construct models for their catalytic centers. The present review tends to be an up to date summary of what has been achieved by mechanistic studies with small molecular phosphodiesters.

Introduction

Nucleic acids are polymeric diesters of phosphoric acid that store and transfer biological information. In biological systems, the diester linkages bridging 3′-O of one nucleoside to the 5′-O of the next one are cleaved by a variety of enzymes [1]. The phosphodiester bonds of DNA are hydrolyzed, depending on the enzyme, either to a 3′- or 5′-phosphate, whereas the bonds in RNA, with few exceptions (above all RNase H-catalyzed cleavages) undergo transesterification to a 2′,3′-cyclic phosphate that is rapidly hydrolyzed to 2′- and 3′-phosphates (Figure 1). In the absence of any catalyst, the 3′,5′-phosphodiester linkages are remarkably stable under physiological conditions. The half-life for the hydrolysis of an individual phosphodiester bond in DNA has been estimated to be 30 million years at 25 °C, which means that protein enzymes, nucleases, are able to accelerate the phosphodiester cleavage by a factor of 10^{17} [2]. The phosphodiester linkages of RNA are much more labile,

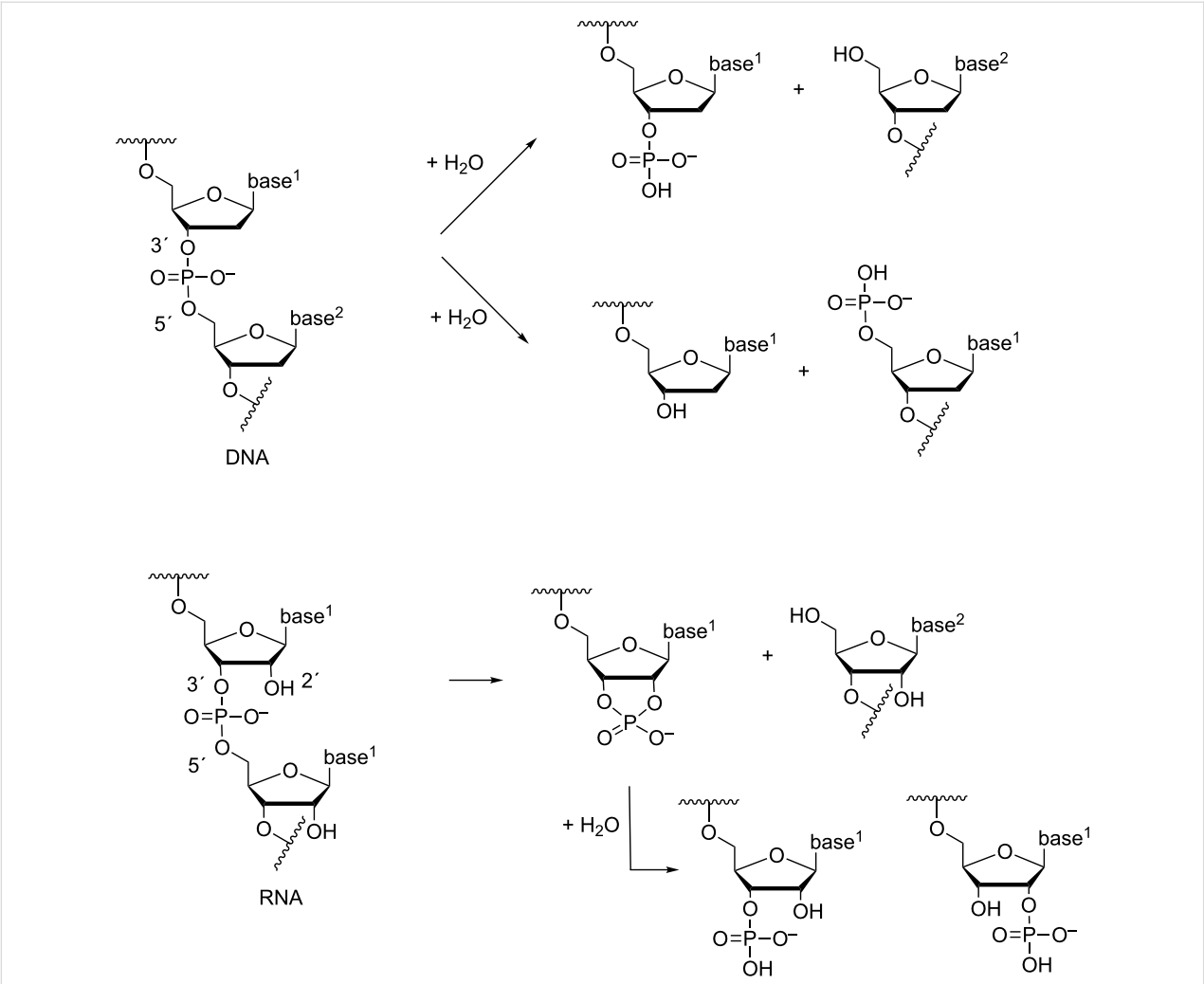


Figure 1: Enzymatic cleavage of phosphodiester linkages of DNA and RNA.

owing to the presence of neighboring hydroxy function that serves as an intramolecular nucleophile resulting in transphosphorylation by departure of the 5′-linked nucleoside [3]. The half-life at pH 6–7 and 25 °C is around 10 years [4,5], the enzymatic cleavage by RNase A being $3{\cdot}10^{11}$ times faster [6]. Interestingly, the RNA phosphodiester bonds are additionally subject to cleavage by RNA itself, viz. by RNA sequences known as ribozymes [7]. The length of these catalytic sequences varies from 70–150 nucleotides of the so-called small ribozymes to hundreds of nucleotides of large ribozymes. Their catalytic efficiency is somewhat more modest than that of protein enzymes.

The remarkable catalytic efficiency has made the action of protein nucleases and ribozymes a subject of intensive mechanistic studies. pH-Rate dependency, X-ray structures, amino acid/nucleotide substitution experiments and the effect of thiosubstitution of phosphate oxygens on the binding of metal ion cofactors have given invaluable information about the residues that participate in substrate binding or contribute to formation of high-energy intermediates or transition states during the PO-bond cleavage by protein nucleases [8] or ribozymes [9,10]. Based on this data, energetics of various pathways from the reactants to products may be compared by computational methods [11-14]. Still, experimental studies with small molecular model compounds play an essential role in mechanistic studies of the enzymatic cleavage of nucleic acids. With small molecules, the importance of various elementary processes, such as proton transfer and metal ion binding, for stabilization of transition states may be elucidated and systematic variation of the basicity of the entering and departing nucleophile enables determination of the position of the transition state on the reaction coordinate. Such data is important on analyzing enzyme mechanisms based on synergistic participation of several catalytic entities. Similar studies are not possible with enzymes, since even a minor change in the structure of enzyme or substrate may have a dramatic effect on the structure and stability

of the enzyme–substrate complex. In addition, the kinetic data obtained with small molecules is useful for testing the validity of computational methods utilized for the generation of energy landscapes for enzyme catalysis [15-17].

Many nucleases are metalloenzymes containing two catalytically active metal ions. Small molecular models offer an excellent tool to study the cooperative action of metal ions and to construct models for catalytic centers [11,18].

Review

Basic principles of phosphoryl transfer reactions

Non-enzymatic cleavage of phosphodiester linkages of nucleic acids proceeds by an intra- (RNA) or intermolecular (DNA) nucleophilic attack on phosphorus. The reaction proceeds via a pentacoordinated species having the structure of a trigonal bipyramid. In case this species represents an energy maximum on a single barrier energy profile, as with S_N2 displacement at carbon, the reaction is called concerted and the pentacoordinated species is a transition state. The reaction is a synchronous displacement (A_ND_N) when bond formation to the entering nucleophile is as advanced as bond fission to the departing nucleophile (A in Figure 2). In case the bond formation is more or less advanced than the bond fission, the reaction still is concerted but has an associative or dissociative nature, respectively. The pentacoordinated species, called pentaoxyphosphorane, may also have a sufficiently long life-time to represent a minimum on the energy profile. The reaction then proceeds in a stepwise manner. It is an associative nucleophilic displacement ($A_N + D_N$) with late transition state if the barrier for breakdown of the phosphorane intermediate to products is higher than the barrier for formation of the intermediate (B in Figure 2). If the barrier for the phosphorane formation is higher than the barrier

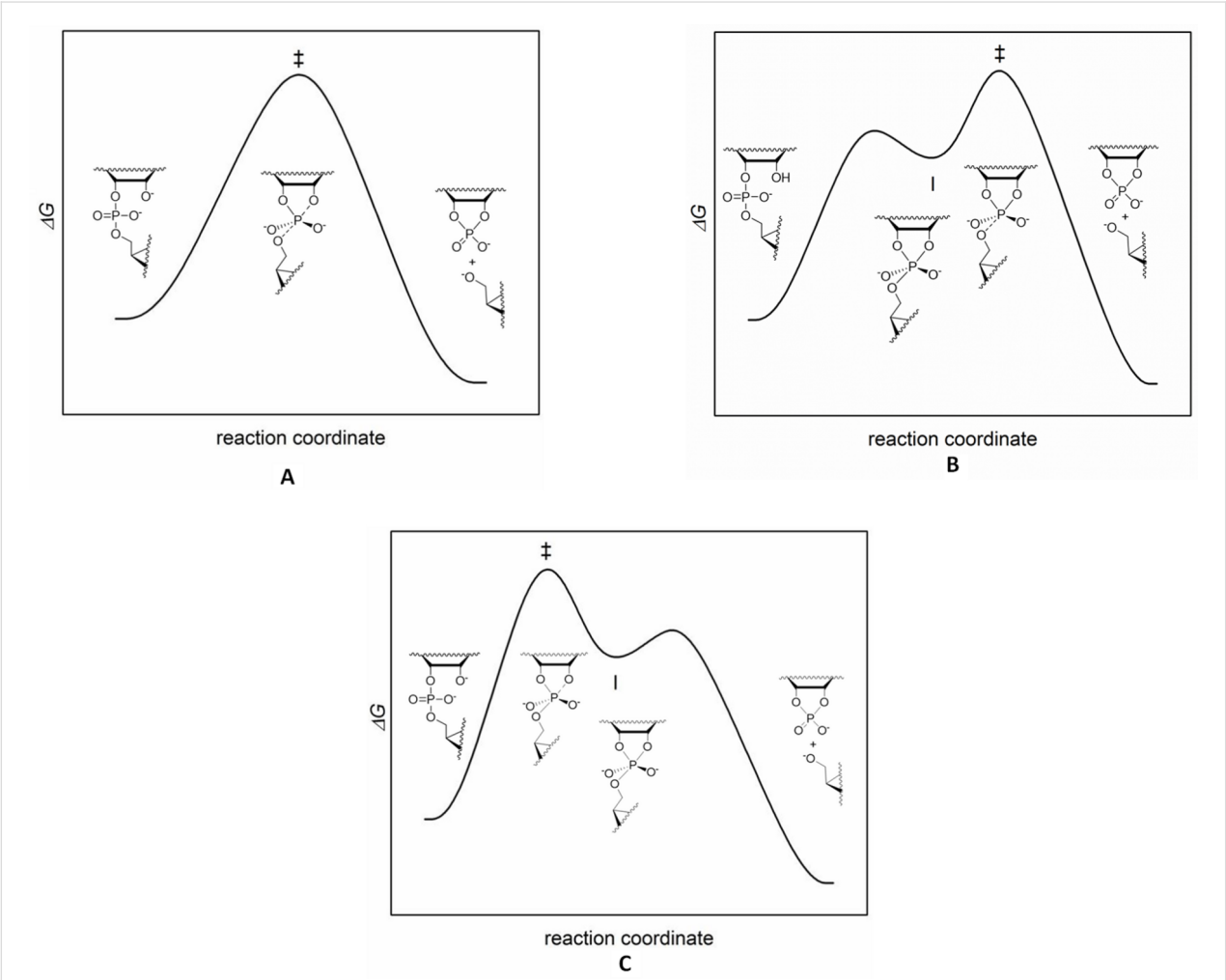


Figure 2: Energy profiles for a concerted A_ND_N (A) and stepwise mechanisms ($A_N + D_N$) with rate-limiting breakdown (B) and rate-limiting formation (C) of intermediate **I** that has a finite life-time. Hydroxide-ion-catalyzed cleavage of RNA has been used to exemplify alternative mechanisms. In reality, the reaction takes place by rate-limiting breakdown of the intermediate (B).

for its breakdown to products, the transition state is early and formation of the phosphorane is rate-limiting (C in Figure 2). The phosphorane intermediate may still have a finite life-time, but experimental distinguishing between this kind of a reaction and a concerted displacement is difficult.

Two of the ligands within the bipyrimidal phosphorane take an apical (*a* in Figure 3) and the rest an equatorial (*e* in Figure 3) position. According to the so-called Westheimer´s rules [19], nucleophiles enter and depart the phosphorane intermediate only through an apical position. Electronegative ligands prefer an apical position, while negatively charged oxygens are locked to an equatorial position. Bulky ligands tend to be equatorial. If two of the oxygen atoms are bridged by an ethylene group, as in the phosphorane obtained by the attack of 2´-OH of RNA on phosphorus, one must be apical and the other equatorial. A sufficiently stable phosphorane may, however, undergo a structural change known as Berry pseudorotation [20]: one of the equatorial ligands remains equatorial, while the rest turn apical and the apical ligands equatorial. Several alternative models for isomerization of trigonal-bipyramidal pentacoordinate compounds have been presented [21], but Berry pseudorotation has almost exclusively used in mechanistic discussion of RNA cleavage.

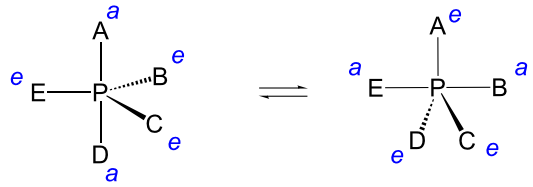


Figure 3: Pseudorotation of a trigonal bipyramidal phosphorane intermediate by Berry pseudorotation [20].

The stability of the phosphorane intermediate largely depends on its state of protonation. The first pK_a value of the acyclic tetraalkoxy monohydroxy phosphorane has been estimated to be 8.6 for an equatorial hydroxy group and 13.5 for an apical group [22]. For a cyclic phosphorane derived from ethylene phosphate, the first pK_a value is 7.9 and the second 14.3, both values referring to an equatorial hydroxy ligand [23]. Accordingly, both neutral phosphorane and its monoanion are present in significant amount at physiological pH. In case a dianionic phosphorane is formed, its protonation to a monoanion expectedly is thermodynamically favored, but it is not clear whether the life-time is long enough to allow this.

The cyclic phosphorane intermediate of RNA cleavage is in neutral form (**IH_2** in Figure 4) sufficiently stable to pseudorotate [24]. According to DFT calculations, the barrier for preudorotation is 10 kcal mol^{-1} lower than the barriers for breakdown of the intermediate [25]. The calculations also suggest the monoanionic form (**IH^-**) to be able to pseudorotate, even more rapidly than the neutral form [26]. The breakdown of the phosphorane is, however, also faster than with neutral phosphorane and, hence, the life-time of the monoanion is shorter. The dianionic phosphorane (**I^{2-}**) is very unstable and cannot pseudorotate, owing to the high barrier for transfer of negatively charged oxygen from equatorial to apical position. Recent DFT calculations suggest the barrier to be about 30 kcal mol^{-1} [27].

While several lines of evidence suggest that the cleavage of the RNA phosphodiester bonds proceeds via a phosphorane intermediate rather than a phosphorane-like transition state [28-30], this is not necessarily the case with DNA that is cleaved by an attack of an external nucleophile. Recent hybrid quantum mechanical/effective fragment potential (QM/EEP) calculations on the hydroxide-ion-catalyzed hydrolysis of diethyl phosphate monoanion, however, suggest that the acyclic phosphorane obtained still is an intermediate [31]. The lifetime for the dianionic pentacoordinated species obtained by the attack of the hydroxide ion on the phosphorus has been argued to represent an energy minimum between the transition states for the attack of HO^- and the departure of EtO^- and to have a lifetime of 1 picosecond. With leaving groups that are less basic than EtO^-, such as 5´-O^- of nucleoside, the lifetime expectedly is shorter. If the leaving group is very good, such as an aryl group, a synchronous concerted mechanism (A_ND_N) may take over the stepwise mechanism ($A_N + D_N$).

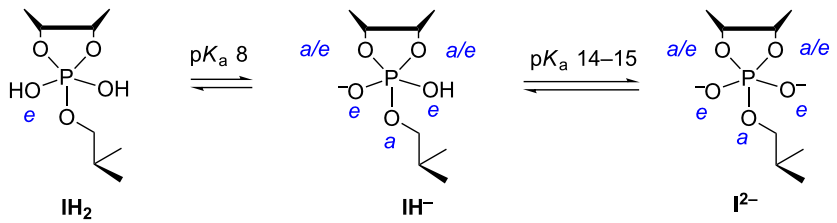


Figure 4: Protolytic equilibria of phosphorane intermediate of RNA transesterification.

Model compounds and experimental tools

Studies with phosphodiester models are aimed at providing firm mechanistic understanding of the hydrolysis and transesterification reactions of nucleic acids. Such information is indispensable for critical evaluation of mechanistic proposals of more complicated enzymatic processes and for the development of artificial cleaving agents that have enzyme-like catalytic properties but are more robust. pH-Rate profiles, linear free energy relationships and kinetic heavy atom isotope effects are the experimental approaches that are, together with construction of multifunctional cleaving agents, most extensively used in mechanistic studies of small molecular phosphodiester models. Kinetic studies over a wide pH-range allow division of observed rate constants to contributions of different ionic forms and, hence, the upper limit for the effect of protonation or deprotonation of a particular atom on the rate is obtained [29,32]. Linear free energy relationships are, in turn, used to determine the position of transition state on the reaction coordinate [33]. The polar property of either entering or departing nucleophile or non-departing groups is altered in a systematic manner and the effect on reaction rate is compared to the effect on the equilibrium of the reaction. In this manner, information about charge distribution in the transition state is obtained; whether the transition state is early (close to starting materials) or late (close to products). A free energy relationship is in principle a plot of activation free energy, $\Delta G^{\ddagger}$ (or log k), against the change in standard free energy of the reaction, ΔG^{o} (or log K_{eq}). The latter quantity is often difficult, sometimes even impossible, to determine. For this reason, $\Delta G^{\ddagger}$ (or log k) is more frequently plotted as a function of the pK_a of the departing (or entering) nucleophile. The slope of the plot, known as a β_{lg} (or β_{nuc}), may have values greater than unity. It does not directly tell the position of transition state on the reaction coordinate. This parameter, the so-called Leffler´s α, is, however, obtained as a ratio of β_{lg}/β_{eq} or β_{nuc}/β_{eq}, if a reasonably reliable estimate for the β value of the equilibrium reaction, β_{eq}, is available. As long as cleavage of phosphodiesters is concerned, β_{eq} = 1.74 reported for the phosphoryl transfer of phosphono monoanion is usually used as the reference value for the equilibrium reaction [34]. Likewise, the occurrence of the proton transfer as part of the rate limiting step may be evaluated by altering the acidity of the proton donor (or acceptor). Plotting of log k against the pK_a of the proton donor (or acceptor) gives the Brönsted α (β for the acceptor) that refers to the extent of proton transfer in the transition state.

The kinetic heavy atom isotope effect (KIE) is a most useful tool for mechanistic studies, especially since it may be used as well in enzymatic and non-enzymatic reactions [35,36]. Replacing a single atom in the substrate with its heavy isotope has so small influence on structure that enzyme–substrate interaction is not distorted, which is the case with other structural modifications. Kinetic isotope effect is defined as the ratio of the rate constants obtained with the light and heavy isotope containing compound, KIE = $^{light}k/^{heavy}k$. When this ratio is greater than unity, the isotope effect is called normal, otherwise inverse. KIE refers to the difference in bonding that takes place on going from ground state to transition state. The effect is a primary KIE when the isotopically labelled atom is directly involved in bond making or bond breaking in the rate-limiting step. In case the isotopic substitution occurs further in the molecule, the KIE is secondary. The primary KIE is usually normal (>1), while the secondary can be either normal or inverse. The reason is that KIE consists of two contributions, a temperature independent (TIF) and temperature dependent (TDF) factor [37]. As regards the primary KIEs, the motion along the reaction coordinate is the predominant source of KIE. The KIE for this process is normal and largely dominated by TIF. With secondary KIEs, motion along the reaction coordinate is less important and changes in TDF-dependent vibrational modes of the transition state start to play a role. That is why both normal and inverse effects are possible.

The kinetic solvent isotope effect (KSIE) is another mechanistic tool frequently used to distinguish between alternative mechanisms. KSIE is an indication of a kinetically significant proton transfer that takes place on going from initial to transition state and shows up as reactivity difference in experiments made in H_2O and D_2O solutions of equal pL (L = H or D). The proton transfer may, however, take place either in pre-equilibrium or rate-limiting stage. Distinguishing between these alternatines is possible, if the equilibrium isotope effect for the pre-equilibrium may be reliably estimated. In case no KSIE is observed, no proton transfer takes place in the rate-limiting step. Proton inventory studies are used to examine how many protons are transferred in the rate-limiting step. In this technique, rate constants are determined as a function of isotopic ratio n, and the shape of a plot k_n/k_o vs n gives information on the proton transfer processes. Unfortunately, interpretation of the data is not always straightforward, owing to possible contribution of the equilibrium isotope effect that refers to binding of the catalyst to the phosphate group [27,38].

Dinucleoside-3´,5´-monophosphates are obvious small molecular models with which to study the cleavage of phosphodiester linkages in nucleic acids. Kinetic studies with these compounds are, however, somewhat laborious, since HPLC chromatography has to be used to analyze the content of samples withdrawn at suitable intervals. That is why many research groups prefer to use a simpler model, 2-hydroxypropyl *p*-nitrophenyl phosphate (HPNP; **1**, Figure 5), the hydrolysis of which can be followed by UV-spectrophotometry. A lot of useful observations have

been done with this simple model. One should, however, bear in mind that the *p*-nitrophenoxy group is a 10^8 times better leaving group than a 5′-linked nucleoside and, hence, the rate limiting step of these two reactions can well be different, as discussed later in more detail below. In addition, the acyclic structure only poorly mimics the ribofuranosyl structure of the 3′-linked nucleoside. The acyclic analog **2**, for example, is cleaved under basic conditions 500 times less readily than a normal diribonucleoside-3′,5′-monophosphate [39]. A small molecular catalyst may accelerate the cleavage of **1** by stabilizing a rotamer that favors intramolecular attack of the neighboring hydroxy function on phosphorus, while this kind of acceleration evidently plays a minor role, if any, with ribonucleoside 3′-phosphodiesters. Finally, phosphate migration in **1** takes place between a primary and secondary hydroxy group, whereas with ribonucleoside 3′-phosphodiesters both hydroxy functions are secondary. Accordingly, extrapolation of the results obtained with **1** to the cleavage of nucleic acids is not straightforward. Care should be exercised to avoid misinterpretations.

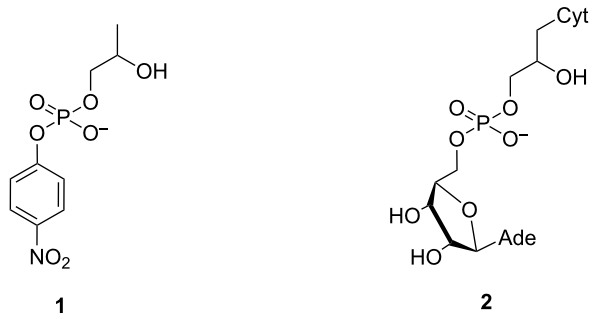


Figure 5: Structures of acyclic analogs of ribonucleosides.

Oligonucleotides containing a thiosubstituted nucleotide are extensively used in mechanistic studies of protein nucleases and ribozymes. Rate accelerating 3′-bridging substitution has been used to find out whether the chemical step really is rate-liming and 5′-substitution to verify that some small ribozymes utilize general acid catalysis [40]. The underlying idea behind the latter application is that protonation of the leaving group by a general acid is not needed with 5′-thiosubstituted analogs, since the sulfide ion is a much better leaving group than the alkoxide ion. Most extensively used thiosubstitution, however, is replacement of either one of the non-bridging oxygens with sulfur, which allows stereochemical studies based on the so-called rescue effect [41,42]. When non-bridging oxygen that participates in binding of Mg^{2+} is replaced with sulfur, the activity drops, but may be restored by using a soft Lewis acid, such as Mn^{2+} or Zn^{2+}. The necessary background information for the studies with thiosubstituted oligonucleotides has been obtained by comparative studies with similar analogs of dinucleoside-3′,5′-monophosphates [43].

Cleavage of RNA by Brönsted acids and bases

Buffer-independent reactions

The predominant buffer-independent reactions of RNA phosphodiester linkages at physiological pH (pH 6–8) are pH-independent isomerization to 2′,5′-bonds (red line in Figure 6) and hydroxide-ion-catalyzed transesterification to a 2′,3′-cyclic phosphate by departure of the 5′-linked nucleoside, followed by subsequent hydrolysis to a mixture of 2′- and 3′-phosphates (blue line in Figure 6) [44,45]. These reactions are approximately as fast at pH 7, the isomerization being faster under more acidic and cleavage under more basic conditions. The occurrence of isomerization inevitably shows that the monoanionic phosphorane, most likely obtained by the attack of 2′-OH on the phosphorus atom with concomitant transfer of the proton to the non-bridging oxygen [46,47], is able to pseudorotate at physiological pH. It is not quite clear whether the pseudorotation takes place through the monoanionic species or kinetically invisible protonation to more stable neutral phosphorane. DFT calculations suggest that the monoanionic form really is stable enough to pseudorotate and the breakdown of the intermediate to 2′- or 3′-phosphodiesters is approximately as fast as the pseudorotation [25]. According to the same calculations, the exocyclic fission of the intermediate to a 2′,3′-cyclic phosphate, leading to pH-independent cleavage, is much slower (Scheme 1). The rate of this reaction (black line in Figure 6) is only 2% of the interconversion rate of 2′,5′- and 3′,5′-diesters

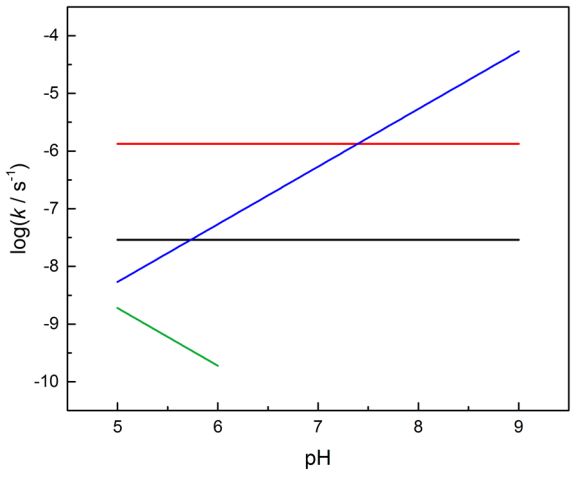
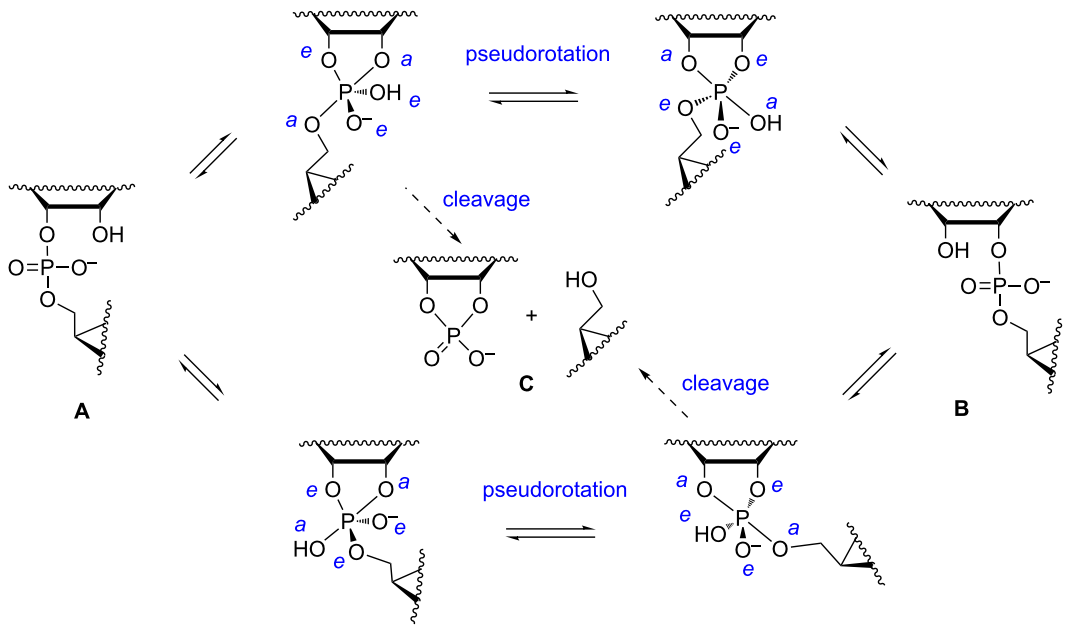


Figure 6: First-order rate constants for buffer-independent partial reactions of uridyl-3′,5′-uridine at pH 5–9 and 90 °C. Hydronium-ion-catalyzed isomerization (green), hydroxide-ion-catalyzed cleavage (blue), pH-independent cleavage (black), pH-independent isomerization (red). Based on the data from ref. [44].

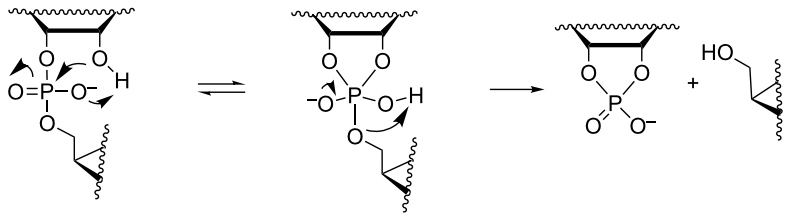


Scheme 1: pH- and buffer-independent cleavage and isomerization of RNA phosphodiester linkages. Observed first-order rate constant for the cleavage (k_{cl}) refers to transesterification of **A** + **B** to **C**, and observed rate constant for isomerization (k_{is}) to mutual isomerization of **A** and **B**, the values for the forward and reverse reactions being almost equal.

[44]. Studies with various uridine 3′-alkylphosphates have, however, verified the existence of this reaction [48].

The mechanism of the pH-independent cleavage reaction has been elucidated by comparative studies of β_{lg} values. While the isomerization rate is almost independent of the polar nature of the esterified alcohol, the cleavage rate is markedly increased with the increasing electronegativity of the alkyl group. For example, the ratio of k_{cl}/k_{is} is 0.014 and 1.8 with the ethyl and 2,2,2-trichloroethyl esters, respectively [48]. The $\beta_{lg} = -0.59$ is more negative than the $\beta_{lg} = -0.12$ of the acid-catalyzed cleavage, proceeding by departure of neutral alcohol, but less negative than the $\beta_{lg} = -1.28$ of the hydroxide-ion-catalyzed reaction where the departing group is an alkoxide ion [49]. Accordingly, the departing oxygen atom seems to become protonated concerted with rate-limiting rupture of the P–OR bond. The essential mechanistic features, hence, are proton transfer to non-bridging oxygen concerted with the attack of 2′-OH, which increases the nucleophilicity of O2′ and stabilizes the phosphorane intermediate, and proton transfer from the non-bridging oxygen to the departing oxygen, which destabilizes the phosphorane and stabilizes the leaving group (Scheme 2). Combined QM/MM simulations have lent support for this interpretation [47]. With triester analogs, such as uridine 3′-diethyl phosphate, the latter intramolecular proton transfer is not possible and the ratio k_{cl}/k_{is} is much smaller than with the diester analog, around 10^{-5} [50]. Since the barrier for the endocyclic cleavage of the phosphorane intermediate is more than 10 kcal mol^{-1} lower than that for the exocyclic cleavage, it is not clear whether a similar proton transfer from a phosphorane hydroxy ligand to the departing oxygen occurs concerted with the fission of P–O2′ and P–O3′ bonds or does protonation of these oxygens take place after the bond fission.

The hydroxide-ion-catalyzed cleavage that dominates at pH >7.5, proceeds by pre-equilibrium deprotonation of the



Scheme 2: Mechanism for the pH- and buffer-independent cleavage of RNA phosphodiester linkages.

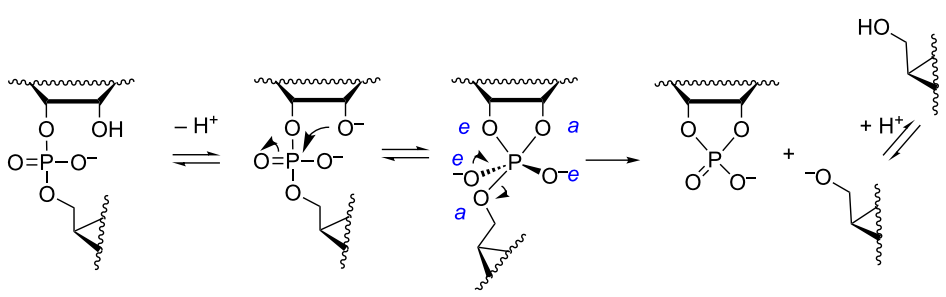
2′-OH and subsequent attack of the 2′-oxyanion on the phosphorus atom of a monoanionic phosphodiester linkage, giving a dianionic phosphorane that decomposes to 2′,3′-cyclic phosphate by departure of the 5′-linked nucleoside as an alkoxide ion (Scheme 3). The stability of the dianionic phosphorane has been studied by experimental and computational methods. As mentioned above, the β_{lg} value of the reaction of uridine 3′-alkyl phosphates is very negative, −1.28, suggesting that the cleavage of the P–O5′ bond is rather advanced in the transition state. However, the β_{lg} value obtained with uridine 3′-aryl phosphates is much less negative, −0.54 [51]. When the data of alkyl and aryl esters is included in the same free energy plot, a break at pK_a of 12.4 occurs, i.e., close to the pK_a of the attacking 2′-OH [52]. A free energy plot exhibiting a breakpoint at the pK_a of the attacking nucleophile is usually taken as a rather compelling evidence of a change in the rate-limiting step [33], in this case from the formation of the phosphorane intermediate with aryl esters to breakdown of this intermediate with alkyl esters. The results of DFT calculations lend further support to this interpretation and suggest that the 2,2,2-trichloroethoxy group is an example of an alkyl leaving group where the barrier for the formation of phosphorane intermediate still is slightly higher than the barrier for its departure [15].

Assuming that the β_{eq} = −1.7 reported for the phosphoryl transfer of phosphono monoanion [34] is valid for the hydroxide-ion-catalyzed cleavage of RNA phosphodiester bonds, the highly negative β_{lg} value, −1,28, means that Leffler′s α referring to the fraction of total bond cleavage is 0.7. The β_{nuc} value, in turn, helps to evaluate how advanced the formation of the P−O2′ bond is. This parameter has been determined by incorporating 2′-*C*-X-uridines (X = H, Me, CFH_2, CF_2H, CF_3) into an oligodeoxyribonucleotide and plotting the cleavage rate against the pK_a of the 2′-OH [53]. The value obtained, β_{nuc} = 0.75, means that the P–O2′ bond is approximately half formed (Leffler′s α ≈ 0.4–0.5) in the transition state.

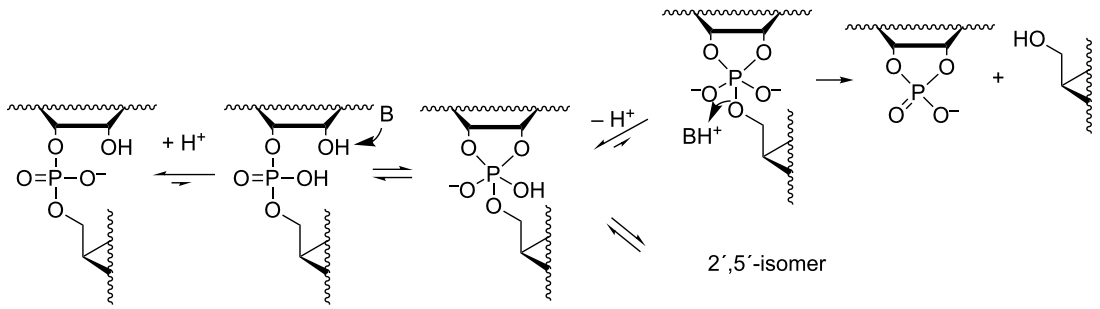
The isotope effects determined for the cleavage of 3′,5′-UpG at pH 14, i.e., under conditions where the attacking 2′-OH is almost completely deprotonated, lend further support for the mechanism in Scheme 3 [54-56]. No solvent D_2O isotope effect occurs, consistent with rapid pre-equilibrium deprotonation of the attacking 2′-OH. For the departing 5′-O, the ^{18}O KIE is normal, $^{16}k_{lg}/^{18}k_{lg}$ = 1.034 ± 0.004, and for the attacking 2′-O^-, the KIE is inverted, $^{16}k_{nuc}/^{18}k_{nuc}$ = 0.984 ± 0.004 [54]. Both effects are large and consistent with advanced P–O5′ fission and P–O2′ formation in the transition state. For comparison, with uridine 3′-(*p*-nitrophenyl phosphate), the leaving group KIE expectedly is small, $^{16}k_{lg}/^{18}k_{lg}$ = 1.0059 ± 0.0004, indicating that the departure of the aryloxy group is not markedly advanced [57]. The secondary KIE for the replacement of the non-bridging oxygen of the attacked phosphate is almost negligible, $^{16}k_{O1P}/^{18}k_{O1P}$ = 0.999 ± 0.001 [16].

Buffer-catalyzed reactions

While the mechanisms of buffer-independent reactions prevailing at physiological pH are rather well established, the buffer-catalyzed reactions still appear to be open to various mechanistic interpretations. The main reason for this is experimental difficulty. The buffer-dependent rate is rather modest compared to the buffer-independent rate. High buffer concentration has to be used and this makes elimination of salt and co-solute effects difficult. Since histidine residues are known to play a central role in the catalytic center of RNase A [58], one of the most extensively studied protein nucleases, catalysis by imidazole/imidazolium ion (Im/ImH^+) buffers has been of special interest. The pioneering studies were carried out by the group of Breslow [59]. Their mechanistic suggestion is depicted in Scheme 4. Im is argued to catalyze the attack of 2′-OH on phosphorus by serving as a general base, but only if the phosphodiester linkage has undergone rapid initial protonation. In other words, a monoanionic phosphorane is obtained by a specific acid/general base mechanism that is experimentally equivalent to general acid catalysis. The monoanionic phosphorane is stable enough to pseudorotate and may, hence, undergo isomerization to the 2′,5′-diester without additional catalysis. The cleavage reaction is, in turn, suggested to take place by pre-equilibrium deprotonation of the phosphorane intermediate, fol-



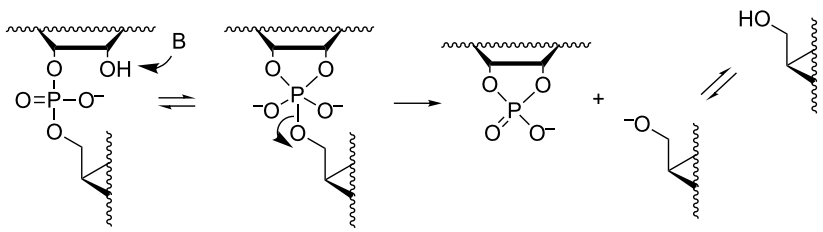
Scheme 3: Hydroxide-ion-catalyzed cleavage of RNA phosphodiester linkages.



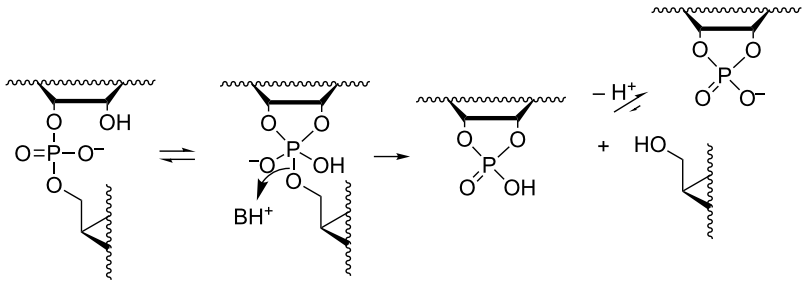
Scheme 4: Anslyn's and Breslow's mechanism for the buffer-catalyzed cleavage and isomerization of RNA phosphodiester linkages [59].

lowed by general acid-catalyzed fission of the P–O5´ bond; experimentally a general base catalysis is observed. An interesting feature of the mechanism is that both the formation and breakdown of the phosphorane intermediate proceed through a minor ionic form in a pre-equilibrium mixture. The mole fraction of neutral phosphodiester, for example, is in imidazole buffers of the order of 10^{-6} (pK_a of phosphodiester ≈ 1). This means that protonation of the phosphodiester linkage must facilitate the nucleophilic attack on phosphorus by at least a factor of 10^6. As regards deprotonation of monoanionic phosphorane, the pK_a is around 14 [23], which means that deprotonation should accelerate the general acid-catalyzed departure of the 5´-linked nucleoside by a factor of 10^7. The mechanistic proposal has partly been based on Breslow's studies on hydrolysis of 4-*tert*-butylcatechol cyclic phosphate by regioisomers of β-cyclodextrins bearing two imidazole groups [60]. This reverse reaction of the cyclization of 4-*tert*-butylcatechol 2-*O*-monophosphate has been shown to proceed via a monoanionic (monoprotonated) phosphorane and, hence, argued to lend support for the mechanism in Scheme 4. This mechanism has been criticized [61-63], but also defended by a reinvestigation [64]. According to the additional studies, the original mechanistic suggestion is in principle valid, but has to be supplemented with a general base-catalyzed reaction through a dianionic phosphorane transition state (Scheme 5) that takes place in parallel with the stepwise reaction through a phosphorane monoanion (Scheme 4).

The group of Kirby has suggested a somewhat simpler mechanism based on two concurrent reactions: rapid initial formation of a monoanionic phosphorane that undergoes rate-limiting general acid-catalyzed cleavage (Scheme 6) and the general



Scheme 5: General base-catalyzed cleavage of RNA phosphodiester bonds.



Scheme 6: Kirby´s mechanism for the buffer-catalyzed cleavage of RNA phosphodiester bonds [65].

base-catalyzed reaction through a dianionic phosphorane transition state [65].

To avoid the contribution of buffer-independent catalysis by hydroxide ions, the buffer-catalyzed cleavage of RNA models has been studied in 80% aq DMSO (v/v). The autoprotolysis constant of water is suppressed by four orders of magnitude (pK_w = 18.38) on going from water to this mixture [66], whereas the pK_a values of amines experience only a modest change [67]. Accordingly, general acid/base catalysis may be studied with amine buffers at much lower hydroxide ion concentrations than in water. This technique was first applied by the group of Yatsimirsky to cleavage of a HPNP [38]. In 0.1 mol L^{-1} piperidine buffer, for example, the buffer-catalyzed reaction was 10^3-fold faster than the buffer-independent reaction. The observed rate constant showed both first- and second-order dependence on the buffer concentration, $k_{obs} = k_1[B] + k_2[B][BH^+]$. The Brönsted β value for the first-order term was 0.77 and this reaction was suggested to be a general base-catalyzed formation of dianionic phosphorane (Scheme 5). The second-order term, which was important especially in guanidine and amidine buffer, was assumed to refer to binding of BH^+ to the anionic phosphodiester linkage more or less concerted with the general base-catalyzed attack of the 2′-OH. The situation seems, however, to be rather different with dinucleoside-3′,5′-monophosphates. The buffer-catalyzed reaction of UpU is not so much faster than the buffer-independent reaction, in 0.1 mol L^{-1} piperidine buffer only 4-fold faster [68]. No second-order dependence of rate on buffer concentration was observed. It should be, however, noted that kinetic measurements in the most interesting guanidine and amidine buffers failed, evidently owing to partial decomposition of the buffer constituents during the prolonged incubation at 90 °C. Both cleavage and isomerization were observed, but only the cleavage was subject to buffer catalysis, viz. general base catalysis. In aqueous solution, second-order dependence of rate on buffer concentration has never been reported.

Besides imidazole, guanidine and primary amines have received special interest as cleaving agents of RNA [69]. Guanidine is the side-chain functionality of arginine, an active component of the catalytic center of some nucleases, e.g., Staphylococcal nuclease [70] and topoisomerase [71]. Additionally, it is a substructure of guanine base that in hammerhead [72,73] and hairpin [74] ribozymes participates in proton transfer from the attacking 2′-OH to non-bridging phosphoryl oxygen. Primary amines are, in turn, used to mimic the action of the ε-amino group of lysine. Both guanidine and primary amino groups are basic functions that at physiological pH are present as guanidinium and ammonium ions. These ions tend to reduce electron density in their vicinity, inductively through bonds and electrostatically through space, or they may serve as weak general acids. The guanidine group may additionally participate in proton shuttling through various tautomeric forms [75] and the amino group through bifurcated H-bonds.

The first experimental observation on the ability of guanidinium containing entities to cleave RNA dates back to the early 1990s. The group of Anslyn [76] showed that compound **3** that incorporated two 2-aminoimidazolinium groups, accelerated at high micromolar concentrations the imidazole-promoted cleavage of RNA by one order of magnitude, whereas its monomeric congener **4** was ineffective (Figure 7). No detailed mechanism was suggested, but binding of **3** to the non-bridging oxygens and the departing 5′-*O* was assumed to stabilize the phosphorane intermediate and possibly protonating the departing oxygen. The second milestone on the way to guanidine-based cleaving agents was the finding that tris[2-(benzimidazol-2-ylamino)ethyl]amine (**5**) could rather rapidly degrade RNA [77]. The first-order rate constant for the cleavage of an individual phosphodiester linkage of a 30-mer RNA sequence was $3.3 \cdot 10^{-6}$ s^{-1} at [**5**] = 1 mmol L^{-1} and 37 °C. Aggregation of **5** with RNA prevented detailed mechanistic studies. The catalyst was, however, active even in the non-aggregated state, though possibly somewhat less efficient. The pK_a value of the 2-aminobenzimidazolium ion is about 7, being exceptionally low for a guanidinium compound. This low basicity was suggested to be a central factor behind the catalytic activity.

A clarification of the mechanism of guanidine-based catalysis has more recently been attempted by anchoring a 2,4-diamino-1,3,5-triazine core to the N3 of uracil bases of UpU by two side

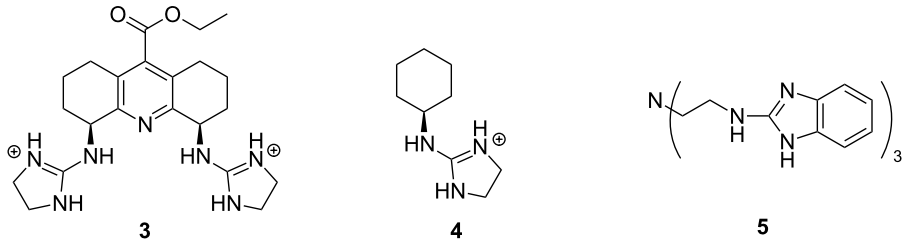
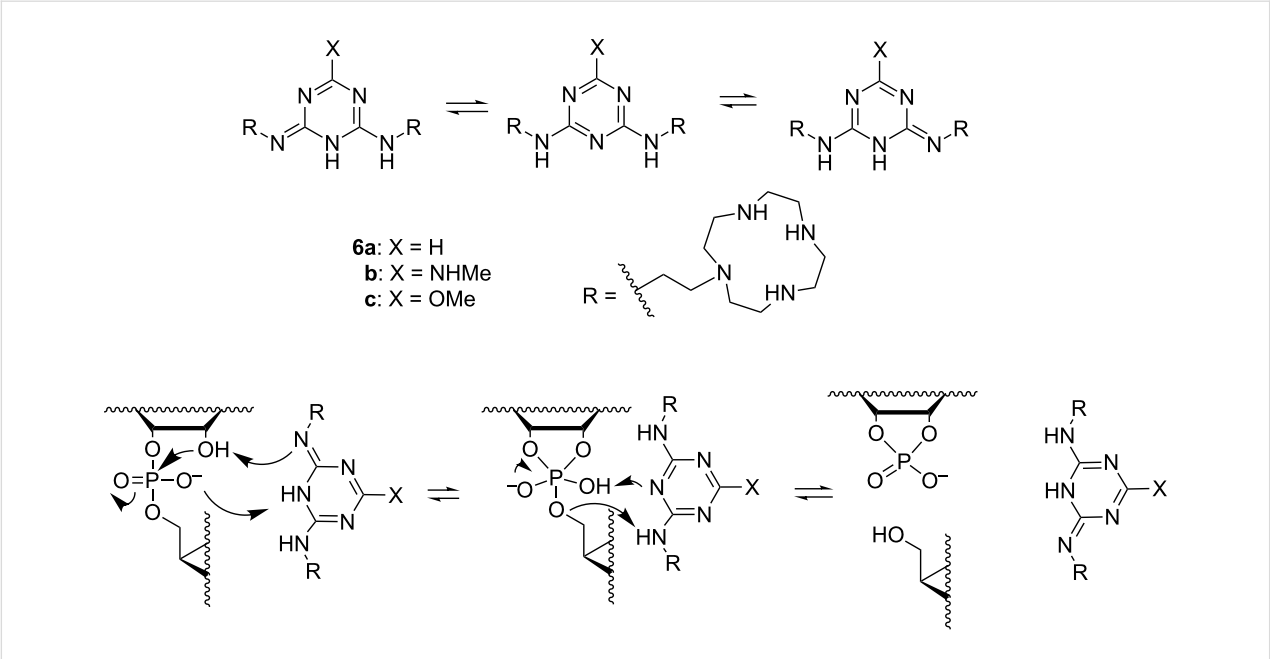


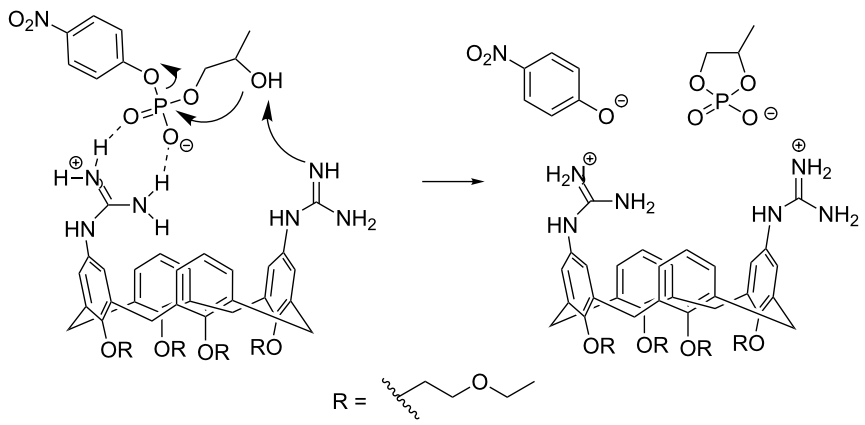
Figure 7: Guanidinium-group-based cleaving agents of RNA.

arms, each bearing a Zn^{2+}–cyclen complex (Scheme 7) [78]. The ternary complex of Zn^{2+}, UpU and **6a** was shown to be more stable than any of the binary complexes of these species. Within this ternary complex, the triazine core could interact with the phosphodiester linkage and via various tautomeric forms facilitate the proton transfer between the attacking 2′-OH, non-bridging phosphate oxygen and departing 5′-O. The scaffold still was flexible enough to allow both cleavage and isomerization of the phosphodiester linkage. In the pH range 6–8, where the triazine core remained neutral (pK_a = 3.96), the cleavage rate was pH-independent and the acceleration at pH 7 was 30-fold compared to the buffer-independent cleavage of UpU. At pH 6, the acceleration was 100-fold. By contrast, isomerization was not accelerated. The catalytic efficiency was not sensitive to the basicity of the triazine core. More basic 6-NHMe (**6b**; pK_a = 5.28) and less basic 6-OMe (**6c**; pK_a = 3.54) substituted compounds were as efficient catalysts as their unsubstituted counterpart. Scheme 7 shows the mechanism suggested to explain the insensitivity to basicity of the general base. Increasing basicity of **6** was argued to favor the pre-equilibrium proton transfer from the 2′-OH to **4**, but at the same time **4** is weakened as a general acid that donates proton to the departing 5′-O in the rate-limiting step. The leaving group effect of the triazine-catalyzed cleavage was studied with uridine 3′-(alkyl phosphates) by using as a catalyst a truncated version of **6**, bearing only one anchoring side-arm [79]. The β_{lg} = −0.7 was of the same order of magnitude as the one, −0.59, reported for the pH- and buffer-independent cleavage, where water molecules mediate the proton shuttling.

Cooperative catalysis by two guanidine groups has been demonstrated by calix[4]arene derivatives **7** bearing the guanidine groups at the upper rim and *O*-(2-ethoxyethyl) groups at the lower rim [80]. The role of the latter groups was to improve solubility to hydroxylic solvents and to rigidify the calixarene system into the so-called *cone* conformation. HPNP (**1**) was used as RNA model and the reactions were carried out in 80% aq DMSO. On using a bis(guanidine)-substituted compound as a catalyst, the maximal cleavage rate was observed at pH 10.4, where only one of the two guanidines was protonated. The 1,3-distal isomer was twice as effective as its 1,2-vicinal counterpart. At 3 mmol L^{-1} concentration, the cleavage rate was 300-fold compared to the hydroxide-ion-catalyzed background reaction. It was suggested that the protonated guanidinium group binds to the phosphate group and facilitates as an electrophilic catalyst the general base-catalyzed attack of the hydroxy function on phosphorus (Scheme 8). Similar results were obtained on using diphenylmethane as a scaffold **8** (Figure 8) [81]. A cyclohexylidene or adamantylidene substituent on the methylene carbon moderately enhanced the catalytic activity. Interestingly, the calix[4]arene-based agent **7** catalyzed the cleavage of dinucleoside-3′,5′-monophosphates in 80% DMSO even more efficiently than the cleavage HPNP, the acceleration compared to the background reaction being in most favorable cases more than 10^4-fold [78]. No saturation with the catalyst in the low millimolar range could be observed. More recent DFT calculations have led to the conclusion that replacement of the *p*-nitrophenoxide leaving group with a less electronegative nucleoside oxyanion converts the mechanism more



Scheme 7: Tautomers of triazine-based cleaving agents and cleavage of RNA phosphodiester bonds by these agents [78].



Scheme 8: Cleavage of HPNP by 1,3-distal calix[4]arene bearing two guanidine groups [80].

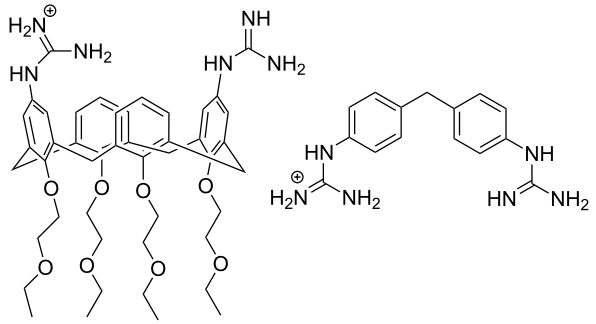


Figure 8: Bifunctional guanidine/guanidinium group-based cleaving agents of RNA.

associative, which results in more marked acceleration compared to the background reaction [27]. Dinucleoside phosphates containing uracil or guanine base were cleaved exceptionally fast [82]. No mechanistic explanation was given. Interestingly, these two bases may undergo deprotonation under mildly basic conditions ($pK_a \approx 9$) in contrast to adenine and cytosine.

Aliphatic amines are poor catalysts for the cleavage of RNA. The second-order rate constant for the ethylenediamine-catalyzed cleavage of ApA has been reported to be $1.2 \cdot 10^{-6}$ L mol^{-1} s^{-1} at pH 8 and 50 °C [83]. Cyclic polyamines are somewhat better catalysts (Figure 9). The tetracation of 1,4,16,19-tetraoxa-7,10,13,22,25,28-hexaazacyclotriacontane (**9**) cleaves ApA almost 20 times as fast as ethylenediamine, the second-order rate constant being $2 \cdot 10^{-5}$ L mol^{-1} s^{-1} at 50 °C [84]. The reason for this enhanced activity remains obscure. One may tentatively assume that the multiple positive charges play a role by stabilizing electrostatically the phosphorane intermediate and the departing 5′-alkoxide ion. 1,4-Dioxa-7,10,13-triazacyclopentadecane (**10**), a smaller congener of **9**, was catalytically inactive.

The tetracation of 1,3-bis(1,4,7,10-tetraazacyclododecan-1-ylmethyl)benzene (**11a**) catalyzes the cleavage, and also the isomerization, of UpU at physiological pH [85], the second-order rate constants for the cleavage and isomerization being $1.75 \cdot 10^{-2}$ L mol^{-1} s^{-1} and $1.5 \cdot 10^{-2}$ L mol^{-1} s^{-1} at 90 °C, respectively. The catalysis seems to be base moiety selective,

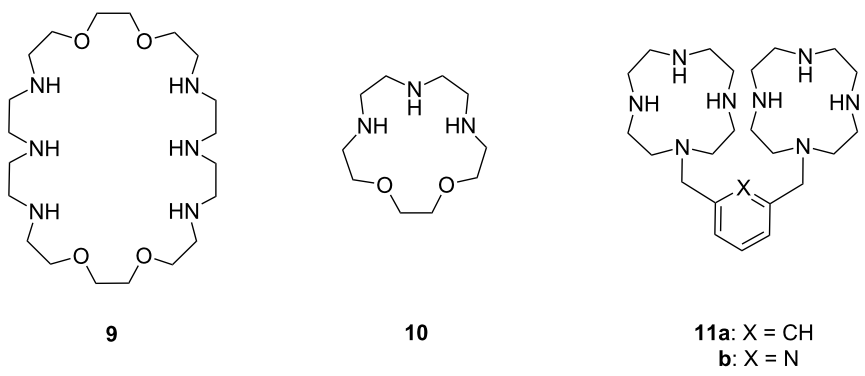
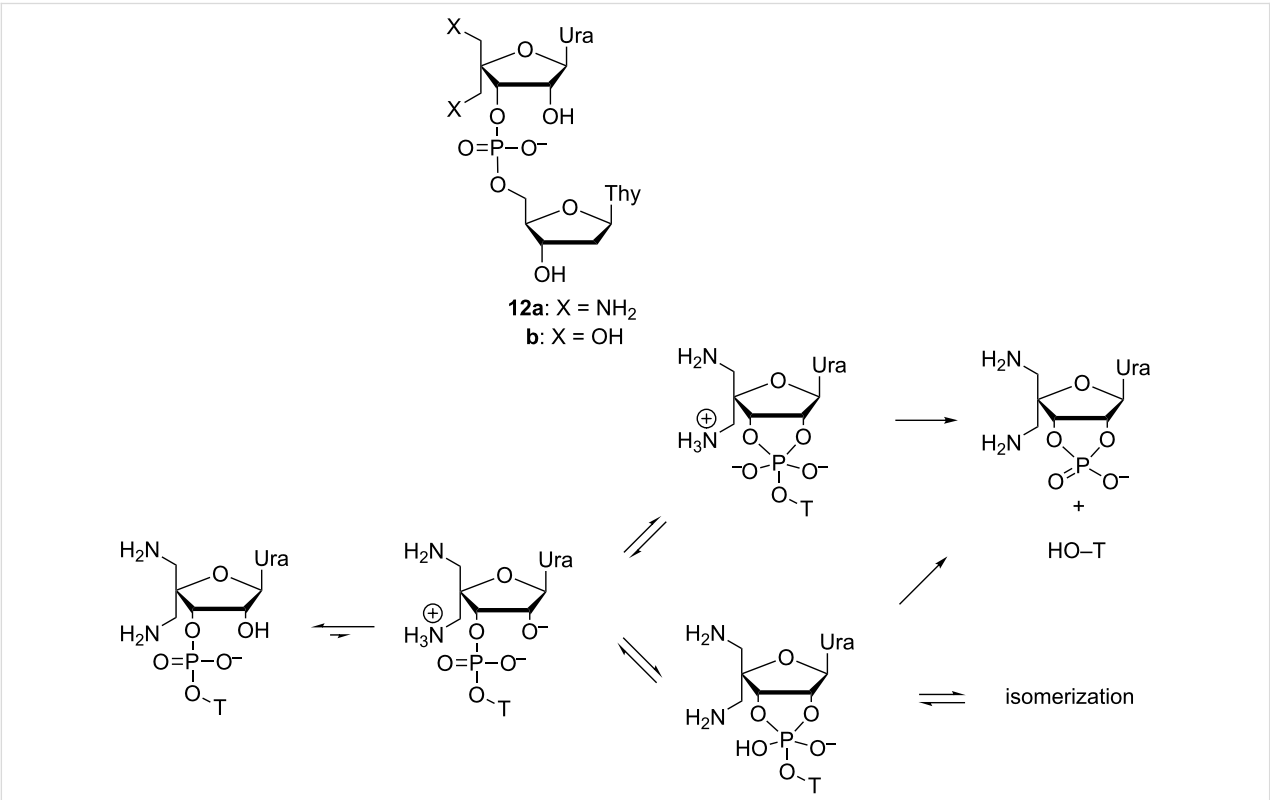


Figure 9: Cyclic amine-based cleaving agents of RNA.

since ApA is not cleaved. It has been suggested that one doubly charged cyclen moiety anchors the catalyst by hydrogen bonding to the carbonyl groups of uracil base and the other cyclen serves as an electrophilic catalyst by interacting with the phosphodiester linkage. The tetra- and penta-cations of 2,6-bis(1,4,7,10-tetraazacyclododecan-1-ylmethyl)pyridine (**11b**) have given similar results.

The possible role of the lysine ε-amino group in the catalytic center of RNase A has been elucidated by incorporating an amino group covalently in the vicinity of the scissile phosphodiester linkage of the model compound. For this purpose, compound **12a** bearing two aminomethyl groups at C4′ was prepared and its reactions were compared to the reactions of UpU [86] and 4′-hydoxymethyl-UpT (**12b**) [87]. The pK_a values for the mono- and diammonium ions of **12a** were determined to be 7.2 and 5.8, respectively. At pH 3–5, i.e., under conditions where both amino groups were protonated, both the cleavage and 3′,5′→2′,5′ isomerization of **12a** were pH-independent and almost two orders of magnitude faster than the corresponding reactions of UpU or **12b**. Since both reactions were accelerated, the ammonium ions were assumed to stabilize the common phosphorane intermediate, most likely by protonation of the initially formed phosphorane monoanion to a neural species. The proton transfer is thermodynamically favorable since the first pK_a value of the neutral phosphorane expectedly is around 8 [23].

At pH > 9, the cleavage of **12a** is hydroxide-ion-catalyzed and as fast as the respective reaction of UpU and **12b**. Over a narrow pH range 7.5–8.5, where both amino groups still are deprotonated, the behavior of **12a**, however, differs from that of UpU or **12b**; another pH-independent cleavage occurs [86]. This reaction is one order of magnitude faster than the pH-independent cleavage of **12a** at pH 3–5, i.e., when both amino groups are protonated. Compared to the pH-independent cleavage of UpU, the acceleration is 10^3-fold. It has been suggested, that the reaction proceeds through a minor tautomer having the 2′-OH deprotonated and one of the amino groups protonated, in spite of the fact that the mole fraction of this species is as low as 10^{-5}. The 2'-O^-, however, is at least a 10^6 times better nucleophile than 2′-OH [32,88]. A dianionic phosphorane is obtained that gives the cleavage products without any kinetically visible catalysis. Concurrent with this cleavage reaction, a proton transfer from protonated aminomethyl group to non-bridging oxygen takes place more or less concerted with the PO-bond formation. A monoanionic phosphorane that is stable enough to pseudorotate is formed and, hence, isomerization takes place, although less rapidly than the cleavage (Scheme 9).



Scheme 9: Mechanism for the pH-independent cleavage and isomerization of model compound **12a** in the pH-range 7.5–8.5 [86].

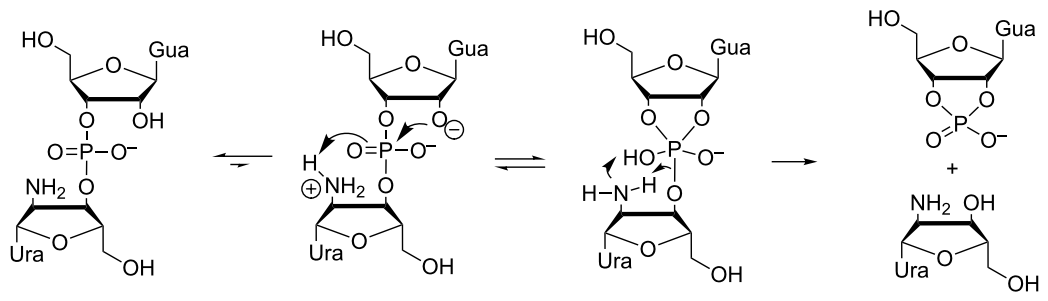
Likewise, the unexpectedly fast pH-independent cleavage of guanylyl-3′,3′-(2′-amino-2′-deoxyuridine) has been accounted for by intermediary formation of a highly reactive minor tautomer (Scheme 10) [89]. The pK_a value of the amino group is surprisingly low, 4.9 at 90 °C. Both the zwitterionic (amino group protonated) and monoanionic (amino group neutral) species undergo a pH-dependent cleavage, the former at pH 3–4 and the latter at pH 6–8. Both reactions give 2′-amino-2′-deoxyuridine as the sole free nucleoside, indicating that the attacking nucleophile in both cases is the 2′-OH of the guanylyl moiety. The pH independent cleavage of the monoanion is, however, one order of magnitude faster than the cleavage of the zwitterion. This observation has led to the conclusion that the monoanion reacts through a minor tautomer having the 2′-OH deprotonated and the amino group protonated. The protonated amino group may facilitate the attack of the 2′-oxyanion by H-bonding to one of the non-bridging oxygens concerted, but upon elongation of the P–O3′ bond, the basicity of this non-bonding oxygen is decreased and the basicity of the departing O3′ is increased. Owing to this change, the H-bond to phosphate is weakened and H-bonding to O3′ is strengthened. While the reaction at pH 6–8 is 100-times faster than the cleavage of guanylyl-3′,3′-(2,5-di-*O*-methyluridine), the isomerization reaction is not accelerated by the amino substitution and, hence, only cleavage is detected at pH > 4.

Cleavage of RNA phosphorothiolates and phosphorothioates

As discussed in the introductory part, phosphorothiolate oligonucleotides containing a bridging 3′- or 5′-thiosubstitution, are used as mechanistic probes of enzyme catalysis. Non-bridging thiosubstitution, in turn, creates R_P and S_P diastereomeric phosphorothioate linkages which have extensively been used for elucidation of the stereochemical course of enzymatic reactions and stereochemical requirements for Mg^{2+} binding. That is why, comparative kinetic studies with phosphorothioate analogs of phosphodiesters are of interest.

Bridging 3′S-substitution accelerates the hydroxide-ion-catalyzed cleavage of the phosphodiester linkage (Scheme 3) by more than two orders of magnitude, in spite of the fact that sulfur is less electronegative than oxygen and, hence, a weaker withdrawer of electrons from phosphorus [90,91]. According to theoretical calculations, the reaction is accelerated since a less strained five-membered ring is formed upon the attack of 2′-OH on phosphorus and since the polarizability of sulfur is higher than that of oxygen [16]. The heavy atom isotope effect measurements with *S*-(2-hydroxypropyl) *O*-(*m*-nitrobenzyl) phosphorothiolate have shown that the effect for the attack of the OH group, $^{18}k_{nuc}$ = (1.1188 ± 0.0055), is large, suggesting an early transition state where the PO bond formation is not markedly advanced [92]. The leaving group effect, $^{18}k_{lg}$ = (1.0118 ± 0.0003), is small but still present consistent with modest progress of the leaving group departure. In striking contrast to the situation with their oxygen counterparts, the 2′,3′-cyclic phosphorothiolate is clearly accumulated [90,93]. At pH 3–5, pH-independent isomerization of the 3′,5′- to 2′,5′-phosphorothiolate is faster than cleavage and 50 times as fast as the isomerization of its oxygen analog [93]. In other words, monoanionic 3′-thiophosphorane is stable enough to pseudorotate.

5′-Thiosubstitution accelerates the hydroxide-ion-catalyzed cleavage even more markedly than the 3′-substitution, the cleavage rate being from 10^4- to 10^5-fold compared to the oxygen analog [94,95]. With *O*-(2-hydroxypropyl) *S*-(3-nitrobenzyl) phosphorothiolate, $^{18}k_{nuc}$ = 1.0245 ± 0.0047 is normal while the leaving group heavy atom KIE, $^{34}k_{lg}$ = 1.0009 ± 0.0001, is very small, 1.0009 ± 0.0001, consistent with an early transition state with advanced formation of the PO bond and without appreciable lengthening of the PS bond [92]. In other words, the transition state resembles the transition of ribonucleoside 3′-aryl phosphates rather than 3′-alkyl phosphates, which is expected on the basis of 10^5-fold lower basicity of sulfide ions compared to alkoxide ions.



Scheme 10: Mechanism for the pH-independent cleavage of guanylyl-3′,3′-(2′-amino-2′-deoxyuridine) at pH 6-8 [89].

The effect of non-bridging thiosubstitution on the cleavage rate is modest compared to the bridging substitutions. Phosphoromonothioates react by 100% inversion, the thioeffect, k_O/k_S, for the R_P and S_p isomer being 1.3 and 0.8, respectively [96,97]. Thiosubstitution tends to stabilize the dianionic phosphorane intermediate, but at the same the solvation of the phosphorane is weakened, and these two opposing influences largely cancel each other [98-100]. The solvation, hence, plays a much more important role than with 3′S- and 5′S-substitutions, evidently for the reason that the sulfur in non-bridging position is anionic and the charge is more dispersed than with oxygen. The leaving group effect is very similar to that with the oxygen phosphodiesters, the β_{lg} values for the alkyl and aryl esters of uridine 3′-phosphate being 1.24 [101] and 0.55 [102], respectively. This also applies to the general base-catalyzed cleavage. For the imidazole-catalyzed reaction, the β_{lg} value of uridine 3′-aryl phosphorothioates and 3′-arylphosphates are 0.63 and 0.59, respectively [102]. The thio effect, k_O/k_S, is somewhat greater than in specific base catalysis, ranging from 1.2 to 3.6. Altogether, the effect of non-bridging thiosubstitution on the kinetics of RNA phosphodiesters remains very modest, which makes thioates useful model compounds for the studies of rescue effect in the catalysis by large ribozymes.

Under physiological conditions, pH-independent reactions via a monoanionic phosphorane (Scheme 2) compete with the hydroxide-ion-catalyzed cleavage. At pH 5–7, these reactions even predominate [97]. Monoanionic thiophosphorane is sufficiently stable to pseudorotate, but the isomerization is moderately retarded, k_O/k_S, being 5 and 7 with the R_P and S_P diasteromers, respectively. The cleavage, in turn, is accelerated: $k_O/k_S(R_P) = 0.1$ and $k_O/k_S(S_P) = 0.3$. In addition, desulfurization takes place under these conditions. The hydrogen sulfide ion is 10^5 times less basic than the hydroxide ion and, hence, able to compete with the sugar oxyanions as a leaving group upon breakdown of the thiophosphorane intermediate (the bond energies of P–O and P–S bonds are 86 kcal mol^{-1} and 55 kcal mol^{-1}, respectively [103]). Although no desulfurization takes place at high pH, this reaction represents 80% of the disappearance of Up(s)U under neutral conditions.

Replacing both of the non-bridging oxygens in a phosphodiester linkage with sulfur does not markedly change the behavior compared to phosphoromonothioates. The thio effect, k_O/k_S, is 2.8 for the hydroxide-ion-catalyzed reaction, 0.2 for the pH-independent cleavage and 8 for the pH-independent isomerization [104].

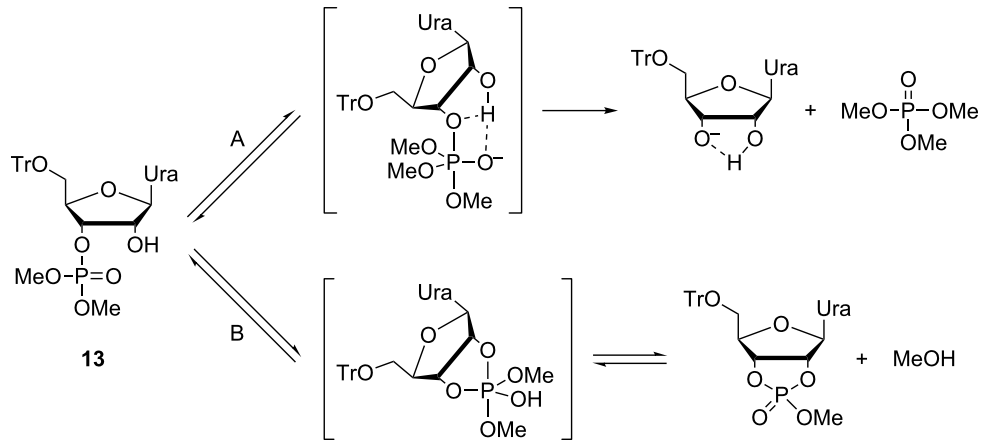
Models for the cleavage by large ribozymes

Transesterification reactions catalyzed by the large ribozymes (group I and II introns, the lariat capping ribozyme, the spliceosome and RNAse P) share a common mechanism that sets them apart from reactions catalyzed by small ribozymes or protein enzymes [42,105]. Perhaps most strikingly, the large ribozymes do not make use of the vicinal 2′-OH as a nucleophile but instead fold into an elaborate tertiary structure that allows an external nucleophile to attack the phosphorus atom of the scissile phosphodiester linkage [106,107]. The leaving group, in turn, is the 3′- rather than the 5′-oxygen. Finally, unlike many small ribozymes, large ribozymes are obligate metalloenzymes, activating the phosphodiester substrate by direct coordination of Mg(II) to the non-bridging oxygens [108-110]. All of these features present unique challenges to the design of relevant model systems.

As discussed above, non-enzymatic cleavage of RNA phosphodiester linkages proceeds exclusively by attack of the vicinal 2′-OH. No other nucleophile, including solvent water or hydroxide ion, is able to compete. The large ribozymes have to provide a solvent-free environment that suppresses the nucleophilic attack of the vicinal 2′-OH by intrachain H-bonding and promotes the attack of an external nucleophile by appropriate preorganization, or the RNA chain is locked to a conformation where intrachain in-line attack is not possible. Several approaches have been developed to simulate these conditions with small molecular models.

The solvent-free environment of the catalytic core of large ribozymes has been mimicked in small molecular model systems by performing the reactions in an organic solvent, rather than water. For example, intermolecular attack on a ribonucleoside 3′-phosphotriester has been observed in methanol and in a mixture of methanol and dichloromethane when methoxide ion at a high concentration was used as the nucleophile (Scheme 11) [111]. A phosphotriester, rather than a phosphodiester, was chosen as a model for better solubility in organic media as well as for higher reactivity. Regarding the overall charge, phosphotriesters can be considered to be mimics of the monoprotonated phosphodiesters.

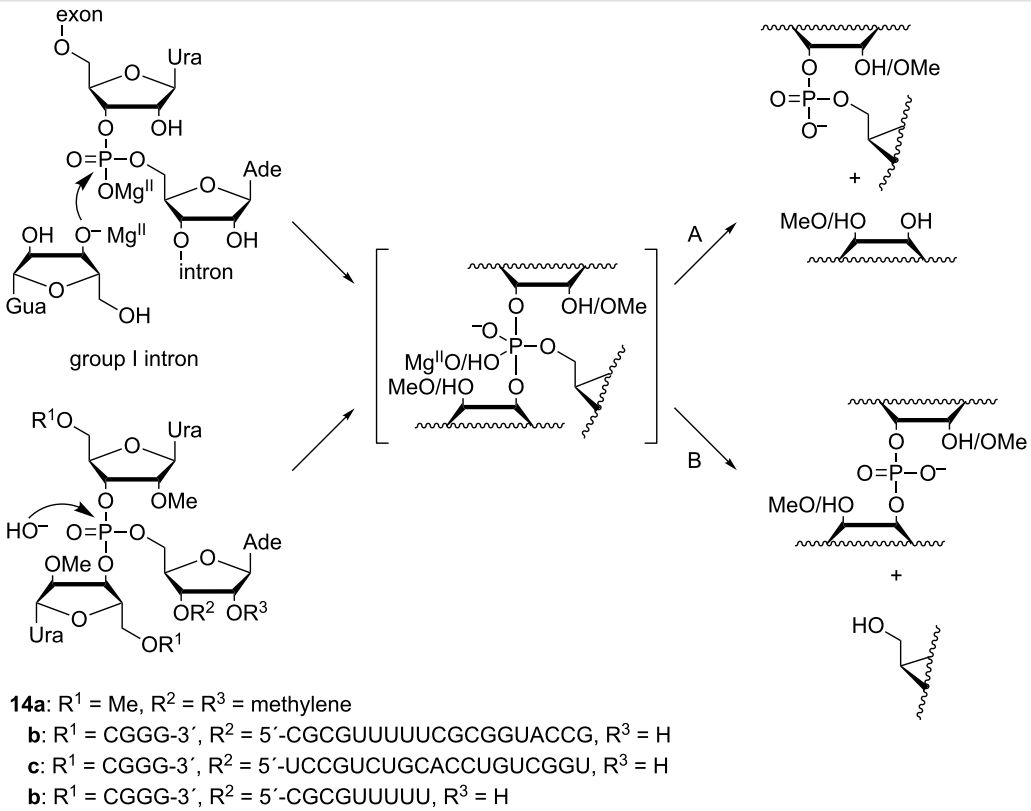
An attack by methoxide (Scheme 11, route A) leads to release of uridine in mixtures of methanol and dichloromethane. The intramolecular attack of 2′-OH undoubtedly is much faster than the intermolecular attack of methoxide (Scheme 1, route B), but the resulting 2′,3′-cyclic triester is reverted back to the starting material by the attack of methoxide, the equilibrium in dry methanol being overwhelmingly on the side of the acyclic triester **13**. In aqueous solution, closely related triesters react exclusively by route B [88,112]. Methanolysis of the arabino and 2′-deoxyribo analogs of **13** was 30-fold slower, underlining the importance of the *cis*-diol system [111]. Apparently, the 2′-OH acts as an electrophilic catalyst which is stabilizing



Scheme 11: Cleavage of uridine 3′-dimethyl phosphate by A) intermolecular attack of methoxide ion and B) intramolecular attack of 2′-OH [111].

the negative charge developing on the phoshorane intermediate and/or the departing 3′-oxygen by H-bonding.

Hydrolysis of phosphotriesters is the reverse reaction of the attack of alcohol on phosphodiesters, the key reaction catalyzed by large ribozymes. These reactions, hence, proceed through the same pentacoordinated phosphorane intermediate or transition state. Accordingly, the impact of various factors, such as intramolecular hydrogen bonding and the secondary structure around the scissile phosphate, can be studied with phosphotriester models. Hydroxide-ion-catalyzed hydrolysis of trinucleoside 3′,3′,5′-monophosphates **14a**–**d**, for example, has been used as a model reaction for transesterification of group I and II introns (Scheme 12) [113,114]. In these models, methylation of



14a: R^1 = Me, R^2 = R^3 = methylene
b: R^1 = CGGG-3′, R^2 = 5′-CGCGUUUUUCGCGGUACCG, R^3 = H
c: R^1 = CGGG-3′, R^2 = 5′-UCCGUCUGCACCUGUCGGU, R^3 = H
b: R^1 = CGGG-3′, R^2 = 5′-CGCGUUUUU, R^3 = H

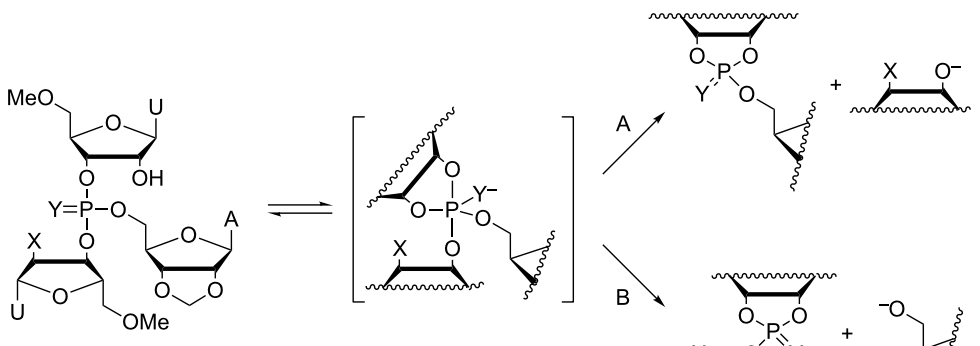
Scheme 12: Transesterification of group I introns and hydrolysis of phosphotriester models proceed through a similar intermediate or transition state that can decompose by A) P–O3′ or B) P–O5′ bond fission.

the 2′-OH group of the two 3′-linked nucleosides was necessary to prevent them from acting as intramolecular nucleophiles.

The pentacoordinated intermediate or transition state obtained by the attack of hydroxide on **14a**–**d** may decompose by cleavage of either P–O3′ (Scheme 12, route A) or P–O5′ bond (route B), yielding a 3′,5′- or a 3′,3′-phosphodiester, respectively. The ribozyme reaction follows exclusively route A [115-117], whereas hydrolysis of the model compounds (**14a**–**d**) proceeds by both routes [113,114]. With the simplest model, compound **14a**, comprising only the three nucleosides directly linked to the scissile phosphate, P–O5′ cleavage (route B) accounts for 15% of hydroxide-ion-catalyzed hydrolysis, independent of the reaction temperature (3–90 °C). The product distribution of the oligonucleotide models, **14b**–**d**, on the other hand, was temperature-dependent, the proportion of P–O5′ cleavage ranging from approximately 3% (at 3 °C) to approximately 20% (at 90 °C). Furthermore, **14b**–**d** reacted approximately 6-fold slower than **14a**. Evidently base stacking specifically retards cleavage of the P–O5′ bond. It is interesting to note that in the catalytic core of group I introns, the scissile phosphodiester linkage is embedded within a double-helical stem [118,119], where base stacking is undoubtedly stronger than in the oligonucleotide models (**14b**–**d**). Unfortunately, studying double-helical model systems was precluded by the strongly denaturing alkaline conditions required for the hydroxide-ion-catalyzed reaction to prevail.

Besides steric constraints of the catalytic core, stabilization of the departing 3′-oxyanion by an H-bond donated by the vicinal 2′-OH group has been proposed as an explanation for the overwhelming predominance of the P–O3′ over the P–O5′ cleavage in the reactions of large ribozymes [120-123]. Rate acceleration by a vicinal hydrogen bond donor in the leaving group has, indeed, been observed in the intramolecular cleavage of ribonucleoside 3′-phosphodiesters [89,124] as well as in the intermolecular methanolysis of ribonucleoside 3′-phosphotriesters discussed above. However, while consistent with stabilization of the leaving group, these results are open to another interpretation, viz. stabilization of the phosphorane intermediate. Hydrolytic reactions of ribonucleoside 3′-phosphotriesters featuring two different leaving groups have been studied to distinguish between these two alternatives [125-128]. Specific acceleration of departure of the leaving group with a vicinal hydrogen bond donor (Scheme 13, route A) would suggest stabilization of the leaving group, whereas equal acceleration of both of the parallel reactions (routes A and B) would be more consistent with stabilization of the common intermediate.

In both the phosphate and the phosphorothioate series, cleavage of the model triesters with a free 2′-OH group in the 3′-linked departing nucleoside **15a** and **16a** was approximately 30-fold faster than the respective reaction of the 2′-O-methylated analogues **15b** and **16b** [125,126,128]. A 2′-trifluoroacetamido group proved somewhat more activating, compound **15c** being hydrolyzed approximately 50-fold faster than **15a** [127]. In the case of **15a** and **16a**, both P–O3′ and P–O5′ cleavage (Scheme 13, routes A and B, respectively) were equally facilitated, suggesting that the 2′-OH donates a hydrogen bond to non-bridging oxygen of the phosphorane intermediate, rather than the departing 3′-oxygen. With **15c**, on the other hand, specific acceleration of P–O3′ cleavage was observed, consistent with hydrogen bond stabilization of the leaving group.



15a: X = OH, Y = O
b: X = OMe, Y = O
c: X = $NHCOCF_3$, Y = O
16a: X = OH, Y = S
b: X = OMe, Y = S

Scheme 13: Cleavage of trinucleoside 3′,3′,5′-monophosphates by A) P–O3′ and B) P–O5′ bond fission.

Isomerization of the internucleosidic phosphodiester linkages is not observed with ribozymes but the respective reaction of model compounds is still useful when making mechanistic interpretations, as it shares a common intermediate with cleavage. With the model triesters **15a**–**c** and **16a**,**b**, isomerization becomes hydroxide-ion-catalyzed already at pH 2 and is much faster than cleavage under neutral and alkaline conditions. Isomerization of the phosphate models was too fast to be measured but with the phosphorothioate models, comparision of the rates of hydroxide-ion-catalyzed isomerization of **16a** and **16b** was possible [128]. Interestingly, **16a** was isomerized an order of magnitude faster than **16b**, offering perhaps the most compelling piece of evidence for hydrogen bond stabilization of the phosphorane (or thiophosphorane) intermediate.

Steric constraints imposed by the tertiary structure of the large ribozymes undoubtedly have a profound effect on the course of the ribozyme-catalyzed reactions and such effects are notoriously difficult to duplicate in small molecular models. For example, the apparent discrepancy between the results obtained with simple triester models and modified ribozymes on the effect of the 2′-OH of the departing 3′-linked nucleoside could be explained in terms of an intricate hydrogen bonding network at the catalytic core of the large ribozymes [120,129-131]. On the other hand, even the simple expansion of a trinucleoside phosphotriester (such as **15b**) with short homothymine oligonucleotide arms stabilized the phosphotriester core toward hydroxide-ion-catalyzed cleavage by an order of magnitude and completely suppressed P–O5′ cleavage [132]. Even higher stabilizations were observed with more elaborate phosphate-branched oligonucleotide models [133] but the data did not allow unambiguous correlation of structure and stability. Clearly, as the model systems start to approach the large ribozymes in complexity, the results may become more relevant but at the same time more difficult to interpret.

Cleavage of DNA by Brönsted acids and bases

The sugar-phosphate backbone of DNA is known to be extremely stable at pH 7 and 25 °C. In fact, no reliable estimate for the half-life of the cleavage of an individual 3′,5′-phosphodiester linkage is available. The estimate for the fission of a P–O bond, based on hydrolysis of dineopentyl phosphate, is $7{\cdot}10^{-16}$ s^{-1}, corresponding to a half-life of 31 million years [2]. Most likely, the cleavage of the C5′–O bond in DNA is somewhat faster. For comparison, 99% of the hydrolysis of dimethyl phosphate proceeds under neutral conditions by nucleophilic attack on carbon leading to C–O bond cleavage [134]. Since C5′ is relatively open for a nucleophilic attack, C–O bond cleavage may take place with DNA phosphodiester linkages. In addition, depurination and various base moiety modifications may well lead to sugar ring opening that allow chain cleavage by elimination [135].

The hydrolysis of dineopentyl phosphate, taken as a model of P–O bond cleavage in DNA, is pH independent over a wide pH range from pH 7 to 12 [2], in striking contrast to cleavage of RNA which turns hydroxide-ion-catalyzed already at pH 5 [44]. Either, water attacks on the phosphorus atom of the dineopentyl phosphate monoanion, possibly by concerted proton transfer to one of the non-bridging oxygens, or hydroxide ion attacks neutral dineopentyl phosphate. In both cases the reaction takes place through a monoanionic pentacoordinated species, which may have a finite life-time. Computational calculations have provided considerable evidence for the former of these mechanistic alternatives [136].

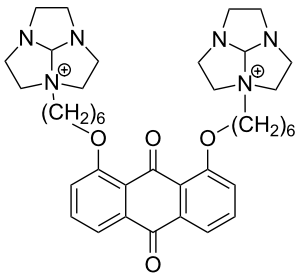
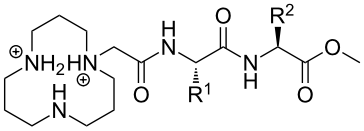
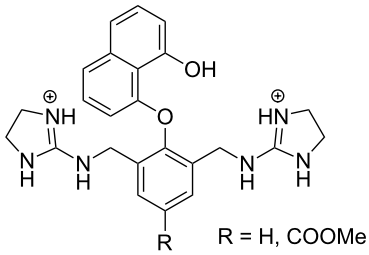
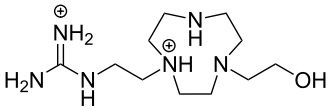
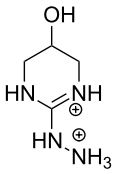
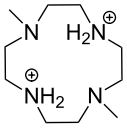
Owing to the extremely high stability of DNA phosphodiester linkages at physiological pH, no mechanistic studies with dimeric DNA fragments have been carried out. Instead, plasmic supercoiled DNA consisting of thousands of base pairs is usually used as a target on developing various cleaving agents. Cleavage of even one phosphodiester linkage may lead to electrophoretically detectable relaxation of the supercoiled structure (Form I), first to a circular DNA (Form II) by bond cleavage within one of the chains and then to a linear form (Form III) by cleavage of both strands. Table 1 depicts structures of nonmetallic agents shown to cleave supercoiled DNA at physiological pH in aqueous solution by a hydrolytic mechanism. Cleavage by a radical mechanism has usually been excluded by showing that radical scavengers do not retard the reaction or by showing that the linearized (Form III) plasmid is a substrate of ligases. Otherwise the mechanistic information is scanty. The common feature of the cleaving agents is a dicationic structure. In addition, the agent may contain an aromatic moiety that enhances intercalation (**18**, **20**) or a hydroxy function that can serve as an intracomplex nucleophile (**20**–**22**). With the latter compounds, the guanidinium type structure has been assumed to interact with the non-bridging phosphoryl oxygens and, hence, facilitate the attack of the covalently attached hydroxy function.

Metal-ion-promoted cleavage of nucleic acids

General

Many metal ions and their complexes enhance the cleavage of phosphodiester bonds. In some cases the process is catalytic and the metal ion catalyst converts an excess of substrate into products. True catalysis with multiple turnover is generally observed with bis(*p*-nitrophenyl) phosphate (BNPP, **23a**, Figure 10) [143,144], a widely used simple model compound mimicking DNA phosphodiester bonds, and sometimes with HPNP (**1**) [145]. Usually, though, it is not the case, as the prod-

Table 1: Cleavage of supercoiled DNA by nonmetallic cleaving agents.

compound	structure of the cleaving agent	efficiency of cleavage	ref.
17		plasmid pBR322 conversion to Form III was detected upon 2 h incubation with **17** (200 mmol L^{-1}) in tris buffer at pH 7.2 and 37 °C.	[137]
18	 R^1 = H, iBu R^2 = H, Me, CH_2OH, CH_2Ph, $CH_2C_6H_4OH$	half-life for the cleavage of plasmid pUC19 to Form II reported to be 3.3 h at physiological pH.	[138]
19	 R = H, COOMe	50–60% of plasmid was converted to Form II upon 48 h incubation with **19** (200 mmol L^{-1}) in HEPES buffer at pH 7.2 and 37 °C.	[139]
20		37% of plasmid pUC19 was converted to Form II upon 20 h incubation with 10 mmol L^{-1} **20** in HEPES buffer at pH 7.0 and 37 °C.	[140]
21		half-life for the conversion of plasmid pUC 19 to Form II 4.3 h (tris buffer pH 7.2) at saturating concentrations of **21**.	[141]
22		half-life for the conversion of plasmid pUC 19 to Form II reported to be 18 h (Tris buffer pH 6.0, 37 °C) at saturating concentrations of **22**.	[142]

ucts bind to the catalyst much more strongly than the starting material. The catalyst is consumed, and the process is, strictly speaking, not catalytic. These terms are, however, used throughout the review along with more correct expressions to promote and to enhance. The rate-enhancement by metal aqua ions on the hydrolysis of DNA models and transesterification of RNA models generally is rather modest, as is shown by the chosen representative examples in Table 2. Among divalent metal ions, Zn^{2+} and Cu^{2+} are usually the most efficient ones. Alkaline and alkaline earth metal cations show only a slight rate-enhancement, whereas trivalent lanthanide ions are generally more efficient catalysts than divalent metal ions [146-148].

In addition to the rather modest rate enhancement, studies with metal aqua ions are limited by precipitation of catalysts as hydroxides, in some cases even at neutral pH [157]. While in the case of divalent metal ions the formation of an insoluble hydroxide decreases catalytic activity, lanthanide aqua ions form gel-like material of unknown structure that is catalytically more active than aqua ions [148,158]. The reaction order in

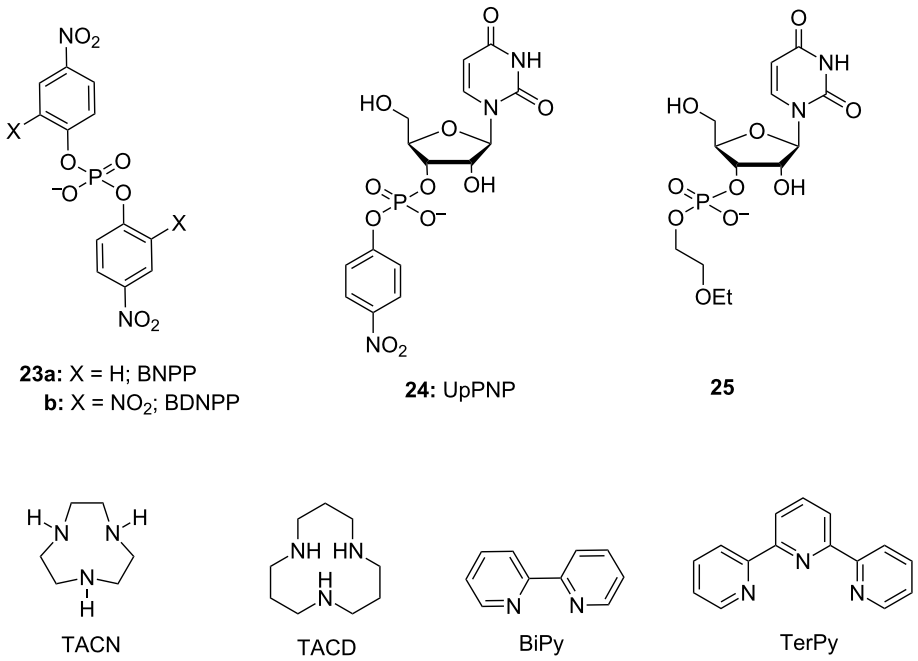


Figure 10: Model compounds (**23**–**25**) and metal ion binding ligands used in kinetic studies of metal-ion-promoted cleavage of nucleic acids.

Table 2: Catalytic activity ($k_{rel} = k_{obs}/k_{uncat}$) of chosen metal ions and their complexes under given conditions ([catalyst], pH and temperature).

catalyst	k_{rel} BNPP (**23a**)	k_{rel} HPNP (**1**)	k_{rel} UpPNP (**24**)	k_{rel} NpN (UpU or **25**)[a]
Zn^{2+}(aq)		150; 0.5 mmol L^{-1}, pH 7.00, 37 °C[b]	33; 10 mmol L^{-1}, pH 5.9, 25 °C[c]	32; UpU, 1 mmol L^{-1}, pH 7.00, 80 °C[b]
Cu^{2+}(aq)	27; 0.1 mmol L^{-1}, pH 6.50, 75 °C[d]			
Eu^{3+}(aq)		7700; 0.5 mmol L^{-1}, pH 7.00, 37 °C[b]		475; UpU, 1 mmol L^{-1}, pH 7.00, 80 °C[b]
Cu^{2+}-TerPy	ND[e]	52; 2 mmol L^{-1}, pH 7.0, 25 °C[f]	179; 10 mmol L^{-1}, pH 6.6, 25 °C[c]	2164, UpU, 10 mmol L^{-1}, pH 6.6, 90 °C[c]
Cu^{2+}-BiPy	2000; 1 mmol L^{-1}, pH 6.50, 75 °C[d]	144; 2 mmol L^{-1}, pH 7.0, 25 °C[g]	116; 10 mmol L^{-1}, pH 6.6, 25 °C[c]	291, UpU, 10 mmol L^{-1}, pH 6.6, 90 °C[c]
Cu^{2+}-TACN	5700; 2 mmol L^{-1}, pH 7.0, 50 °C[h]	298; 2 mmol L^{-1}, pH 7.0, 25 °C[h]		
Zn^{2+}-TACD	10000; 10 mmol L^{-1}, pH 8.5, 35 °C[i]	450; 0.20 mmol L^{-1}, pH 7.0, 25 °C, 50% MeCN[j]	58; 10 mmol L^{-1}, pH 5.9, 25 °C[c]	410; **25**, 2 mmol L^{-1}, pH 6.6[c]

[a]The pK_a of the leaving group alcohol in **25** is the same as in dinucleoside monophosphates; [b]from ref. [146] ; [c]from ref. [149]; [d]from ref. [150]; [e]no catalysis has been observed as discussed in ref. [151]; [f]from ref. [152]; [g]from ref. [153]; [h]from ref. [154]; [i]from ref. [155]; [j]from ref. [156].

lanthanide and hydroxide ion concentration approaches three when reaching the pH where precipitation starts. Furthermore, the remarkably large rate enhancement is observed only when the gel is being formed during the course of the phosphoester cleavage.

The solubility problem can be, to some extent, overcome by the use of sufficiently stable metal ion complexes. The ligand affects the catalytic activity of metal ion and many Zn^{2+} and Cu^{2+} complexes are more efficient as catalysts than the corresponding aqua ions (Table 2). Zn^{2+} complexes of polyaza-

macrocycles such as 1,5,9-triazacyclododecane (TACD), 1,4,7-triazacyclononane (TACN), and their derivatives [159,160], as well as Cu^{2+} complexes of terpyridine (TerPy), bipyridine (BiPy) and their derivatives, are among the most frequently studied species. In the case of lanthanide ions, the situation is opposite. Complex formation decreases the observed catalytic activity, at least partly due to blocked gel formation. Furthermore, lanthanide complexes with neutral ligands tend to be unstable and ligands with side arms that encapsulate the lanthanide ions are required [161,162]. Ligands with negatively charged side arms form the most stable complexes, but a negative charge generally decreases the catalytic activity. In addition to improved solubility, a ligand may enable ligation of the metal complex to various structures. This is necessary in a number of applications, which are outside the scope of the present review.

As suggested by Breslow [163] and Chin [164] already in early 1990's, a second metal ion [165-167] or a hydrogen bond forming substituent [168-171] can markedly enhance the catalytic activity. As an example, **26a** is a 79 times more efficient catalyst for HPNP cleavage than **26b** devoid of amino groups [168] and the rate-accelerating effect of the second metal ion center in **27b** is even more prominent when compared to **28d** [167]. A similar effect has been observed on using BNPP as a substrate: **28a** promotes the hydrolysis of BNPP 230 times as efficiently as **28b** [172] and k_{cat}/k_0 values reported for hydrolysis promoted by **29a** and **29b** are 640 and 250 times higher than that for the unsubstituted complex **29c** [173]. The higher cleaving activity partially results from stronger interactions with the substrate, but also from enhanced catalytic efficiency [173]. The importance of the factors may vary depending on the structure [143,167]. As an example, the observed rate enhancement by the bimetallic complex **27b** and the mononuclear **28c** are equal, but inhibition studies by an unreactive substrate analog shows that while **27b** binds more strongly, **28c**, when bound, is more efficient as a catalyst (Figure 11) [167].

The most intensively studied bimetallic catalysts for the cleavage of RNA models are **30** (Figure 12) and **27a** introduced by Morrow [166] and Williams [145], respectively. Complex **30** at 2 mmol L^{-1} concentration reduces the half-life of the cleavage of UpU to about one week at pH 7.0 and 25 °C [174] and **27a** is even more efficient: the half-life of UpU cleavage is only seven hours in the presence of 1 mmol L^{-1} **27a** at pH 6.5 and 25 °C [175]. **27a** and its Co^{2+} analog are unique among metal ion catalysts in that they modestly enhance also the interconversion of 3′,5′- and 2′,5′-dinucleoside monophosphates [175,176]. Catalysis on the hydrolysis of DNA models by these complexes has not been studied or is less significant than in the case of RNA models. Interestingly, very fast cleavage of highly activated DNA analog, bis(2,4-dinitrophenyl phosphate) (BDNPP; **23b**), has been observed in the presence of Tb^{3+}, Eu^{3+} and Gd^{3+} complexes of ligand **31** in water/acetonitrile mixtures. Half-life less than 1 second has been reported for Eu^{3+}-**31** at 1 mmol L^{-1} concentration at pH 7.0 and 25 °C [144]. The rate-enhancement compared to the background reaction is approximately 10^6-fold. Larger non-enzymatic rate-enhancing effects have been obtained only in anhydrous methanol and ethanol with HPNP and its analog as substrates [177]. Kinetic data obtained with bifunctional catalysts is collected in Table 3.

Even though many metal ion catalysts promote the cleavage of phosphodiester bonds, **27a** is the only catalyst that is known to enhance the mutual 3′,5′- to 2′,5′ isomerization of RNA phos-

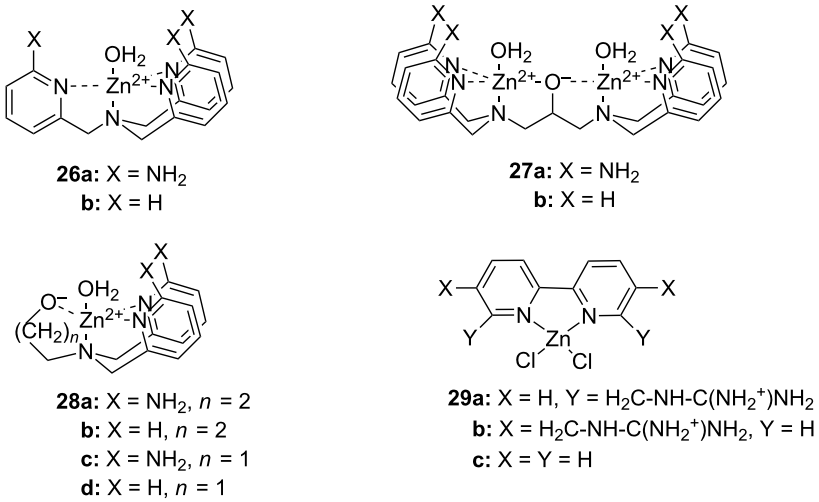


Figure 11: Zn^{2+}-ion-based mono- and di-nuclear cleaving agents of nucleic acids.

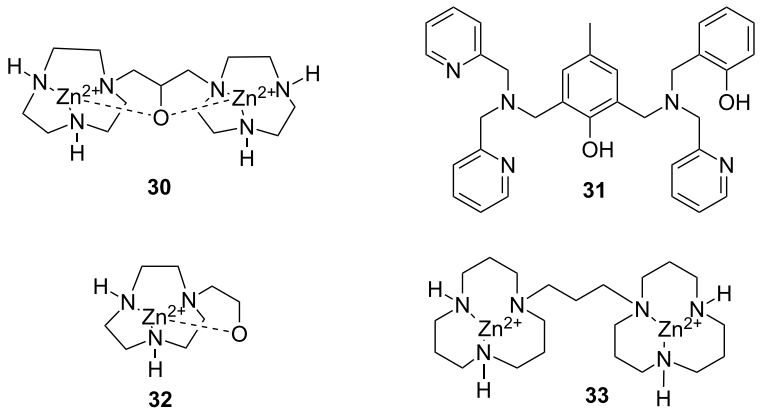


Figure 12: Miscellaneous complexes and ligands used in kinetic studies of metal-ion-promoted cleavage of nucleic acids.

Table 3: Pseudo first-order rate constants (s^{-1}) of phosphoester hydrolysis/transesterification in the presence of bimetallic and monometallic complexes (1 mmol L^{-1}) under neutral conditions.

Catalyst	HPNP	NpOAr	NpN
32	$1.3 \cdot 10^{-6}$ (pH 7,61)[a]		
30	$2.5 \cdot 10^{-4}$ (pH 7,61)[a]	0.1 (pH 7.5)[b] NPP	$9.9 \cdot 10^{-7}$ (pH 7.4)[c]
28c	$4.6 \cdot 10^{-5}$ (pH 7.4)[d]		
27a	$5.3 \cdot 10^{-2}$ (pH 7,4)[e]		$2.6 \cdot 10^{-5}$ (pH 6.5)[f]

[a]From ref. [178]. Calculated from the second-order rate constant determined as the slope of k_{obs} vs *c*(complex) plot. [b]From ref. [179]. Calculated from the second-order rate constant estimated from Figure 1. [c]From ref. [174]. Calculated from the second-order rate constant determined as $k_2 = k_{cat}/K_m$. [d]From ref. [167]. Calculated from the second-order rate constant determined as the slope of k_{obs} vs c(complex) plot. [e]From ref. [168]. Calculated from the second-order rate constant determined as $k_2 = k_{cat}/K_m$. Second-order rate constants determined as the slope of k_{obs} vs c(complex) plot. [f]Observed pseudo first-order rate constants from ref. [175].

phodiester bonds [175,176]. As discussed in the foregoing, isomerization is the predominant reaction of dinucleoside monophosphates and related nucleoside 3′-alkyl phosphates with a poor leaving group in the absence of metal ion catalysts at pH < 7, whereas activated phosphodiesters are not isomerized. There are two obvious reasons for the lack of isomerization in the presence of metal ion catalysts. Firstly, when the phosphorane intermediate obtained is dianionic, it is too unstable to pseudorotate. Evidently metal ion binding does not sufficiently stabilize the intermediate, or it retards pseudorotation. Alternatively, the departure of the leaving group by the exocyclic fission may be so efficiently enhanced that isomerization via the endocyclic cleavage cannot compete with it. The first step of the reaction may become rate-limiting or the reaction becomes a concerted process.

The catalysis of phosphate migration by **27a** is modest in comparison to the cleavage reaction. At a concentration of 1 mmol L^{-1} **27a** promotes the isomerization of UpU by a factor of 150, while the cleavage is accelerated up to 10^6-fold [175,176]. Studies with a non-cleavable phosphonate analog have, however, verified the rate-acceleration of isomerization. Evidently, **27a** and its Co^{2+} and Cu^{2+} analogs stabilize the phosphorane to such an extent that pseudorotation can take place, probably through multiple interactions between the catalyst and the phosphorane. Consistent with this assumption, thiophilic Zn^{2+} accelerates the isomerization of phosphoromonothioate analog of UpU, although again the acceleration of isomerization is modest compared to the acceleration of cleavage, at $[Zn^{2+}]$ = 5 mM 6.4- and 410-fold, respectively [180].

Parameters describing the catalytic activity

The rate enhancing effects of metal ion catalysts can be described in several different ways that may give a different impression on the catalytic power of a given complex. A straightforward way to describe the efficiency of a metal ion catalyst is to give the ratio of pseudo first-order rate constants obtained in the presence and in the absence of the catalyst, as done in Table 2. Problems may, however, arise when the background reaction is slow. Rate constants under neutral conditions often have to be estimated by linear extrapolation from the rate constants measured under alkaline conditions without knowing whether the logarithmic rate constant really is linearly related to pH over the wide pH range employed. One should

bear in mind that the shape of the pH-rate profile depends on polar nature of the leaving group [48,181]. Likewise, comparison between rate constants determined at different pH and catalyst concentration may easily lead to errors, if experimental data on dependence of rate on catalyst concentration at various pH values is not available, which very often is the case. In summary, care should be exercised on comparing the catalytic efficiencies of various catalysts.

Michaelis–Menten kinetics (Equation 1) has often been applied to metal-ion-catalyzed cleavage, particularly the cleavage of HPNP [143,145,182,183]. Parameters K_m (in mol L^{-1}) and k_{cat} (in s^{-1}) are the dissociation constants of the catalyst-substrate complex and the first-order rate constants for the breakdown of the catalyst-substrate complex to products. $[S]_0$ and $[catalyst]_0$ stand for the initial concentrations of the substrate and catalyst. The ratio k_{cat}/K_m, hence, is the measure of catalytic efficiency. This ratio actually is equal to the second order rate constant for the metal ion catalytic reaction, i.e., the slope of k_{obs} vs [catalyst] plot.

$$\text{Initial rate} = k_{cat}[\text{catalyst}]_0[S]_0 / \left(K_m + [S]_0\right) \quad (1)$$

The ratio of k_{cat}/k_0, where k_0 is the first-order rate constant for the uncatalyzed reaction, is sometimes used to describe the efficiency of a given catalyst. Values thus obtained are impressive, but may give an unrealistic impression, as k_{cat} refers to situation where all the substrate molecules are quantitatively bound to the catalyst; a situation that is rarely achieved. Comparison of k_{cat}/K_m values shown in Table 4 puts the catalytic activity of even the most efficient metal ion catalysts into perspective. It can be seen that while the rate-enhancement obtained by bimetallic complexes is fairly impressive, it still falls far behind the catalytic activity of enzymes. Although the K_m term referring to the substrate binding is of the same order, the k_{cat} are several orders of magnitude smaller than those for enzyme catalysis. Sometimes catalytic activity is expressed as kinetic effective molarity that is defined as the ratio between the first-order rate constant of an intracomplex reaction and the second-order rate constant of the corresponding intermolecular reaction.

As mentioned above, catalytic efficiency may be expressed by k_{cat}/K_m. Accordingly, it is of interest to understand to what extent each of these parameters contribute to the observed catalytic effect of various metal-based catalysts. Metal aqua ions and simple metal ion complexes generally bind monoanionic phosphodiesters only weakly. A frequently applied method to estimate the K_m value is inhibition of the cleavage with an unreactive structural analog of the substrate that binds to the metal ion catalyst approximately as tightly at the substrate [178]. Usually, HPNP is used as the substrate and dimethyl or diethyl phosphate as the inhibitor. The K_i values, dissociation constants of the catalyst–inhibitor complex, are then assumed to correlate with the K_m values. According to these studies, bifunctional catalysts generally bind to the inhibitor more strongly than their monomeric counterparts. Complexes **26a**,**b**, **27a**,**b** and **28c** offer an illustrative example of the stabilizing effect of increasing number of functional groups. The monomeric Zn^{2+} complex **26b** binds considerably less readily, K_i = 0.13 mol L^{-1}, than its amino substituted analog **26a**, K_i = 0.01 mol L^{-1} [168]. Monomeric complex **28c** binds surprisingly weakly (K_i = 0.15 mol L^{-1}), but the corresponding dimer, **27b**, binds much more tightly (K_i = 0.009 mol L^{-1}) [167]. Additional amino groups still increase the affinity; the K_i value for **27a** is 0.32 mmol L^{-1} [145]. Likewise, the dinuclear Zn^{2+} complex of **34** (Figure 13) binds more tightly than the mononuclear Zn^{2+} complex of **35**, the K_m values being 0.007 mol L^{-1} and 0.0184 mol L^{-1}, respectively [183]. One should, however, bear in mind that the structure of substrate may also play a role. For instance, dependence of the cleavage rate of BDNPP (**23b**) and HPNP (**1**) on concentration of **36** suggests that binding to BDNPP is weaker than binding to HPNP [182].

Nucleic acid bases offer additional potential coordination sites for metal ion complexes, resulting in tighter substrate binding.

Table 4: Kinetic parameters for the catalysis of the HPNP cleavage by bimetallic complexes. Experimental details are described in the text.

catalyst	substrate	k_{cat} / s^{-1}	K_m / mol L^{-1}	[k_{cat}/K_m] / L mol^{-1} s^{-1} (= k_2)
27a[a]	HPNP	0.017	$3.2\cdot10^{-3}$	53
30[b]	HPNP	$4.1\cdot10^{-3}$	0.016	0.25
33 + 1 equiv MeO^- in MeOH[c]	HPNP			$2.75\cdot10^5$
33 + 1 equiv MeO^- in MeOH[c]	BNPP	0.041	$0.37\cdot10^{-3}$	111
Tb^{3+}-**31**[d]	BDNPP	18	0.006[e]	3000
RNase A[f]	HPNP	$7.9\cdot10^2$	$7.9\cdot10^{-3}$	$1.0\cdot10^5$

[a]From ref. [145]; [b]from ref. [178]; [c]from ref. [177]; [d]from ref. [144]. Data refer to 75% MeCN in water; [e]given as K_1 = 166 mol^{-1} L (= $1/K_m$); [f]ref. [184].

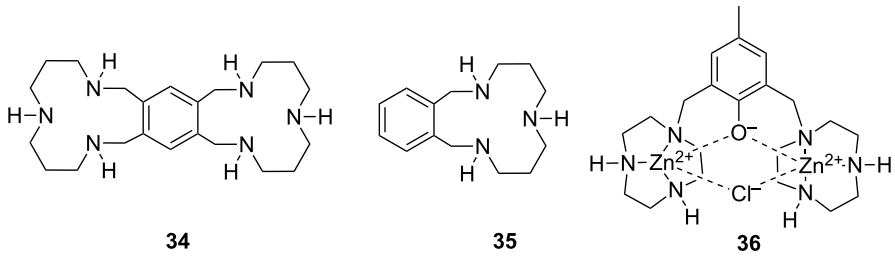


Figure 13: Azacrown ligands **34** and **35** and dinuclear Zn^{2+} complex **36** used in kinetic studies of metal-ion-promoted cleavage of nucleic acids.

Uracil and guanine bases, in particular, are potential coordination sites as they undergo deprotonation around pH 9. Interaction of **30** with uracil bases has been suggested to be fairly strong [174]. According to kinetic inhibition studies, UpU is recognized almost one order of magnitude more efficiently than HPNP. In addition, uridine has been shown to inhibit the cleavage of HPNP promoted by Zn^{2+}-polyazamacrocycle complexes [185].

pH-Rate profiles

Determination of pH-rate profile is very often the first experiment employed to study the mechanism of a reaction. Plots of k_{obs} (or $k_2 = k_{obs}$/[catalyst]) against pH are generally sigmoidal [151,159,182] or bell-shaped [162,169,172,186,187] for metal-ion-promoted reactions, independently of the type of substrate. Sigmoidal profile has been attributed to a catalyst with one dissociable functional group, whereas a bell-shaped profile has been taken as an indication of two such groups [160]. pK_a values determined on the basis of pH-rate profiles usually agree well with the values obtained potentiometrically for the catalyst complexes [159,168]. These results are often interpreted as an indication of the mono-deprotonated complex being the active catalyst and a metal-bound hydroxy or alkoxy group being involved in the reaction. Consistent with this, metal complexes with lowest pK_a values are usually the most efficient catalysts at a fixed pH [148,188].

The descending part of a bell-shaped pH-rate profile has been taken as an indication of a second deprotonation that renders the catalyst inactive. Most logical explanation for the inactivation is release of the substrate: the hydroxide ion and the substrate compete for the metal ion and at sufficiently high concentration of hydroxide ions the binding starts to weaken [162,168]. With a multifunctional catalyst, the decreasing catalytic activity may also result from deprotonation of a functional group directly involved in the catalysis. A third factor, rarely considered in this context, is decreasing stability of the catalyst complex. Formation of precipitates is sometimes observed at higher pH's [166,183], but inactivation of the catalyst may take place already before visible precipitation. Reaction time is also crucial; complexes that are efficient catalysts in reactions of HPNP over a wide pH-range may become inactivated on a time scale required to follow reactions of non-activated substrates.

Another fact that complicates the mechanistic interpretations on the basis of pH-rate profiles is that the background reaction usually is base-catalyzed. Even though the observed first-order or second-order rate constants increase upon increasing pH, the catalytic activity of metal ion complexes may actually decrease. This is clearly seen with the pH-rate profile reported for HPNP cleavage promoted by **27b** [167]. In addition, when quantitative data on the pH-dependence of binding equilibrium is not available, the concentration of catalyst–substrate complex at a given pH is not known and, hence, the reaction system is not accurately defined. Despite the shortcomings discussed above, it is clear that deprotonation at pH close to pK_a of a metal bound aqua ligand plays a significant role in catalysis and it often serves as the basis of mechanistic conclusions.

Effect of substrate structure; β_{lg} values

The results in Table 2 show that the rate-enhancement observed for three RNA models, viz. HPNP, nucleoside 3′-(*p*-nitrophenyl phosphate) and dinucleoside-3′,5′-monophosphate, are within the same magnitude, the largest values being more often obtained with HPNP. There is one clear exception: with Cu^{2+}-Terpy the largest rate-enhancement is obtained with a dinucleoside-3′,5′-monophosphate or nucleoside 3′-alkyl phosphate with an equally poor leaving group. This may possibly be attributed to dimerization of Cu^{2+}-TerPy under the experimental conditions; different substrates seem to respond differently in dimer formation [153].

Despite the apparent similarity of the overall influences, differences in the behavior between alkyl and aryl esters are accounted when the susceptibility to the polar nature of the leaving group is considered [189]. β_{lg} values collected in Table 5 show that there are differences between different types of catalysts (Figure 14) as well as between substrates. Values obtained with nucleoside alkyl esters are generally modestly negative on using metal aqua ions as a catalyst

Table 5: β_{lg} values for cleavage reactions of phosphodiesters promoted by metal ion catalysts.

substrate	catalyst / conditions	β_{lg}	ref.
MePAr	**37**	−1.38 ± 0.01	[193]
MePAr	**38**	−1.2 ± 0.1	[187]
3′-UMP aryl esters	10 mmol L^{-1} $Zn(NO_3)_2$, pH 5.9, 25 °C	−0.9 ± 0.2	[190]
3′-UMP aryl esters	10 mmol L^{-1} Zn-TACD, pH 7.5, 25 °C	−0.81 ± 0.07	[189]
3′-UMP alkyl esters	10 mmol L^{-1} $Zn(NO_3)_2$, pH 5.6, 90 °C	−0.32 ± 0.04	[190]
3′-UMP alkyl esters	2 mmol L^{-1} $ZnCl_2$, pH 5.6, 90 °C	−0.36 ± 0.02	[176]
3′-UMP alkyl esters	10 mmol L^{-1} $NiNO_3$, pH 5.6, 90 °C	−0.54 ± 0.03	[191]
3′-UMP alkyl esters	10 mmol L^{-1} Zn-TACD, pH 6.6, 90 °C	−0.6 ± 0.1	[191]
3′-UMP alkyl esters	2 mmol L^{-1} Zn-TACN, pH 6.6, 90 °C	−0.51 ± 0.04	[191]
3′-UMP alkyl esters	2 mmol L^{-1} Zn-cyclen, pH 6.6, 90 °C	−0.71 ± 0.06	[191]
3′-UMP alkyl esters	2 mmol L^{-1} Ni-TACD, pH 6.6, 90 °C	−0.58 ± 0.04	[191]
3′-UMP alkyl esters	1 mmol L^{-1} **27a**	−0.92 ± 0.07	[176]

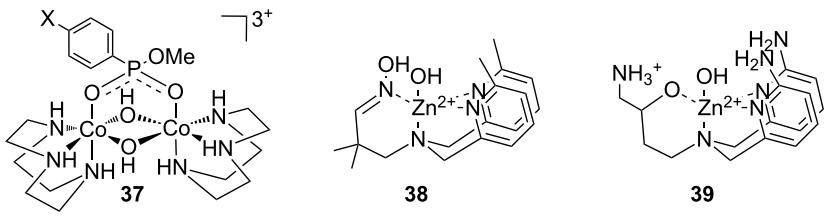


Figure 14: Metal ion complexes used for determination of β_{lg} values of metal-ion-promoted cleavage of RNA model compounds.

[149,176,189,190]. In this respect, the reaction resembles acid-catalyzed transesterification of nucleoside phosphodiesters [49], and the similarity has been taken as an indication of protonation of the leaving group in the rate-limiting step [190]. β_{lg} values obtained with Ni^{2+} or metal ion complexes are slightly more negative than that obtained with Zn^{2+}, but they still are clearly less negative than the value reported for the alkaline cleavage, viz. −1.28 at 90 °C [49]. The values evidently reflect varying degree of protonation that, in turn, depends on the acidity of aqua ligand of the complex and the coordination geometry around the metal cation. The fairly negative value of −0.92 obtained in the presence of 1 mmol L^{-1} **27a** has been compared [176] to the value, −0.94, reported for the pH-independent reaction of nucleoside 3′-(dialkyl phosphate)s [50]. In the latter reaction the leaving group departs as alcohol with concerted proton transfer from a general acid.

The β_{lg} values of the cleavage of aryl esters are more negative than those obtained with nucleoside alkyl esters [149,176,189,191], typically around −0.9. They are also more negative than the values obtained in the absence of metal ion catalysts, −0.58 [192] and −0.54 [51] for the hydroxide-ion-catalyzed cleavage of alkyl aryl phosphates and nucleoside aryl phosphates, respectively. In the case of the Co^{3+}-complex-promoted cleavage of alkyl aryl phosphates, the markedly negative β_{lg} has been attributed to significant bond strain, resulting from a formation of a four-membered ring upon nucleophilic attack of the bridging hydroxo ligand on phosphorus [192]. As regards nucleoside aryl phosphates, the most logical explanation is that metal ion binding stabilize the phosphorane intermediate and, hence, shifts the transition state towards the products obtained by departure of aryloxy anions. In other words, the concerted mechanism with rate-limiting formation of the phosphorane that operates in the absence of a metal ion catalyst is altered towards a stepwise mechanism. In summary, with nucleoside aryl phosphates, the metal-ion-promoted cleavage is more sensitive than the background reaction to the electronegativity of the leaving group (−0.9 vs −0.5), whereas with alkyl phosphates the situation is the opposite (−0.5 vs −1.3). This essentially means that the rate-enhancing effect of metal ions, when expressed as k_{cat}/k_0, increases when an aryl leaving group becomes better or an alkyl leaving group becomes poorer [189].

The use of dinucleoside-3′,5′-monophosphates as model compounds brings about an additional feature not present in simpler model compounds; two nucleic acid bases provide additional binding sites for catalysts. Catalysis by monometallic species is fairly insensitive to the base composition: rate constants of 15

different dinucleoside monophosphates differed within a factor of two in the presence of 10 mmol L^{-1} Zn^{2+} at pH 5.1 and 90 °C [194]. In contrast, catalysis by Cu^{2+}-TerPy is markedly base moiety selective: among four dinucleoside monophosphates studied, an 8-fold difference was observed between the most (ApA) and least (UpU) reactive substrates [165]. With more complex catalysts, the differences can be even larger: a 500-fold reactivity difference has been reported for a trinuclear calix[4]arene-based Cu^{2+} catalyst, UpU and CpA being the most and least reactive, respectively [155]. Bifunctionalized calix[4]arene bearing Cu^{2+}-TACN and a guanidinium group also show marked selectivity. GpA is 130 times more reactive than CpA [171]. A dimeric catalyst with two Cu^{2+}-TerPy units favors, in turn, ApA as the substrate [165]. In contrast to these results, rate-enhancement by **27a** is fairly insensitive to base composition: among five different 3,5-dinucleoside monophosphates studied, only a 3.5-fold difference was observed [176]. Preferred binding of Zn^{2+} azacrown chelates to uracil has been exploited in developing di- and trinuclear base moiety selective cleaving agents for RNA [195,196].

Heavy atom and solvent isotope effects

Heavy atom isotope effects lend further support for the view that the transition state of metal-ion-promoted cleavage of RNA is late compared to the hydroxide-ion-catalyzed cleavage (Table 6). While the $^{18}k_{lg}$ value for specific base-catalyzed cleavage of UpG is 1.0343, the same isotope effect for the Zn^{2+}-promoted reaction is 1.015, still normal but considerably smaller and, hence, consistent with more rigid bonding to the leaving group [197]. The ^{18}O isotope effect for the attacking nucleophile is inverse for the metal-ion-catalyzed reaction, $^{18}k_{nuc}$ = 0.986. The values are consistent with a late transition state, with significant bond formation between the nucleophile and phosphorous [197]. When dinuclear Zn^{2+} complex **30** is used as a catalyst and HPNP as a substrate $^{18}k_{lg}$ = 1.0113 and $^{18}k_{nuc}$ = 0.9874 [198]. The values closely resemble those obtained with UpG and differ more markedly from those of the hydroxide-ion-catalyzed cleavage of HPNP. Accordingly, Zn^{2+}-promoted cleavage of both UpG and HPNP appears to proceed via a similar late transition state, whereas mechanisms of the hydroxide-ion-catalyzed reactions are different: HPNP is cleaved by rate limiting formation and UpG by rate limiting breakdown of the phosphorane intermediate.

The secondary ^{15}N isotope effect (^{15}k) for the nitro group of *p*-nitrophenol leaving group is particularly useful, for it can be regarded as a measure of the charge development on the leaving group oxygen. The value of 1.0013 observed for the Cu^{2+}-TACN-promoted reaction of ethyl *p*-nitrophenyl phosphate (EtPNP) has been attributed to 46% bond cleavage in the transition state [200]. A value of the same magnitude has been observed for the transesterification of HPNP-promoted by **30** [194]. The value of 1.0002 for the specific base-catalyzed reaction has been considered insignificant and consistent with reaction where the formation of the phosphorane is rate-limiting.

The kinetic solvent isotope effect (KSIE), in turn, shed light to any kinetically significant proton transfer that occurs in a pre-equilibrium or rate-limiting step. In case no KSIE is observed, no proton transfer takes place. k_H/k_D values close to unity are generally considered as an indication of a nucleophilic mechanism. In practice, the interpretation of the results is much more complicated, for the total effect observed may consist of opposing contributions. For example, an inverse equilibrium isotope effect (EIE) on deprotonation of a metal bound L_2O ligand (L is H or D in any combination) and a normal EIE on deprotonation of the attacking nucleophile may result in an observed KSIE close to unity. Interactions with hydrogen bonding groups may also contribute to the observed KSIE, a fact that is often ignored when KSIE values are interpreted, even in cases where such a group significantly enhances the catalytic activity under consideration (e.g., [170]).

Table 6: Heavy-isotope effects determined in the presence and absence of metal ion catalysts.

catalyst	substrate	$^{18}k_{nuc}$	$^{18}k_{lg}$	$^{18}k_{NB}$	$^{15}k_{NO2}$	ref.
30[a]	HPNP (**1**)	0.9874[b]	1.0113		1.0015	[198]
HO^- [c]	HPNP (**1**)	1.0079[b]	1.0064		1.0002	[198]
30	HPNP (**1**)	0.9926[d]	1.0042[d]			[200]
H_2O	HPNP (**1**)	1.0182[d]	1.0021[d]			[200]
Zn^{2+} [e]	UpG	0.986	1.015	1.0007		[197]
HO^- [f]	UpG	0.997	1.0343	0.999		[197]
CuTACN[g]	EtPNP[h]				1.0013	[199]
HO^-	EtpNP				1.0016	[199]

[a]pH 7.8 HEPES buffer, 40 °C. [b]Observed values have been corrected for the calculated EIE for deprotonation of HPNP. [c]pH 10.1 CHES buffer, 67 °C. [d]Based on DFT calculation. [e]10 mmol L^{-1} $ZnNO_3$, pH 7, 90 °C; [f]pH 12, 90 °C; [g]pH 7.2, 70 °C, [h]ethyl *p*-nitrophenyl phosphate.

Often conditions are chosen to avoid any ambiguity resulting from pre-equilibrium proton transfer in order to obtain a KSIE that refers to the catalytic step only. For example, the KSIE of 1.43 reported for the transesterification of HPNP has been determined at pH 10.5 that is well above the kinetic pK_a of the catalyst [159]. According to the authors, the nucleophile is totally deprotonated both in H_2O and D_2O. If this is the case, the KSIE reflects the nucleophilic attack that inevitably takes place in the reaction, but gives no information on how the reactive ionic form has been formed. In case a significant KSIE is observed at pL < pK_a of the catalyst but not at pL markedly higher than the pK_a of the catalyst, a proton transfer is involved in a pre-equilibrium process [169,201].

An exceptionally large KSIE of 13.2 has been reported for the transesterification of a dinucleoside monophosphate, UpG, in the presence Zn^{2+} [197]. There may be other contributing factors, such as interactions to nucleic acid bases, but a very likely explanation stems from precipitation of Zn^{2+} lyoxo species under the experimental conditions. Examples of KSIEs determined for metal-complex-promoted cleavage of DNA and RNA models are listed in Table 7.

Zhang et al. [197] have additionally carried out proton inventory studies on Zn^{2+}-promoted transterification of UpG. The curve k_n/k_0 vs isotopic ratio n was strikingly similar in shape to the one obtained for lyoxide-ion-catalyzed reaction. According to the authors, these curves were consistent with two normal fractionation factors: a large equilibrium effect due the deprotonation of the nucleophile, and another normal effect resulting from the solvation of the transition state.

Medium effects

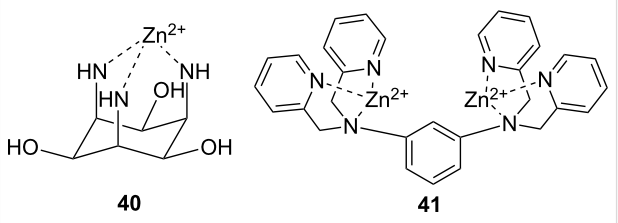
The solvent composition may have a dramatic effect on the rate of metal-ion-complex-promoted reactions, either rate acceleration or deceleration. The most impressive rate-enhancing effect has been reported for the cleavage of activated phosphodiesters by the dinuclear Zn^{2+} complex **33** in the presence of 1 equiv of alkoxide ion in methanol [177] and ethanol [202]. Rate-enhancements up to 10^{12} in comparison to the corresponding background reactions have been observed with HPNP and methyl *p*-nitrophenyl phosphate (MePNP) in methanol [177,203]. In ethanol, the rate enhancement is even higher and the difference increases as the pK_a of the leaving group increases [202,203]. The significant rate enhancements result from stronger binding of the catalyst to substrate and from the reduced permittivity of the medium that allows closer contacts with and within the catalyst. Monomeric Zn^{2+}-TACD complexes, for example, have been observed to act cooperatively at high concentration [177], in striking contrast to the behavior in water. Likewise, the dimeric catalyst **33** cleaves HPNP much more effectively than its monomeric counterpart in methanol but not in water [204]. Any structural change that expectedly weakens association, diminish the rate-enhancing effect of medium. Complex **41** (Figure 15) with a more rigid structure is clearly less efficient than **33**(MeO^-) as a catalyst in methanol [205] and N-methylation of various azacrown-based complexes markedly decreases their catalytic efficiency in methanol [206].

Owing to very efficient cleavage of HPNP in the presence **33** (MeO^-) in methanol, binding of the catalyst to substrate becomes rate limiting [205]. The efficiency of the binding events has been evaluated by using colored Cu^{2+} analog of **33** (MeO^-) as a catalyst [207]. The colorimetric analysis showed that binding is a two-step process. The first of these is very fast and the rate is linearly dependent on the catalyst concentration. The second is a concentration-independent rearrangement that forms the active species with dinuclear Cu^{2+} coordination. The rate constants for the latter step are almost equal with MePNP and HPNP, 0.57 s^{-1} and 0.72 s^{-1}, respectively. As the rate con-

Table 7: Solvent isotope effects reported for reactions of phosphodiesters in the presence of metal ion catalysts.

catalyst	substrate	conditions/reaction	KSIE	ref.
Cu^{2+}-TACN	EtPNP[a]	pH 9	k_H/k_D = 1.14	[199]
39	BNPP (**23a**)	catalysis by a mono-deprotonated species	$k_{2,H}/k_{2,D}$ = 0.8	[169]
36	BNPP (**23a**)	pL = 7.9	k_H/k_D = 1.26	[182]
Tb^{3+}-**31**	BDNPP (**23b**)	pL = 7, 75% MeCN	k_H/k_D = 1.14	[144]
Cu^{2+}-TerPy	cAMP[b]	catalysis by a mono-deprotonated species	$k_{2,H}/k_{2,D}$ = 1	[151]
40	HPNP (**1**)	pH 10.5	k_H/k_D = 1.43	[159]
36	HPNP (**1**)	pL = 7.3	k_H/k_D = 2.76	[182]
30	UpPNP (**24**)	pL > 9	$k_{c,H}/k_{c,D}$ = 0.8	[201]
Zn^{2+}	UpEtoEt (**25**)	pL = 5.6, 90 °C	k_H/k_D = 2.7	[176]
27a	UpEtoEt (**25**)	pL = 6.5, 25 °C	k_H/k_D = 2.7	[176]
Zn^{2+}	UpG	pL = 7.0, 90°C	k_H/k_D = 13.2	[197]

[a]Ethyl *p*-nitrophenyl phosphate; [b]adenosine 2′,3′-cyclic phosphate.



40 41

Figure 15: Metal ion complexes used in kinetic studies of medium effects on the cleavage of RNA model compounds.

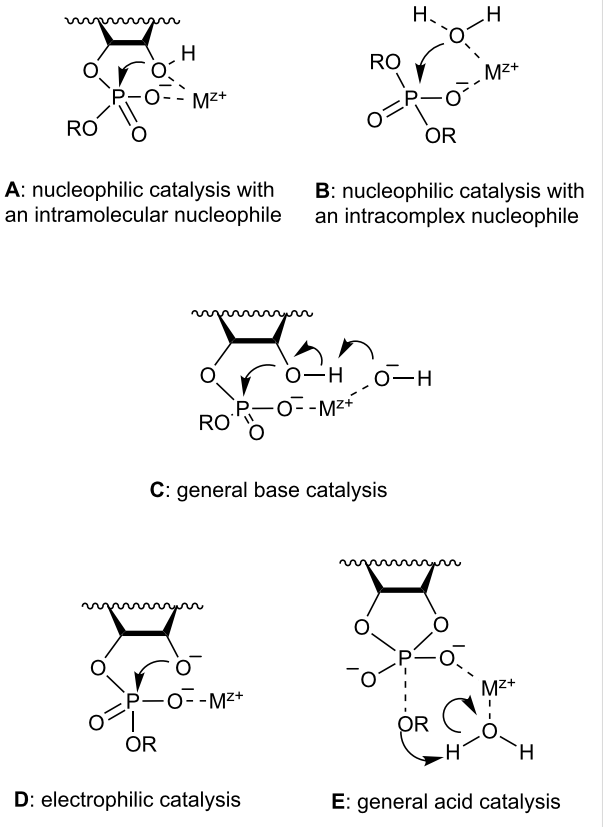
stant for the chemical cleavage of HPNP under the same conditions is 0.7 s^{-1}, the latter binding step is rate-limiting. With the less reactive DNA analog, MePNP, the chemical cleavage step still is clearly rate-limiting.

In contrast to alcohols, DMSO and acetonitrile have been shown to retard the metal-complex-promoted cleavage of phosphodiesters. The effect of DMSO has been utilized to distinguish between general base-catalyzed and specific base-catalyzed reaction routes, as specific base-catalyzed reactions are suppressed in DMSO rich mixtures, owing to suppressed autoprotolysis of water [208]. Second order rate constants for the metal-ion-promoted reactions have been determined in 80% aqueous DMSO in different buffers keeping the buffer ratio constant but increasing the total buffer concentration. When the rate constants are plotted against the buffer ratio or the concentration of the base form, the shape of the plots indicates whether either a specific base or a general base-catalyzed reaction is suppressed. According to such an analysis, all metal ions studied enhance the specific base-catalyzed reaction of HPNP, whereas the general base-catalyzed reaction is assisted only by Mg^{2+} and Na^{+}. KSIE values of 0.25 and 0.36 have been determined for the specific base-catalyzed reactions in the presence of Mg^{2+} and Ca^{2+}, respectively and a value of 1.23 for the Mg^{2+}-assisted general base-catalyzed reaction.

Despite the inhibition, organic co-solvents are often used to improve the solubility of the substrate or the catalyst [143,144,171]. In some cases the inhibition is strong enough to completely prevent the catalysis, although conflicting reports also exist. While Zn^{2+}-TACD has been reported to catalyze the cleavage of HPNP efficiently in 50% aqueous acetonitrile [156], complete inactivation of Cu^{2+} and Zn^{2+} complexes of a related catalyst **35** was observed in the same medium [183]. The authors have speculated that the cyano group of acetonitrile binds the catalysts hence occupying one or more coordination sites of the catalysts.

Mechanistic conclusions

Despite extensive studies, no universally accepted mechanism for metal ion catalysis has been found. There is, however, a fairly unanimous understanding of the importance of deprotonation event at pH close to that of the pK_a of a metal bound aqua ligand. Three different basic mechanisms have been proposed to explain the need for deprotonation: intracomplex nucleophilic catalysis (**A** and **B** in Scheme 14), intracomplex general base catalysis (**C**) and electrophilic (**D**) or general acid (**E**) catalysis on an intermediate obtained by a specific base-catalyzed reaction. Intermolecular general base or nucleophilic mechanisms are not considered feasible, since the catalysis by metal ion species is much more significant than by organic bases or nucleophiles.



Scheme 14: Alternative mechanisms for metal-ion-promoted cleavage of phosphodiesters.

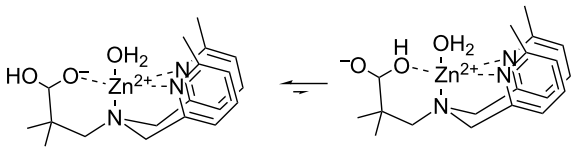
The nucleophilic mechanism in this context involves a nucleophilic attack by a group coordinated to the metal ion catalyst. In case of DNA type substrates [159,160,182,198], the nucleophile is likely to be a metal bound hydroxo ligand (**B**), whereas with RNA type of substrates the nucleophile is the neighboring OH group on the substrate (**A**). Metal ion catalysts enhance deprotonation of the nucleophile by coordination, and since the pK_a values of metal-bound H_2O and alcohols are likely to be of the same order of magnitude, the pH-dependence for reactions of both types of substrates is generally similar.

The nucleophilic mechanism is widely accepted for the reactions of DNA model compounds, such as BNPP [145,160,169,199], but also for HPNP [159,160]. KSIE values are close to unity, which is generally regarded as an evidence of a nucleophilic mechanism. Furthermore, it has been reported that under conditions where the metal-bound aqua ligand is completely deprotonated, the catalytic activity of metal ion catalysts increases with increasing pK_a, as long as complexes of similar type (tridentate vs tetradentate) are concerned. This has been suggested to indicate that the catalytic activity at high pH depends on nucleophilicity of the metal-bound hydroxy ligand [159]. Tetradentate complexes are less efficient catalysts than tridentate ones of similar acidity, consistent with the need of a free hydroxo ligand to act as a nucleophile [160].

The dependence on the pK_a of the catalysts is similar in reactions of BNPP and HPNP, and the KSIE of 1.45 determined for the transtesterification of HPNP at pH 10.5 is within the range typical for nucleophilic catalysis [159]. In contrast, bimetallic complex **36** has been suggested to enhance the reaction BNPP by different mechanisms [182]. pH-Rate profiles for the reactions of the two substrates are different suggesting that different deprotonation events are involved. Furthermore, KSIE effects determined under the same conditions point to different mechanisms: while that for the reaction of BNPP is typical for nucleophilic catalysis, a value of 2.76 determined for the reaction of HPNP is of a magnitude typical to general base catalysis.

A metal-ion-bound hydroxide or alkoxide ion certainly is a weaker nucleophile than their free counterparts. Still virtually all metal-ion-based catalysts for the cleavage of BNPP are based on the attack of a metal-ion-bound nucleophile. Only rather recently, it has been shown that by carefully ligand design a situation may be achieved, where an unbound alkoxy group serves as a powerful nucleophile [209]. The key feature is a hydrated aldehyde group locked by a proper position Zn^{2+} coordinated additionally to three nitrogen atoms within ligand **42** (Figure 16). The *gem*-diol system may be coordinated to the central ion through alkoxy oxygen, but also through hydroxy oxygen, leaving the alkoxy function free to serve as a nucleophile. Although the latter species is a minor tautomer, its reactivity is high enough to overcome the unfavorable equilibrium.

Electrophilic catalysis or Lewis acid catalysis (**D**) has repeatedly been suggested for the reactions of RNA type substrates. The phosphate-bound metal ion catalyst activates the substrate towards nucleophilic attack, the nucleophile being neutral or deprotonated depending on the pH. The sigmoidal or bell-shaped pH-rate profiles can be understood by considering the effects of increasing pH on both the catalyst and substrate. The



42

Figure 16: Nucleic acid cleaving agents where the attacking oxyanion is not coordinated to metal ion.

proportion of the anionic nucleophile, and hence, the efficiency of the nucleophilic attack is increased as long as the pK_a value of the secondary OH group around pH 12 is reached. On passing the pK_a value of catalyst aqua ligands, generally at pH 7–9, binding to the phosphodiester group is weakened and the electrophilic contribution of the catalysis is lost. This mechanism has been proposed to be utilized, for example, by the most efficient bifunctional catalysts **30** and **27a** of the transesterification of HPNP. Williams et al. [168] have justified their mechanistic choice by studying the two kinetically equivalent mechanisms: deprotonation of neutral substrate by a deprotonated complex that acts a general base, and specific base-catalyzed reaction of a substrate activated by the aqua form of the catalyst. Because dimethyl phosphate inhibits the reaction more strongly at a lower pH, where the proportion of the aqua form is higher, it has been concluded that the inhibitor competes for the aqua form. This has been taken as an evidence of the electrophilic mechanism, where the aqua form is the active catalyst (**D**).

Transesterification of nucleoside phosphoesters, UpNP and UpU, by **30** has also been suggested to proceed by mere electrophilic catalyzed pathway [170]. Similar pH-dependence with three different types of substrates has been taken as an indication of similar ionic forms being important in the reactions. Furthermore, a KSIE value of 0.8 has been determined for **30**-promoted reaction of UpNP at pL > 9, which shows that no proton transfer takes place in the reaction, when the formation of the phosphorane is rate-limiting (**D**). A proton transfer to assist the departure of the poor leaving group of UpU has been rejected on the basis of microscopic reversibility.

In contrast to catalysis by **30**, two different mechanisms have been proposed for the **27a**-promoted reactions of HPNP and UpU. While HPNP with a good leaving group is most probably cleaved without general acid/base catalysis (**D**) [167], a KSIE of 2.7 for the **27a**-catalyzed reaction of UpEtOEt with a poor leaving group suggests proton transfer in the rate-limiting step [176]. Since the KSIE for the **27a**-promoted phosphate migra-

tion is close to unity [176], the proton transfer most probably enhances the breakdown of the phosphorane intermediate to cleavage products. In other words, general acid catalysis appears to be involved (**E**).

General base catalysis by a metal-bound hydroxo ligand (**C**) is the most obvious way of interpreting the sigmoidal or bell-shaped pH-rate profiles**.** The rate of reaction increases as the proportion of the hydroxo form of the catalyst is increased. At pH < pK_a of the metal-bound aqua ligands, the hydroxo form of the catalyst is the strongest base. KSIE values determined under such conditions fall within the range usually attributed to general base catalysis [176,182]. It has, however, been recognized that a kinetically equivalent specific base – general acid catalysis, i.e., pre-equilibrium deprotonation of the attacking nucleophile followed by general acid-catalyzed breakdown of the intermediate, appears more feasible when the substrate has a poor leaving group. Consistent with this suggestion, modestly negative β_{lg} values have been observed for metal-ion-promoted reactions of nucleoside 3'-alkyl phosphates [176,190]. In this respect, metal-ion-promoted reactions resemble more the acid-catalyzed reaction than the base-catalyzed. Furthermore, an analysis of the effect of the acidity of the leaving group alcohol on the catalysis by various metal ion complexes shows that the most acidic catalysts fail to promote the transesterification of the substrates with most basic leaving groups [189]. Results obtained with ^{18}O experiments on Zn^{2+}-catalyzed reaction of UpG [197] may also be taken as an indication of catalysis mechanism that affects the departure of the leaving group.

The preceding discussion shows that all three basic mechanistic alternatives are firmly supported by experimental evidence. Theoretical calculations based on density functional theory do not solve the controversy, either [160,200,210]. All theoretical studies generally support a concerted reaction mechanism and indicate a number of important interactions to the nucleophile, phosphate and leaving group. Many of the studies concentrate also on the deprotonation of the nucleophile and both pre-equilibrium [200,211] and concerted processes [183,210] have been predicted. Regardless of timing, the nucleophile may also be coordinated to a metal ion [160,200,210].

Most probably the mechanism depends on both the substrate and the catalyst. Consistent with this, there are examples showing that two different types of substrates may be cleaved by two different mechanisms in the presence of the same catalyst. Furthermore, an analysis of the effect of the acidity of the leaving group in nucleoside phosphodiesters shows, that, generally, a more efficient catalysis is observed when there is an imbalance between the properties of a nucleophile and the leaving group. Results in Table 2 suggest that this may be extended even further and the extremes on the scale would be DNA model BNPP with no intramolecular nucleophile and a good leaving group, and dinucleoside monophosphates with a favorably positioned nucleophile and a poor leaving group. It would seem logical to assume that catalysis required in corresponding reactions is different.

Beyond the scope of the present review are the nanostructured cleaving agents that show cooperativity between the catalytic functions on particle surface [212] and sequence-selective cleaving agents that consist of an artificial cleaving agent conjugated to a sequence recognizing moiety [213]. Finally, it is worth noting that in spite of extensive studies of the metal-ion-promoted cleavage of nucleic acids, the applications still are scanty. There is only one patiently developed application that deserves to be highlighted, viz. the manipulation of large genomes by Ce^{4+}-promoted cleavage followed by enzymatic ligation. The description of this fascinating technique is, however, outside the scope of the present paper. Recent reviews on the subject [214-216] are recommended.

Conclusion

Experimental studies with small molecular model compounds of nucleic acids allow evaluation of the importance of various elementary processes, such as proton transfer and metal ion binding, for stabilization of transition states and systematic variation of the basicity of the entering and departing nucleophile enables determination of the position of transition state on the reaction coordinate. Such data is helpful on analyzing mechanisms of enzymatic processes. Studies with RNA models have been more extensive than those with DNA models. The predominant buffer-independent reactions of RNA 3',5'-phosphodiester linkages under neutral conditions are approximately as fast pH-independent isomerization to 2′,5′-bonds and hydroxide-ion-catalyzed transesterification to a 2′,3′-cyclic phosphate. The kinetics and mechanisms of these reactions are rather well known. By contrast, the detailed mechanisms of buffer-catalyzed reactions still seem to be open to various interpretations of kinetic data. Catalysis by multifunctional agents containing amino, imidazole and guanidine groups have received special attention, owing to presence of such functions at the side chains of catalytically important amino acids in nucleases. The mechanistic studies on cleavage of DNA are scanty. The very high stability of the phosphodiester bonds within DNA has clearly limited the interest. The metal-ion-promoted cleavage of both RNA and DNA has recently received increasing interest. Extensive studies have led to a number of mechanistic suggestions, but more systematic studies with various substrates and catalysts are still needed to draw firm mechanistic conclusions.

ORCID® iDs

Satu Mikkola - https://orcid.org/0000-0002-5149-0252
Tuomas Lönnberg - https://orcid.org/0000-0003-3607-3116
Harri Lönnberg - https://orcid.org/0000-0002-8494-4062

References

1. Yang, W. *Q. Rev. Biophys.* **2011,** *44,* 1–93. doi:10.1017/S0033583510000181
2. Schroeder, G. K.; Lad, C.; Wyman, P.; Williams, N. H.; Wolfenden, R. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103,* 4052–4055. doi:10.1073/pnas.0510879103
3. Brown, D. M.; Magrath, D. I.; Neilson, A. H.; Todd, A. R. *Nature* **1956,** *177,* 1124–1125. doi:10.1038/1771124a0
4. Kaukinen, U.; Lyytikäinen, S.; Mikkola, S.; Lönnberg, H. *Nucleic Acids Res.* **2002,** *30,* 468–474. doi:10.1093/nar/30.2.468
5. Linjalahti, H.; Mikkola, S. *Chem. Biodiversity* **2007,** *4,* 2938–2947. doi:10.1002/cbdv.200790243
6. Thompson, J. E.; Kutateladze, T. G.; Schuster, M. C.; Venegas, F. D.; Messmore, J. M.; Raines, R. T. *Bioorg. Chem.* **1995,** *23,* 471–481. doi:10.1006/bioo.1995.1033
7. Lilley, D. M. J.; Eckstein, F., Eds. *Ribozymes and RNA catalysis;* Royal Society of Chemistry: Cambridge, UK, 2008.
8. Cleland, W. W.; Hengge, A. C. *Chem. Rev.* **2006,** *106,* 3252–3278. doi:10.1021/cr050287o
9. Lilley, D. M. J. *Philos. Trans. R. Soc., B* **2011,** *366,* 2910–2917. doi:10.1098/rstb.2011.0132
10. Ward, W. L.; Plakos, K.; DeRose, V. J. *Chem. Rev.* **2014,** *114,* 4318–4342. doi:10.1021/cr400476k
11. Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. *Acc. Chem. Res.* **2015,** *48,* 220–228. doi:10.1021/ar500314j
12. Dissanayake, T.; Swails, J. M.; Harris, M. E.; Roitberg, A. E.; York, D. M. *Biochemistry* **2015,** *54,* 1307–1313. doi:10.1021/bi5012833
13. Elsässer, B.; Fels, G.; Weare, J. H. *J. Am. Chem. Soc.* **2014,** *136,* 927–936. doi:10.1021/ja406122c
14. Kamerlin, S. C.; Sharma, P. K.; Prasad, R. B.; Warshel, A. *Q. Rev. Biophys.* **2013,** *46,* 1–132. doi:10.1017/S0033583512000157
15. Huang, M.; York, D. M. *Phys. Chem. Chem. Phys.* **2014,** *16,* 15846–15855. doi:10.1039/C4CP01050G
16. Wong, K.-Y.; Gu, H.; Zhang, S.; Piccirilli, J. A.; Harris, M. E.; York, D. M. *Angew. Chem., Int. Ed.* **2012,** *51,* 647–651. doi:10.1002/anie.201104147
17. Chen, H.; Giese, T. J.; Huang, M.; Wong, K.-Y.; Harris, M. E.; York, D. M. *Chem. – Eur. J.* **2014,** *20,* 14336–14343. doi:10.1002/chem.201403862
18. Desbouis, D.; Troitsky, I. P.; Belousoff, M. J.; Spiccia, L.; Graham, B. *Coord. Chem. Rev.* **2012,** *256,* 897–937. doi:10.1016/j.ccr.2011.12.005
19. Westheimer, F. H. *Acc. Chem. Res.* **1968,** *1,* 70–78. doi:10.1021/ar50003a002
20. Berry, R. S. *J. Chem. Phys.* **1960,** *32,* 933–938. doi:10.1063/1.1730820
21. Ugi, I.; Marquarding, D.; Klusacek, H.; Gillespie, P.; Ramirez, F. *Acc. Chem. Res.* **1971,** *4,* 288–296. doi:10.1021/ar50044a004
22. Davies, J. E.; Doltsinis, N. L.; Kirby, A. J.; Roussev, C. D.; Sprik, M. *J. Am. Chem. Soc.* **2002,** *124,* 6494–6599. doi:10.1021/ja025779m
23. Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002,** *124,* 5010–5018. doi:10.1021/ja011373i
24. Lönnberg, H. In *Chemical Biology of Nucleic Acids: Fundamentals and Clinical Applications;* Erdman, V. A.; Markiewicz, W. T.; Barciszewski, J., Eds.; Springer Verlag: Berlin Heidelberg, 2014; pp 41–56. doi:10.1007/978-3-642-54452-1_3
25. López, C. S.; Faza, O. N.; Gregersen, B. A.; Lopez, X.; de Lera, A. R.; York, D. M. *ChemPhysChem* **2004,** *5,* 1045–1049. doi:10.1002/cphc.200400091
26. Lopez, C. S.; Faza, O. N.; de Lera, A. R.; York, D. M. *Chem. – Eur. J.* **2005,** *11,* 2081–2093. doi:10.1002/chem.200400790
27. Salvio, R.; Casnati, A. *J. Org. Chem.* **2017,** *82,* 10461–10469. doi:10.1021/acs.joc.7b01925
28. Perrault, D. M.; Anslyn, E. V. *Angew. Chem., Int. Ed. Engl.* **1997,** *36,* 432–450. doi:10.1002/anie.199704321
29. Oivanen, M.; Kuusela, S.; Lönnberg, H. *Chem. Rev.* **1998,** *98,* 961–990. doi:10.1021/cr960425x
30. Lönnberg, H. *Org. Biomol. Chem.* **2011,** *9,* 1687–1703. doi:10.1039/c0ob00486c
31. Pereira, E. S.; Da Silva, J. C. S.; Brandão, T. A. S.; Rocha, W. R. *Phys. Chem. Chem. Phys.* **2016,** *18,* 18255–18267. doi:10.1039/C6CP01536K
32. Emilsson, G. M.; Nakamura, S.; Roth, A.; Breaker, R. R. *RNA* **2003,** *9,* 907–918. doi:10.1261/rna.5680603
33. Williams, A. *Free Energy Relationships in Organic and Bio-organic Chemistry;* The Royal Society of Chemistry: Cambridge, 2003.
34. Bourne, N.; Williams, A. *J. Org. Chem.* **1984,** *49,* 1200–1204. doi:10.1021/jo00181a011
35. Malier, J. F. *Acc. Chem. Res.* **2001,** *34,* 283–290. doi:10.1021/ar000054d
36. Hengge, A. C. *Acc. Chem. Res.* **2002,** *35,* 105–112. doi:10.1021/ar000143q
37. Melander, L.; Saunders, W. H. *Reaction Rates of Isotopic Molecules;* Wiley: New York, 1980.
38. Corona-Martinez, D. O.; Taran, O.; Yatsimirsky, A. K. *Org. Biomol. Chem.* **2010,** *8,* 873–880. doi:10.1039/B920398B
39. Oivanen, M.; Mikhailov, S. N.; Florentiev, V. L.; Vihanto, P.; Lönnberg, H. *Acta Chem. Scand.* **1993,** *47,* 622–625. doi:10.3891/acta.chem.scand.47-0622
40. Li, N.-S.; Frederiksen, J. K.; Piccirilli, J. A. *Acc. Chem. Res.* **2011,** *44,* 1257–1269. doi:10.1021/ar200131t
41. Frederiksen, J. K.; Piccirilli, J. A. *Methods* **2009,** *49,* 148–166. doi:10.1016/j.ymeth.2009.07.005
42. Lönnberg, T. *Chem. – Eur. J.* **2011,** *17,* 7140–7153. doi:10.1002/chem.201100009
43. Ora, M.; Lönnberg, T.; Lönnberg, H. In *From Nucleic Acid Sequences to Molecular Medicine;* Erdman, V. A.; Barciszewski, J., Eds.; Springer Verlag: Berlin Heidelberg, 2012; pp 47–65. doi:10.1007/978-3-642-27426-8_3
44. Järvinen, P.; Oivanen, M.; Lönnberg, H. *J. Org. Chem.* **1991,** *56,* 5396–5401. doi:10.1021/jo00018a037
45. Kuusela, S.; Lönnberg, H. *J. Chem. Soc., Perkin Trans. 2* **1994,** 2109–2113. doi:10.1039/P29940002109
46. Lopez, X.; Dejaegere, A.; Leclerc, F.; York, D. M.; Karplus, M. *J. Phys. Chem. B* **2006,** *110,* 11525–11539. doi:10.1021/jp0603942
47. Yang, Y.; Cui, Q. *J. Phys. Chem. B* **2009,** *113,* 4930–4939. doi:10.1021/jp810755p
48. Kosonen, M.; Yousefi-Salakdeh, E.; Strömberg, R.; Lönnberg, H. *J. Chem. Soc., Perkin Trans. 2* **1998,** 1589–1595. doi:10.1039/a801670d

49. Kosonen, M.; Yousefi-Salakdeh, E.; Strömberg, R.; Lönnberg, H. *J. Chem. Soc., Perkin Trans. 2* **1997,** 2661–2666. doi:10.1039/a704636g
50. Kosonen, M.; Hakala, K.; Lönnberg, H. *J. Chem. Soc., Perkin Trans. 2* **1998,** 663–670. doi:10.1039/a707095k
51. Davis, A. M.; Hall, A. D.; Williams, A. *J. Am. Chem. Soc.* **1988,** *110,* 5105–5108. doi:10.1021/ja00223a031
52. Lönnberg, H.; Strömberg, R.; Williams, A. *Org. Biomol. Chem.* **2004,** *2,* 2165–2167. doi:10.1039/B406926A
53. Ye, J.-D.; Li, N.-S.; Dai, Q.; Piccirilli, J. A. *Angew. Chem., Int. Ed.* **2007,** *46,* 3714–3717. doi:10.1002/anie.200605124
54. Harris, M. E.; Dai, Q.; Gu, H.; Kellerman, D. L.; Piccirilli, J. A.; Anderson, V. E. *J. Am. Chem. Soc.* **2010,** *132,* 11613–11621. doi:10.1021/ja103550e
55. Virtanen, N.; Polari, L.; Välilä, M.; Mikkola, S. *J. Phys. Org. Chem.* **2005,** *18,* 385–397. doi:10.1002/poc.883
56. Virtanen, N.; Nevalainen, V.; Lehtinen, T.; Mikkola, S. *J. Phys. Org. Chem.* **2007,** *20,* 72–82. doi:10.1002/poc.1130
57. Hengge, A. C.; Bruzik, K. S.; Tobin, A. E.; Cleland, W. W.; Tsai, M.-D. *Bioorg. Chem.* **2000,** *28,* 119–133. doi:10.1006/bioo.2000.1170
58. Raines, R. T. *Chem. Rev.* **1998,** *98,* 1045–1066. doi:10.1021/cr960427h
59. Anslyn, E.; Breslow, R. *J. Am. Chem. Soc.* **1989,** *111,* 4473–4482. doi:10.1021/ja00194a050
60. Breslow, R. *Acc. Chem. Res.* **1991,** *24,* 317–324. doi:10.1021/ar00011a001
61. Haim, A. *J. Am. Chem. Soc.* **1992,** *114,* 8384–8388. doi:10.1021/ja00048a006
62. Kirby, A. J.; Marriott, R. E. *J. Am. Chem. Soc.* **1995,** *117,* 833–834. doi:10.1021/ja00107a034
63. Perrin, C. L. *J. Org. Chem.* **1995,** *60,* 1239–1243. doi:10.1021/jo00110a030
64. Breslow, R.; Dong, S. D.; Webb, Y.; Xu, R. *J. Am. Chem. Soc.* **1996,** *118,* 6588–6600. doi:10.1021/ja9526933
65. Beckmann, C.; Kirby, A. J.; Kuusela, S.; Tickle, D. C. *J. Chem. Soc., Perkin Trans. 2* **1998,** 573–581. doi:10.1039/a707741f
66. Baughman, E. H.; Kreevoy, M. M. *J. Phys. Chem.* **1974,** *78,* 421–423. doi:10.1021/j100597a021
67. Pawlak, Z.; Bates, R. G. *J. Solution Chem.* **1975,** *4,* 817–829. doi:10.1007/BF00650538
68. Lain, L.; Lönnberg, H.; Lönnberg, T. *Org. Biomol. Chem.* **2015,** *13,* 3484–3492. doi:10.1039/C4OB02682A
69. Salvio, R. *Chem. – Eur. J.* **2015,** *21,* 10960–10971. doi:10.1002/chem.201500789
70. Cotton, F. A.; Hazen, E. E. J.; Legg, M. J. *Proc. Natl. Acad. Sci. U. S. A.* **1979,** *76,* 2551–2555. doi:10.1073/pnas.76.6.2551
71. Redinbo, M. R.; Stewart, L.; Kuhn, P.; Champoux, J. J.; Hol, W. G. J. *Science* **1998,** *279,* 1504–1513. doi:10.1126/science.279.5356.1504
72. Martick, M.; Scott, W. G. *Cell* **2006,** *126,* 309–320. doi:10.1016/j.cell.2006.06.036
73. Han, J.; Burke, J. M. *Biochemistry* **2005,** *44,* 7864–7870. doi:10.1021/bi047941z
74. Rupert, P. B.; Ferré-DA'maré, A. R. *Nature* **2001,** *410,* 780–786. doi:10.1038/35071009
75. Perreault, D. M.; Cabell, L. A.; Anslyn, E. V. *Bioorg. Med. Chem.* **1997,** *5,* 1209–1220. doi:10.1016/S0968-0896(97)00051-5
76. Smith, J.; Ariga, K.; Anslyn, E. V. *J. Am. Chem. Soc.* **1993,** *115,* 362–364. doi:10.1021/ja00054a062
77. Scheffer, U.; Strick, A.; Ludwig, V.; Peter, S.; Kalden, E.; Göbel, M. W. *J. Am. Chem. Soc.* **2005,** *127,* 2211–2217. doi:10.1021/ja0443934
78. Lönnberg, T. A.; Helkearo, M.; Jancso, A.; Gajda, T. *Dalton Trans.* **2012,** *41,* 3328–3338. doi:10.1039/c2dt10193a
79. Lönnberg, T.; Luomala, M. *Org. Biomol. Chem.* **2012,** *10,* 6785–6791. doi:10.1039/c2ob25958c
80. Baldini, L.; Cacciapaglia, R.; Casnati, A.; Mandolini, L.; Salvio, R.; Sansone, F.; Ungaro, R. *J. Org. Chem.* **2012,** *77,* 3381–3389. doi:10.1021/jo300193y
81. Salvio, R.; Mandolini, L.; Savelli, C. *J. Org. Chem.* **2013,** *78,* 7259–7263. doi:10.1021/jo401085z
82. Salvio, R.; Cacciapaglia, R.; Mandolini, L.; Sansone, F.; Casnati, A. *RSC Adv.* **2014,** *4,* 34412–34416. doi:10.1039/C4RA05751A
83. Komiyama, M.; Yoshinari, K. *J. Org. Chem.* **1997,** *62,* 2155–2160. doi:10.1021/jo961935u
84. Bencini, A.; Berni, E.; Bianchi, A.; Giorgi, C.; Valtancoli, B. *Supramol. Chem.* **2001,** *13,* 489–497. doi:10.1080/10610270108029464
85. Lain, L.; Lahdenpohja, S.; Lönnberg, H.; Lönnberg, T. *Int. J. Mol. Sci.* **2015,** *16,* 17798–17811. doi:10.3390/ijms160817798
86. Lain, L.; Lönnberg, H.; Lönnberg, T. *Chem. – Eur. J.* **2013,** *19,* 12424–12434. doi:10.1002/chem.201301711
87. Lain, L.; Lönnberg, H.; Lönnberg, T. *Org. Biomol. Chem.* **2015,** *13,* 4737–4742. doi:10.1039/C5OB00400D
88. Kosonen, M.; Lönnberg, H. *J. Chem. Soc., Perkin Trans. 2* **1995,** 1203–1209. doi:10.1039/P29950001203
89. Ora, M.; Linjalahti, H.; Lönnberg, H. *J. Am. Chem. Soc.* **2005,** *127,* 1826–1832. doi:10.1021/ja045060+
90. Liu, X.; Reese, C. B. *Tetrahedron Lett.* **1996,** *37,* 925–928. doi:10.1016/0040-4039(95)02281-3
91. Weinstein, L. B.; Earnshaw, D. J.; Cosstick, R.; Cech, T. R. *J. Am. Chem. Soc.* **1996,** *118,* 10341–10350. doi:10.1021/ja9616903
92. Iyer, S.; Hengge, A. C. *J. Org. Chem.* **2008,** *73,* 4819–4829. doi:10.1021/jo8002198
93. Elzagheid, M. I.; Oivanen, M.; Klika, K. D.; Jones, B. C. N. M.; Cosstick, R.; Lönnberg, H. *Nucleosides Nucleotides* **1999,** *18,* 2093–2108. doi:10.1080/07328319908044866
94. Liu, X.; Reese, C. B. *Tetrahedron Lett.* **1995,** *36,* 3413–3416. doi:10.1016/0040-4039(95)00495-X
95. Thomson, J. B.; Patel, B. K.; Jiménez, V.; Eckart, K.; Eckstein, F. *J. Org. Chem.* **1996,** *61,* 6273–6281. doi:10.1021/jo960795l
96. Almer, H.; Strömberg, R. *Tetrahedron Lett.* **1991,** *32,* 3723–3726. doi:10.1016/S0040-4039(00)79778-4
97. Oivanen, M.; Ora, M.; Almer, H.; Strömberg, R.; Lönnberg, H. *J. Org. Chem.* **1995,** *60,* 5620–5627. doi:10.1021/jo00122a050
98. Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004,** *126,* 7504–7513. doi:10.1021/ja031815l
99. Liu, Y.; Gregersen, B. A.; Lopez, X.; York, D. M. *J. Phys. Chem. B* **2005,** *109,* 19987–20003. doi:10.1021/jp053146z
100. Liu, Y.; Gregersen, B. A.; Hengge, A.; York, D. M. *Biochemistry* **2006,** *45,* 10043–10053. doi:10.1021/bi060869f
101. Ora, M.; Hanski, A. *Helv. Chim. Acta* **2011,** *94,* 1563–1574. doi:10.1002/hlca.201100115
102. Almer, H.; Strömberg, R. *J. Am. Chem. Soc.* **1996,** *118,* 7921–7928. doi:10.1021/ja953399d
103. Hudson, R. F. *Pure Appl. Chem.* **1964,** *9,* 371–386. doi:10.1351/pac196409020371
104. Ora, M.; Järvi, J.; Oivanen, M.; Lönnberg, H. *J. Org. Chem.* **2000,** *65,* 2651–2657. doi:10.1021/jo991632a

105. Takagi, Y.; Warashina, M.; Stec, W. J.; Yoshinari, K.; Taira, K. *Nucleic Acids Res.* **2001,** *29,* 1815–1834. doi:10.1093/nar/29.9.1815
106. Zaug, A. J.; Grabowski, P. J.; Cech, T. R. *Nature* **1983,** *301,* 578–583. doi:10.1038/301578a0
107. Bass, B. L.; Cech, T. R. *Nature* **1984,** *308,* 820–826. doi:10.1038/308820a0
108. Guerrier-Takada, C.; Haydock, K.; Allen, L.; Altman, S. *Biochemistry* **1986,** *25,* 1509–1515. doi:10.1021/bi00355a006
109. Grosshans, C. A.; Cech, T. R. *Biochemistry* **1989,** *28,* 6888–6894. doi:10.1021/bi00443a017
110. Haydock, K.; Allen, L. C. *Prog. Clin. Biol. Res.* **1985,** *172A,* 87–98.
111. Roussev, C. D.; Ivanova, G. D.; Bratovanova, E. K.; Vassilev, N. G.; Petkov, D. D. *J. Am. Chem. Soc.* **1999,** *121,* 11267–11272. doi:10.1021/ja990491r
112. Kosonen, M.; Oivanen, M.; Lönnberg, H. *J. Org. Chem.* **1994,** *59,* 3704–3708. doi:10.1021/jo00092a037
113. Lönnberg, T.; Mikkola, S. *J. Org. Chem.* **2004,** *69,* 802–810. doi:10.1021/jo035094k
114. Lönnberg, T.; Kero, K.-M. *Org. Biomol. Chem.* **2012,** *10,* 569–574. doi:10.1039/C1OB06399E
115. Cech, T. R.; Zaug, A. J.; Grabowski, P. J. *Cell* **1981,** *27,* 487–496. doi:10.1016/0092-8674(81)90390-1
116. McSwiggen, J. A.; Cech, T. R. *Science* **1989,** *244,* 679–683. doi:10.1126/science.2470150
117. Rajagopal, J.; Doudna, J. A.; Szostak, J. W. *Science* **1989,** *244,* 692–694. doi:10.1126/science.2470151
118. Been, M. D.; Cech, T. R. *Cell* **1987,** *50,* 951–961. doi:10.1016/0092-8674(87)90522-8
119. Vicens, Q.; Cech, T. R. *Trends Biochem. Sci.* **2006,** *31,* 41–51. doi:10.1016/j.tibs.2005.11.008
120. Yoshida, A.; Shan, S.-o.; Herschlag, D.; Piccirilli, J. A. *Chem. Biol.* **2000,** *7,* 85–96. doi:10.1016/S1074-5521(00)00074-0
121. Herschlag, D.; Eckstein, F.; Cech, T. R. *Biochemistry* **1993,** *32,* 8299–8311. doi:10.1021/bi00083a034
122. Herschlag, D.; Eckstein, F.; Cech, T. R. *Biochemistry* **1993,** *32,* 8312–8321. doi:10.1021/bi00083a035
123. Roitzsch, M.; Fedorova, O.; Pyle, A. M. *Nat. Chem. Biol.* **2010,** *6,* 218–224. doi:10.1038/nchembio.312
124. Kiviniemi, A.; Lönnberg, T.; Ora, M. *J. Am. Chem. Soc.* **2004,** *126,* 11040–11045. doi:10.1021/ja047568i
125. Lönnberg, T.; Kiiski, J.; Mikkola, S. *Org. Biomol. Chem.* **2005,** *3,* 1089–1096. doi:10.1039/B500054H
126. Lönnberg, T.; Korhonen, J. *J. Am. Chem. Soc.* **2005,** *127,* 7752–7758. doi:10.1021/ja050325l
127. Lönnberg, T.; Laine, M. *Org. Biomol. Chem.* **2010,** *8,* 349–356. doi:10.1039/B912042D
128. Lönnberg, T.; Ora, M.; Virtanen, S.; Lönnberg, H. *Chem. – Eur. J.* **2007,** *13,* 4614–4627. doi:10.1002/chem.200601835
129. Hougland, J. L.; Sengupta, R. N.; Dai, Q.; Deb, S. K.; Piccirilli, J. A. *Biochemistry* **2008,** *47,* 7684–7694. doi:10.1021/bi8000648
130. Forconi, M.; Sengupta, R. N.; Piccirilli, J. A.; Herschlag, D. *Biochemistry* **2010,** *49,* 2753–2762. doi:10.1021/bi902200n
131. Strobel, S. A.; Ortoleva-Donnelly, L. *Chem. Biol.* **1999,** *6,* 153–165. doi:10.1016/S1074-5521(99)89007-3
132. Lönnberg, T. *Nucleosides, Nucleotides Nucleic Acids* **2006,** *25,* 315–323. doi:10.1080/15257770500544537
133. Lönnberg, T. A. *ChemBioChem* **2012,** *13,* 2690–2700. doi:10.1002/cbic.201200629
134. Wolfenden, R.; Ridgway, C.; Young, G. *J. Am. Chem. Soc.* **1998,** *120,* 833–834. doi:10.1021/ja9733604
135. Abe, Y. S.; Sasaki, S. *Bioorg. Med. Chem.* **2016,** *24,* 910–914. doi:10.1016/j.bmc.2016.01.016
136. Kamerlin, S. C. L.; Williams, N. H.; Warshel, A. *J. Org. Chem.* **2008,** *73,* 6960–6969. doi:10.1021/jo801207q
137. Wan, S.-H.; Liang, F.; Xiong, X.-Q.; Yang, L.; Wu, X.-J.; Wang, P.; Zhou, X.; Wu, C.-T. *Bioorg. Med. Chem. Lett.* **2006,** *16,* 2804–2806. doi:10.1016/j.bmcl.2006.01.106
138. Wei, L.; Shao, Y.; Zhou, M.; Hu, H.-W.; Lu, G.-Y. *Org. Biomol. Chem.* **2012,** *10,* 8484–8492. doi:10.1039/c2ob25743b
139. Li, Z.-F.; Chen, H.-L.; Zhang, L.-J.; Lu, Z.-L. *Bioorg. Med. Chem. Lett.* **2012,** *22,* 2303–2307. doi:10.1016/j.bmcl.2012.01.075
140. Ullrich, S.; Nazir, Z.; Büsing, A.; Steffer, U.; Wirth, D.; Bats, J. W.; Dürner, G.; Göbel, M. W. *ChemBioChem* **2011,** *12,* 1223–1229. doi:10.1002/cbic.201100022
141. Sheng, X.; Lu, X.-M.; Zhang, J.-J.; Chen, Y.-T.; Lu, G.-Y.; Shao, Y.; Liu, F.; Xu, Q. *J. Org. Chem.* **2007,** *72,* 1799–1802. doi:10.1021/jo0624041
142. Shao, Y.; Ding, Y.; Jia, Z.-L.; Lu, X.-M.; Ke, Z.-H.; Xu, W.-H.; Lu, G.-Y. *Bioorg. Med. Chem.* **2009,** *17,* 4274–4279. doi:10.1016/j.bmc.2009.05.044
143. Bosch, S.; Comba, P.; Gahan, L. R.; Schenk, G. *Inorg. Chem.* **2014,** *53,* 9036–9051. doi:10.1021/ic5009945
144. Camargo, M. A.; Neves, A.; Bortoluzzi, A. J.; Szpoganicz, B.; Fischer, F. L.; Terenzi, H.; Serra, O. A.; Santos, V. G.; Vaz, B. G.; Eberlin, M. N. *Inorg. Chem.* **2010,** *49,* 6013–6025. doi:10.1021/ic100549u
145. Feng, G.; Natale, D.; Prabaharan, R.; Mareque-Rivas, J. C.; Williams, N. H. *Angew. Chem., Int. Ed.* **2006,** *45,* 7056–7059. doi:10.1002/anie.200602532
146. Breslow, R.; Huang, D. L. *Proc. Natl. Acad. Sci. U. S. A.* **1991,** *88,* 4080–4083. doi:10.1073/pnas.88.10.4080
147. Morrow, J. R. *Adv. Inorg. Biochem.* **1994,** *9,* 41–74.
148. Kuusela, S.; Lönnberg, H. *Met. Ions Biol. Syst.* **1996,** *32,* 271–300.
149. Korhonen, H.; Koivusalo, T.; Toivola, S.; Mikkola, S. *Org. Biomol. Chem.* **2013,** *11,* 8324–8339. doi:10.1039/c3ob41554f
150. Morrow, J. R.; Trogler, W. C. *Inorg. Chem.* **1988,** *27,* 3387–3394. doi:10.1021/ic00292a025
151. Jenkins, L. A.; Bashkin, J. K.; Pennock, J. D.; Florián, J.; Warshel, A. *Inorg. Chem.* **1999,** *38,* 3215–3222. doi:10.1021/ic990228r
152. Liu, S.; Hamilton, A. D. *Tetrahedron Lett.* **1997,** *38,* 1107–1110. doi:10.1016/S0040-4039(96)02519-1
153. Liu, S.; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* **1997,** *7,* 1779–1784. doi:10.1016/S0960-894X(97)00303-X
154. Koike, T.; Kimura, E. *J. Am. Chem. Soc.* **1991,** *113,* 8935–8941. doi:10.1021/ja00023a048
155. Tjioe, L.; Joshi, T.; Brugger, J.; Graham, B.; Spiccia, L. *Inorg. Chem.* **2011,** *50,* 621–625. doi:10.1021/ic1018136
156. Cacciapaglia, R.; Casnati, A.; Mandolini, L.; Reinhoudt, D. N.; Salvio, R.; Sartori, A.; Ungaro, R. *J. Am. Chem. Soc.* **2006,** *128,* 12322–12330. doi:10.1021/ja0632106
157. Morrow, J. R.; Buttrey, L. A.; Berback, K. A. *Inorg. Chem.* **1992,** *31,* 16–20. doi:10.1021/ic00027a005
158. Matsumura, K.; Komiyama, M. *J. Biochem.* **1997,** *122,* 387–394. doi:10.1093/oxfordjournals.jbchem.a021765
159. Bonfá, L.; Gatos, M.; Mancin, F.; Tecilla, P.; Tonellato, U. *Inorg. Chem.* **2003,** *42,* 3943–3949. doi:10.1021/ic034139x
160. Bonomi, R.; Saielli, G.; Scrimin, P.; Mancin, F. *Supramol. Chem.* **2013,** *25,* 665–671. doi:10.1080/10610278.2013.830724
161. Chin, K. O. A.; Morrow, J. R. *Inorg. Chem.* **1994,** *33,* 5036–5041. doi:10.1021/ic00100a031

162. Fanning, A.-M.; Plush, S. E.; Gunnlaugsson, T. *Org. Biomol. Chem.* **2015,** *13,* 5804–5816. doi:10.1039/C4OB02384F
163. Chapman, W. H., Jr.; Breslow, R. *J. Am. Chem. Soc.* **1995,** *117,* 5462–5469. doi:10.1021/ja00125a005
164. Wall, M.; Hynes, R. C.; Chin, J. *Angew. Chem., Int. Ed. Engl.* **1993,** *32,* 1633–1635. doi:10.1002/anie.199316331
165. Liu, S.; Hamilton, A. D. *Chem. Commun.* **1999,** 587–588. doi:10.1039/a808195f
166. Iranzo, O.; Elmer, T.; Richard, J. P.; Morrow, J. R. *Inorg. Chem.* **2003,** *42,* 7737–7746. doi:10.1021/ic030131b
167. Feng, G.; Mareque-Rivas, J. C.; Williams, N. H. *Chem. Commun.* **2006,** 1845–1847. doi:10.1039/b514328d
168. Feng, G.; Mareque-Rivas, J. C.; Torres Martín de Rosales, R.; Williams, N. H. *J. Am. Chem. Soc.* **2005,** *127,* 13470–13471. doi:10.1021/ja054003t
169. Bonomi, R.; Saielli, G.; Tonellato, U.; Scrimin, P.; Mancin, F. *J. Am. Chem. Soc.* **2009,** *131,* 11278–11279. doi:10.1021/ja9033236
170. Tjioe, L.; Joshi, T.; Forsyth, C. M.; Moubaraki, B.; Murray, K. S.; Brugger, J.; Graham, B.; Spiccia, L. *Inorg. Chem.* **2012,** *51,* 939–953. doi:10.1021/ic2019814
171. Salvio, R.; Volpi, S.; Cacciapaglia, R.; Casnati, A.; Mandolini, L.; Sansone, F. *J. Org. Chem.* **2015,** *80,* 5887–5893. doi:10.1021/acs.joc.5b00965
172. Livieri, M.; Mancin, F.; Saielli, G.; Chin, J.; Tonellato, U. *Chem. – Eur. J.* **2007,** *13,* 2246–2256. doi:10.1002/chem.200600672
173. He, J.; Sun, J.; Mao, Z.-W.; Ji, L.-N.; Sun, H. *J. Inorg. Biochem.* **2009,** *103,* 851–858. doi:10.1016/j.jinorgbio.2009.02.010
174. O'Donoghue, A.; Pyun, S. Y.; Yang, M.-Y.; Morrow, J. R.; Richard, J. P. *J. Am. Chem. Soc.* **2006,** *128,* 1615–1621. doi:10.1021/ja056167f
175. Linjalahti, H.; Feng, G.; Mareque-Rivas, J. C.; Mikkola, S.; Williams, N. H. *J. Am. Chem. Soc.* **2008,** *130,* 4232–4233. doi:10.1021/ja711347w
176. Korhonen, H.; Mikkola, S.; Williams, N. H. *Chem. – Eur. J.* **2012,** *18,* 659–670. doi:10.1002/chem.201100721
177. Neverov, A. A.; Lu, Z.-L.; Maxwell, C. I.; Mohamed, M. F.; White, C. J.; Tsang, J. S. W.; Brown, R. S. *J. Am. Chem. Soc.* **2006,** *128,* 16398–16405. doi:10.1021/ja0651714
178. Iranzo, O.; Kovalevsky, A. Y.; Morrow, J. R.; Richard, J. P. *J. Am. Chem. Soc.* **2003,** *125,* 1988–1993. doi:10.1021/ja027728v
179. Yang, M.-Y.; Iranzo, O.; Richard, J. P.; Morrow, J. R. *J. Am. Chem. Soc.* **2005,** *127,* 1064–1065. doi:10.1021/ja044617i
180. Ora, M.; Peltomäki, M.; Oivanen, M.; Lönnberg, H. *J. Org. Chem.* **1998,** *63,* 2939–2947. doi:10.1021/jo972112n
181. Oivanen, M.; Lönnberg, H. *Acta Chem. Scand.* **1991,** *45,* 968–971. doi:10.3891/acta.chem.scand.45-0968
182. Montagner, D.; Gandin, V.; Marzano, C.; Erxleben, A. *Eur. J. Inorg. Chem.* **2014,** 4084–4092. doi:10.1002/ejic.201402319
183. Bím, D.; Svobodová, E.; Eigner, V.; Rulíšek, L.; Hodačová, J. *Eur. J. Chem.* **2016,** *22,* 10426–10437. doi:10.1002/chem.201601175
184. Mathews, C. K.; van Holde, K. E.; Ahern, K. G. *Biochemistry,* 3rd ed.; Prentice Hall: New Jersey, 1999.
185. Rossiter, C. S.; Mathews, R. A.; Morrow, J. R. *Inorg. Chem.* **2005,** *44,* 9397–9404. doi:10.1021/ic050892+
186. Linkletter, B.; Chin, J. *Angew. Chem., Int. Ed. Engl.* **1995,** *34,* 472–474. doi:10.1002/anie.199504721
187. Tirel, E. Y.; Williams, N. H. *Chem. – Eur. J.* **2015,** *21,* 7053–7056. doi:10.1002/chem.201500619
188. Kuusela, S.; Lönnberg, H. *J. Phys. Org. Chem.* **1993,** *6,* 347–356. doi:10.1002/poc.610060606
189. Korhonen, H.; Williams, N. H.; Mikkola, S. *J. Phys. Org. Chem.* **2012,** *26,* 182–186. doi:10.1002/poc.2942
190. Mikkola, S.; Stenman, E.; Nurmi, K.; Yousefi-Salakdeh, E.; Strömberg, R.; Lönnberg, H. *J. Chem. Soc., Perkin Trans. 2* **1999,** 1619–1626. doi:10.1039/a903691a
191. Kivioja, T.; Mikkola, S. *Recent Res. Dev. Org. Bioorg. Chem.* **2002,** *5,* 69.
192. Williams, N. H.; Takasaki, B.; Wall, M.; Chin, J. *Acc. Chem. Res.* **1999,** *32,* 485–493. doi:10.1021/ar9500877
193. Williams, N. H.; Cheung, W.; Chin, J. *J. Am. Chem. Soc.* **1998,** *120,* 8079–8087. doi:10.1021/ja980660s
194. Kuusela, S.; Lönnberg, H. *Nucleosides Nucleotides* **1996,** *15,* 1669–1678. doi:10.1080/07328319608002466
195. Wang, Q.; Lönnberg, H. *J. Am. Chem. Soc.* **2006,** *128,* 10716–10728. doi:10.1021/ja058806s
196. Wang, Q.; Leino, E.; Jancsó, A.; Szilágyi, I.; Gajda, T.; Hietamäki, E.; Lönnberg, H. *ChemBioChem* **2008,** *9,* 1739–1748. doi:10.1002/cbic.200800095
197. Zhang, S.; Gu, H.; Chen, H.; Strong, E.; Ollie, E. W.; Kellerman, D.; Liang, D.; Miyagi, M.; Anderson, V. E.; Piccirilli, J. A.; York, D. M.; Harris, M. E. *Chem. Commun.* **2016,** *52,* 4462–4465. doi:10.1039/C5CC10212J
198. Humphry, T.; Iyer, S.; Iranzo, O.; Morrow, J. R.; Richard, J. P.; Paneth, P.; Hengge, A. C. *J. Am. Chem. Soc.* **2008,** *130,* 17858–17866. doi:10.1021/ja8059864
199. Deal, K. A.; Hengge, A. C.; Burstyn, J. N. *J. Am. Chem. Soc.* **1996,** *118,* 1713–1718. doi:10.1021/ja952306p
200. Gao, H.; Ke, Z.; DeYonker, N. J.; Wang, J.; Xu, H.; Mao, Z.-W.; Phillips, D. L.; Zhao, C. *J. Am. Chem. Soc.* **2011,** *133,* 2904–2915. doi:10.1021/ja106456u
201. Yang, M.-Y.; Richard, J. P.; Morrow, J. R. *Chem. Commun.* **2003,** 2832–2833. doi:10.1039/b308644e
202. Liu, C. T.; Neverov, A. A.; Brown, R. S. *J. Am. Chem. Soc.* **2008,** *130,* 16711–16720. doi:10.1021/ja806462x
203. Bunn, S. E.; Liu, C. T.; Lu, Z.-L.; Neverov, A. A.; Brown, R. S. *J. Am. Chem. Soc.* **2007,** *129,* 16238–16248. doi:10.1021/ja076847d
204. Neverov, A. A.; Liu, C. T.; Bunn, S. E.; Edwards, D.; White, C. J.; Melnykchuk, S. A.; Brown, R. S. *J. Am. Chem. Soc.* **2008,** *130,* 6639–6649. doi:10.1021/ja8006963
205. Mohamed, M. F.; Neverov, A. A.; Brown, R. S. *Inorg. Chem.* **2009,** *48,* 11425–11433. doi:10.1021/ic9015965
206. Song, Y.; Zan, J.; Yan, H.; Lu, Z.-L.; Wang, R. *Org. Biomol. Chem.* **2012,** *10,* 7714–7720. doi:10.1039/c2ob25624j
207. Lu, Z.-L.; Liu, C. T.; Neverov, A. A.; Brown, R. S. *J. Am. Chem. Soc.* **2007,** *129,* 11642–11652. doi:10.1021/ja073780l
208. Corona-Martínez, D. O.; Gomez-Tagle, P.; Yatsimirsky, A. K. *J. Org. Chem.* **2012,** *77,* 9110–9119. doi:10.1021/jo301649u
209. Tirel, E. Y.; Bellamy, Z.; Adams, H.; Lebrun, V.; Duarte, F.; Williams, N. H. *Angew. Chem., Int. Ed.* **2014,** *53,* 8246–8250. doi:10.1002/anie.201400335
210. Fan, Y. B.; Gao, Y. Q. *Acta Phys.-Chim. Sin.* **2010,** *26,* 1034–1042.
211. Zhang, X.; Zhu, Y.; Gao, H.; Zhao, C. *Inorg. Chem.* **2014,** *53,* 11903–11912. doi:10.1021/ic501084a
212. Mancin, F.; Prins, L. J.; Pengo, P.; Pasquato, L.; Tecilla, P.; Scrimin, P. *Molecules* **2016,** *21,* No. 1014. doi:10.3390/molecules21081014
213. Ghidini, A.; Murtola, M.; Strömberg, R.; Stultz, E., Eds.; Wiley: Chichester, UK, 2015; pp 158–167.
214. Shigi, N.; Sumaoka, J.; Komiyama, M. *Molecules* **2017,** *22,* No. 1586. doi:10.3390/molecules22101586

215.Komiyama, M. *Chem. Lett.* **2016,** *45,* 1347–1355. doi:10.1246/cl.160786
216.Katada, H.; Komiyama, M. *ChemBioChem* **2009,** *10,* 1279–1288. doi:10.1002/cbic.200900040

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Mechanochemistry of nucleosides, nucleotides and related materials

Olga Eguaogie, Joseph S. Vyle*, Patrick F. Conlon, Manuela A. Gîlea and Yipei Liang

Review

Open Access

Address:
School of Chemistry and Chemical Engineering, Queen's University Belfast, David Keir Building, Stranmillis Road, Belfast BT9 5AG, UK

Email:
Joseph S. Vyle* - j.vyle@qub.ac.uk

* Corresponding author

Keywords:
DNA; green chemistry; mechanochemistry; nucleoside; nucleotide

Beilstein J. Org. Chem. **2018,** *14,* 955–970.
doi:10.3762/bjoc.14.81

Received: 31 December 2017
Accepted: 20 April 2018
Published: 27 April 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

The application of mechanical force to induce the formation and cleavage of covalent bonds is a rapidly developing field within organic chemistry which has particular value in reducing or eliminating solvent usage, enhancing reaction rates and also in enabling the preparation of products which are otherwise inaccessible under solution-phase conditions. Mechanochemistry has also found recent attention in materials chemistry and API formulation during which rearrangement of non-covalent interactions give rise to functional products. However, this has been known to nucleic acids science almost since its inception in the late nineteenth century when Miescher exploited grinding to facilitate disaggregation of DNA from tightly bound proteins through selective denaturation of the latter. Despite the wide application of ball milling to amino acid chemistry, there have been limited reports of mechanochemical transformations involving nucleoside or nucleotide substrates on preparative scales. A survey of these reactions is provided, the majority of which have used a mixer ball mill and display an almost universal requirement for liquid to be present within the grinding vessel. Mechanochemistry of charged nucleotide substrates, in particular, provides considerable benefits both in terms of efficiency (reducing total processing times from weeks to hours) and by minimising exposure to aqueous conditions, access to previously elusive materials. In the absence of large quantities of solvent and heating, side-reactions can be reduced or eliminated. The central contribution of mechanochemistry (and specifically, ball milling) to the isolation of biologically active materials derived from nuclei by grinding will also be outlined. Finally non-covalent associative processes involving nucleic acids and related materials using mechanochemistry will be described: specifically, solid solutions, cocrystals, polymorph transitions, carbon nanotube dissolution and inclusion complex formation.

Introduction

Several definitions of mechanochemistry have been attempted since Ostwald included it as one of four taxa along with thermochemistry, electrochemistry and photochemistry [1]. A general definition commonly cited is that developed by The International Union of Pure and Applied Chemistry (IUPAC) to encompass both the chemical and physical effects of shearing,

stretching or grinding polymeric materials: "[a mechano-chemical reaction is one] induced by the direct absorption of mechanical energy" [2]. The etymology and early history of this field have been reviewed comprehensively by Takacs [1]. Several recent reviews discuss both general aspects of mechanochemistry [3,4] as well as more focussed elements of the subject relevant to the current work including applications in organic synthesis [5-8], green chemistry [9], API formulation [10] and coordination/materials chemistry [11,12]. Some aspects of the current work have also been reviewed recently [13]. However, the impact of mechanochemistry upon biological chemistry and specifically the selective degradation of biopolymers which enables biochemically active materials to be isolated from cell grindates – most notably in Buchner's laboratory [14] – appears not to have been considered.

A recent tutorial review by Andersen and Mack [15] augments an earlier introduction describing both the parameters used to define such chemistry and also how this information is conveyed in synthetic schemes [16]. Stolle has written a comprehensive treatise on the chemical, technological and process parameters which influence the outcome of a ball mill reaction [17]. In this review we have also adopted Hanusa's formalism which distinguishes ball milling from other forms of mechanochemistry [18].

Perhaps most critical to the recent interest in this field has been the ability to deliver consistent and reproducible levels of mechanical energy using commercially-available equipment which, for reactions of nucleosides and related materials has most commonly been the mixer ball mill (MBM – e.g., Figure 1a). Using a MBM, high energy collisions between reactants and one or more balls within a closed vessel (jar) are induced by vibrating the jar through a limited arc (ca. 0.5°) within one plane at up to 60 Hz (more typically 30 Hz). In its single-armed form, this is sometimes referred to as an amalgum mill. Alternatively, grinding actions have been provided using a mechanised mortar mill which mimics the action of hand grinding in a mortar and pestle (Figure 1b), an improvised attritor-type device (Figure 1c) or a planetary ball mill (not shown).

The amount of mechanical energy delivered to the reaction mixture via these collisions is a function of several engineering parameters including: the frequency of vibration; the degree of filling of the vessel (and its shape); the mass of the ball(s); and the hardness of the colliding materials. In order of descending hardness, zirconia, stainless steel, copper and PTFE have all been used to effect mechanochemical transformation of nucleoside or nucleotide substrates. During a study of amide coupling under ball-milling conditions, Lamaty and co-workers showed that deterioration of vessels and balls by physical abrasion and/or chemical leaching gave products in which (depending upon the nature of the jar) iron, chromium, zirconia or PTFE were detected [20]. This has influenced the choice of vessel for nucleoside and nucleotide chemistry as, although considerably cheaper, leaching of iron from stainless steel vessels in the presence of sulfur-containing materials [21] has been found to inhibit the preparation of thionucleoside [22] or thionucleotide [23] analogues. Although grinding using PTFE components delivers less energy due to the material's elasticity and low density (2.1 g cm^{-3}) compared with stainless steel (7.8 g cm^{-3}) or

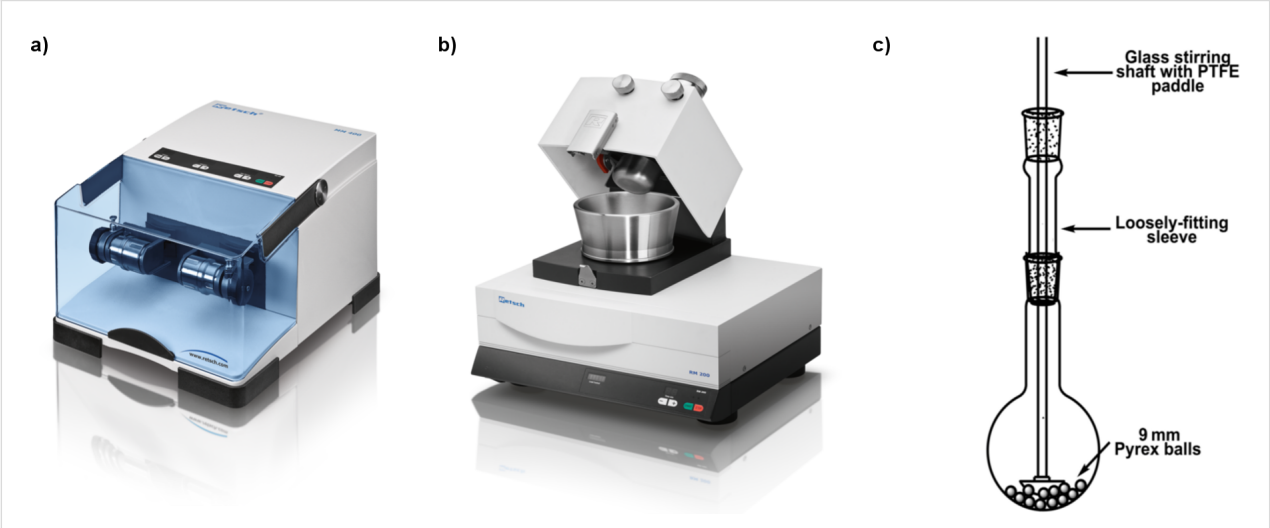


Figure 1: Examples of equipment used to perform mechanochemistry on nucleoside and nucleotide substrates (not to scale). a) Mixer ball mill; b) mortar grinder; c) improvised attritor [19]. Figures a) and b) are reused with the permission of Retsch (https://www.retsch.com); c) is adapted with permission from [19], copyright 2006 American Chemical Society.

zirconia (5.9 g cm^{-3}), PTFE may be required for the preparation of pharmaceutical grade materials which are subject to regulatory approval.

Theoretical models of mechanochemical bond activation are mainly based upon examination and/or modelling the behaviour of single molecules under tension in an atomic force microscope [24-26] and have included relating traditional Arrhenius reaction parameters to applied forces [27,28]. However, early models of macroscopic scale reactivity in the solid-state (such as the formation of a short-lived plasma phase [29]) do not account for observations on the comminution of organic reactants during milling such as changes to the physical form of the mixture (including zones of liquefaction [30] and cohesive states [31]) which are correlated with the progress of the reaction including induction periods of up to 40 minutes [31,32]. Rate enhancements may thus be achieved from very high localised reactant concentrations within which developing intermolecular and intramolecular interactions are formed that can lead to reaction of a nucleoside or nucleotide substrate which would be disfavoured in solution. Although bond disruption via ultrasound-induced cavitation can be considered within the purview of mechanochemistry [33,34], this review will be restricted to the delivery of mechanochemical energy on a macroscopic scale by vibration, grinding and/or crushing actions. Furthermore, the term grinding will be applied throughout even though kneading (often referred to as solvent-drop grinding) is more accurate to describe the process of grinding or milling mixtures of solids and liquids [35]. To date, all but one chemical transformation of solid nucleoside or nucleotide substrates have been performed in the presence of liquids. These may originate either from the use of reagents which are liquids or low-melting solids (which liquify upon grinding) or from the addition of stoichiometric quantities of molecular solvents (or ionic liquids). During subsequent discussions, liquid-assisted grinding (LAG) is used to describe only the latter case.

The minimal level of solvent requirement is particularly advantageous in the context of charged nucleotide substrates as considerable cost, time and energy savings can be gained in the absence of arduous ion-exchange and drying processes required to render these materials soluble in organic solvents. Likewise, significant reductions in solvent processing (especially if these are high boiling and often toxic and/or carcinogenic) is an attractive green chemistry target. In this context, Thorwith et al. compared the amount of energy required to effect permanganate-mediated oxidative self-coupling of *p*-toluidine using different energy inputs. Ball milling was significantly more efficient (up to an order of magnitude) than conventional heating, microwave or ultrasound inputs [36]. The reduction in both solvent and energy input are particularly relevant in fine chemical manufacturing processes which typically have very high E-factors and low energy efficiency [37]. Although mechanochemistry was not involved in redesigning the synthesis of the antiviral prodrug ganciclovir (Figure 2), the high levels of involatile solvent usage typically employed in the solution-based synthesis of such compounds can be gauged by the ability of Roche to eliminate 1.12 million kilograms of solvent per annum [38].

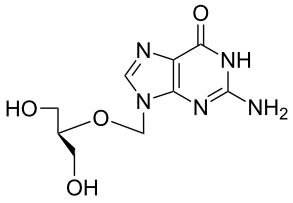


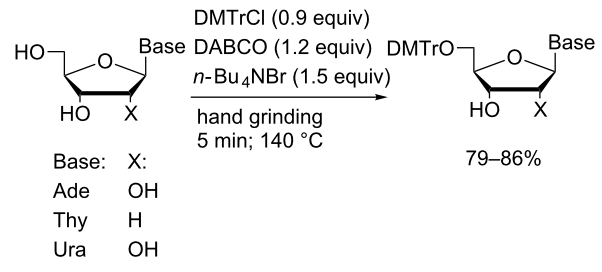
Figure 2: Ganciclovir.

Review

Mechanochemical transformations of nucleosides and related materials involving covalent bonds

Reactions of nucleoside sugar and nucleobase moieties

An early example of the application of mechanochemistry for nucleoside derivatisation was reported by Khalafi-Nezhad and Mokhtari who effected regioselective 5′-protection of ribonucleosides and thymidine using a mortar and pestle with trityl-, monomethoxytrityl- or dimethoxytrityl chloride (Scheme 1) [39].



Scheme 1: Nucleoside tritylation effected by hand grinding in a heated mortar and pestle.

A variety of temperatures and either inorganic or low-melting organic bases were surveyed. Optimal yields were achieved at 140 °C using DABCO by hand-grinding the reaction mixture in molten tetra-*n*-butylammonium bromide (TBAB) for five minutes. In the presence of excess nucleoside (1.1 equiv), the corresponding 5′-trityl ethers of uridine, adenosine or thymi-

dine were isolated in yields up to 86%. Reactions of guanosine or cytidine under these conditions gave rise to mixtures of products from which the corresponding tritylated products could not be isolated.

Subsequently, Patil and Kartha described the gram-scale preparation of 5′-tritylated uridine derivatives in a planetary ball mill (using a steel vessel and balls) in the absence of TBAB [40]. Following extended grinding (600 rpm for 15 hours) of the nucleoside in the presence of excess DABCO and either TrCl or DMTrCl, the products were recovered in 44% and 43% yields, respectively.

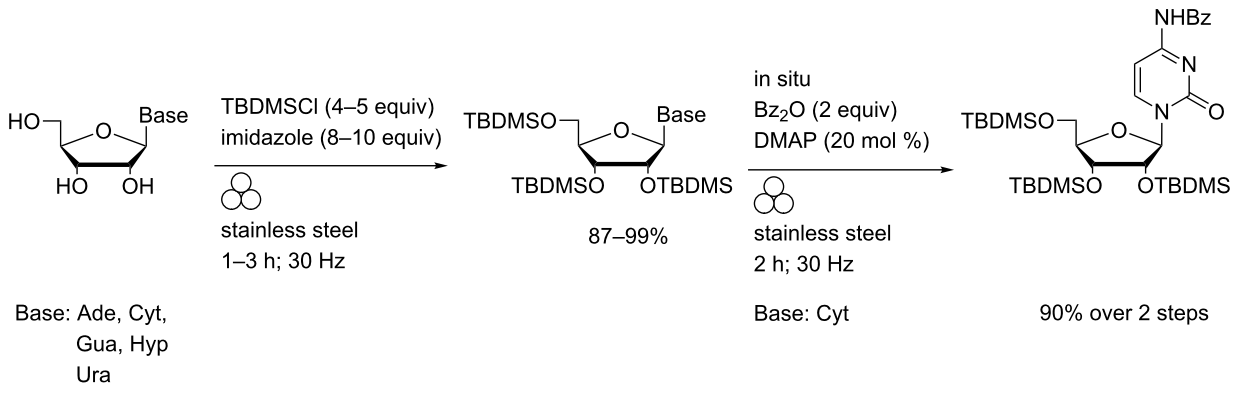
Under solvent-free Corey conditions, rapid and chemoselective persilylation of ribonucleoside hydroxy functions was effected in a mixer ball mill (Scheme 2) [41].

Complete consumption of starting materials was observed within one to three hours and only in the case of adenosine was any (minor) side-product formation found. In all cases, facile purification using a scrubber column enabled pure 2′,3′,5′-tri-*O*-TBDMS-protected nucleosides to be isolated in 87–99% yields. In situ benzoylation of the persilylated cytosine was also effected following addition of benzoic anhydride and catalytic DMAP to the crude reaction mixture and extending the milling time. Quantitative silylation of 5′-*O*-dimethoxytritylthymidine under these conditions was also reported.

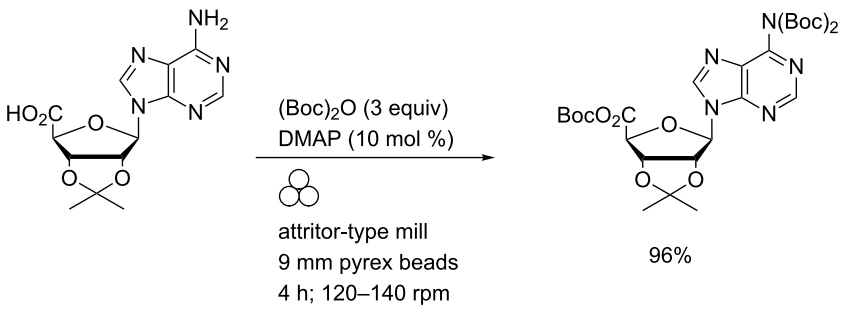
Prompted by the insolubility of adenosine 5′-carboxylic acid derivatives, Sikchi and Hultin contrived an attritor-type mill (Figure 1c) to facilitate the use of neat reagents and vent CO_2 (closed vessels were reported to break) [19]. Efficient gram-scale Boc protection of amine and carboxylic acid functions was thereby effected (Scheme 3).

This chemistry was further applied to the derivatisation of the exocyclic amino functions of hydroxy-protected adenosine and cytidine derivatives (and the corresponding 2′-deoxynucleosides). The majority of these reactions proceeded in excellent yields (90–99%) over one to six hours. In contrast, guanine-derived (deoxy)nucleosides generally required longer to achieve complete reaction and yielded the corresponding O^6,N^2,N^2-tri-Boc derivatives with variable recoveries (25–70%).

The scope of this reaction was extended to unprotected nucleosides by effecting a one-pot, two-step reaction sequence



Scheme 2: Persilylation of ribonucleoside hydroxy groups (and in situ acylation of cytidine) in a MBM.



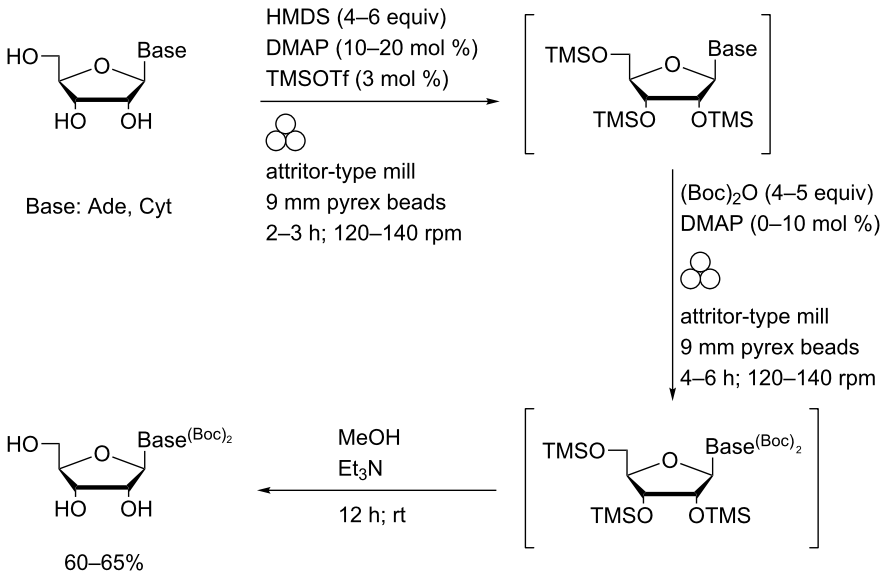
Scheme 3: Nucleoside amine and carboxylic acid Boc protection using an improvised attritor-type mill.

(Scheme 4). Initial transient silylation and subsequent Boc-protection were both performed in the absence of solvent under mechanochemical conditions. In situ methanolysis of the TMS ethers yielded the corresponding base-protected nucleosides.

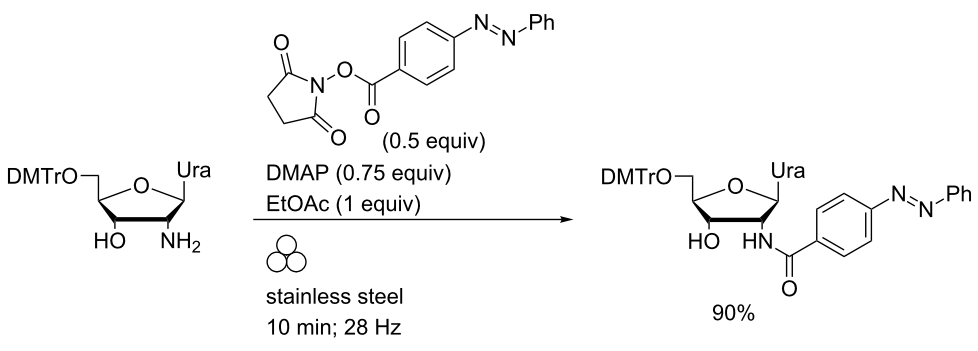
The opaque nature of typical reaction vessels used in ball milling has enabled cleaner reactions of nucleoside analogues with photoreactive materials. Thus, liquid-assisted grinding of the *N*-hydroxysuccinimidyl esters of *o*-, *m*- or *p*-phenylazobenzoic acids with excess D-threoninol or of the *para* isomer with an aminonucleoside in the presence of DMAP and ethyl acetate engendered chemoselective *N*-acylation (Scheme 5) [42]. In the absence of light, azobenzene derivatives were isolated as the pure *E*-isomers.

In their original report, Sharpless and co-workers described the use of copper turnings to promote a regioselective azide–alkyne [3 + 2]-cycloaddition ("click") reaction over 24 hours [43]. High-speed ball milling using a custom-made copper vial and copper ball enabled efficient reaction between propyne-derivatised photoswitches and an azidodeoxynucleoside click partner (Scheme 6) [44].

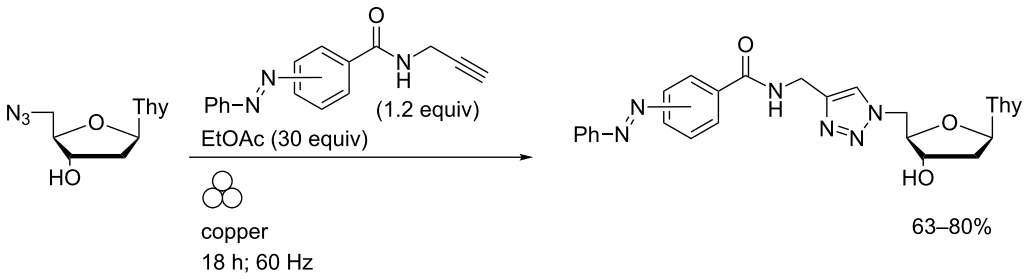
In contrast to the solution-phase (Cu(I)-promoted) reactions, no contamination of the ball milled products by copper salts was found. In an attempt to expedite the LAG reaction, millimol-scale reactions between the *p*-azobenzene-appended alkyne and 5′-azido-5′-deoxythymidine were attempted in a more capacious copper vessel with a 15 mm diameter zirconia ball (Figure 3). Clean and complete click reactions were achieved within 40 minutes at 25 Hz in the presence of ethyl acetate although the integrity of the vessel was compromised and significant levels of metallic copper were removed from the walls during work-up.



Scheme 4: Nucleobase Boc protection via transient silylation using an improvised attritor-type mill.



Scheme 5: Chemoselective *N*-acylation of an aminonucleoside using LAG in a MBM.



Scheme 6: Azide–alkyne cycloaddition reactions performed in a copper vessel in a MBM.

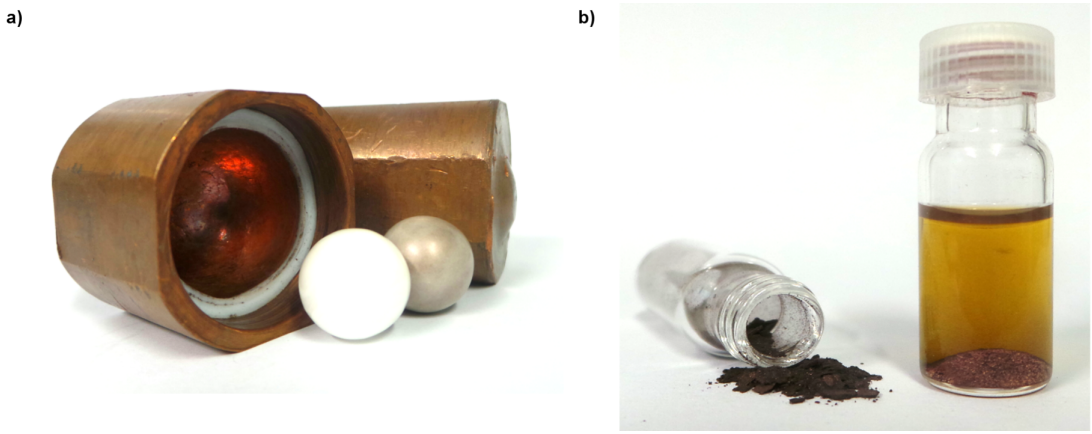
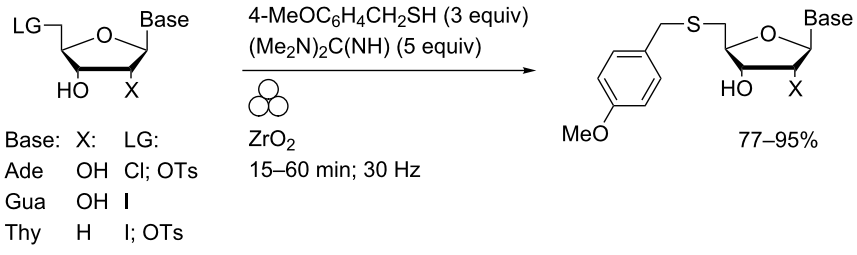


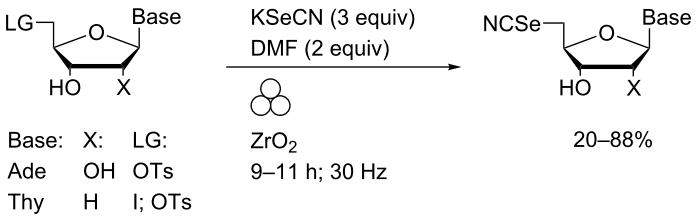
Figure 3: a) Custom-machined copper vessel and zirconia balls used to perform CuAAC reactions (showing: upper half of vessel with PTFE insert (front), pristine ZrO_2 ball, used ZrO_2 ball and lower half of vessel showing deformation of the metal). b) Crude solid ball mill click reaction mixture after removal from copper vessel (left) and during extraction of pure product with DMSO (right).

Expeditious displacement of tosylate or halides from 5′-derivatised nucleosides was achieved using chalogenate nucleophiles in a mixer ball mill using zirconia components [22]. Highly efficient transformations to the corresponding 4-methoxybenzyl thioethers were achieved in 15–60 minutes such that pure products could be isolated without the need for chromatography (Scheme 7). Of particular note was the absence of any observable intramolecular cyclisation of the unprotected purine nucleoside derivatives typical of solution-phase reactions using such substrates.

More variable yields were obtained using potassium selenocyanate which required grinding in the presence of DMF to promote the reaction with adenosine or thymidine derivatives (Scheme 8). No reaction of 5′-chloro-5′-deoxyadenosine was observed.



Scheme 7: Thiolate displacement reactions of nucleoside derivatives in a MBM.



Scheme 8: Selenocyanate displacement reactions of nucleoside derivatives in a MBM.

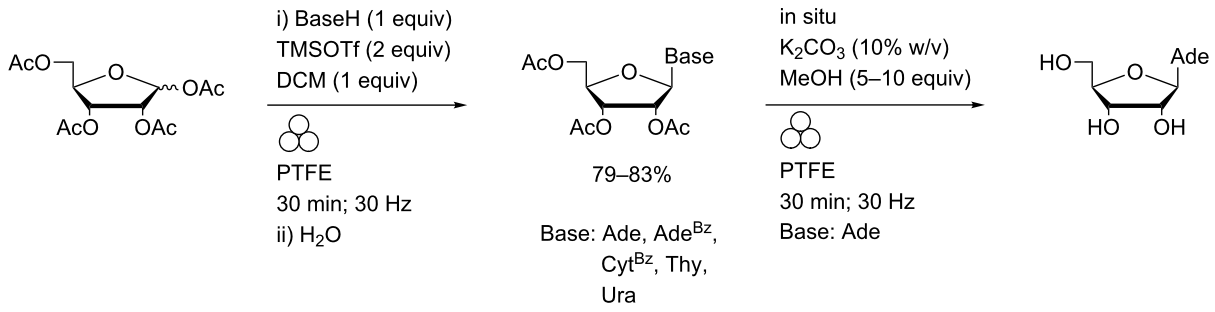
Under these conditions, cyclisation of 5′-tosyladenosine was inferred although rapid and clean reaction of 5′-iodo-5′-deoxyguanosine was apparent in the absence of added solvent – the product from this latter reaction rapidly decomposed during work-up in solution.

Regioselective and stereoselective glycosidation of adenine, N^6-benzoyladenine, N^4-benzoylcytosine, thymine and uracil to the corresponding β-N^9-purine or β-N^1-pyrimidine ribosides was achieved on gram scales under Vorbrüggen-type conditions using LAG (Scheme 9) [45].

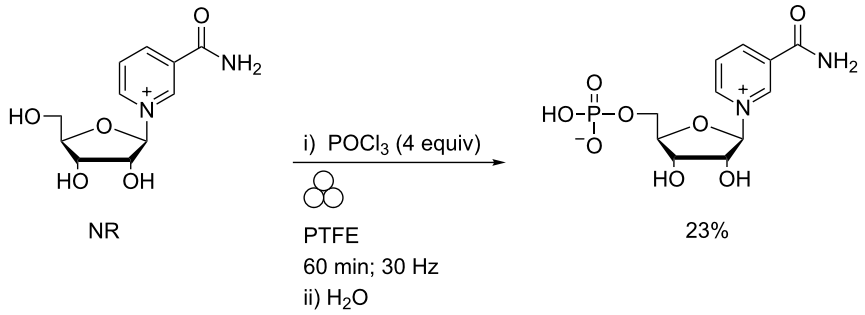
Yields were slightly enhanced following presilylation of the bases in solution prior to ball milling and under these conditions, the corresponding protected 6-chloropurine riboside could also be accessed. Multiple products were formed from N^2-isobutyrylguanine and hypoxanthine but cytosine remained untransformed. In situ deprotection of 2′,3′,5′-tri-*O*-acetyladenosine was also claimed. This chemistry has also been applied to the preparation of a library of ribosylated nicotinamide and nicotinic acid ester derivatives in a mortar grinder or using mixer or planetary ball mills [46,47]. Reaction scales up to 40 g were described and the conditions developed enabled exclusive formation of the β-anomer of nicotinamide riboside (NR) in the absence of toxic bromide salts.

Preparation and reactions of nucleotides and their analogues

Phosphorylation of NR on gram-scales using $POCl_3$ (e.g., Scheme 10), monoalkyl phosphorodichloridates or dialkyl phos-



Scheme 9: Nucleobase glycosidation reactions and subsequent deacetylation performed in a MBM.



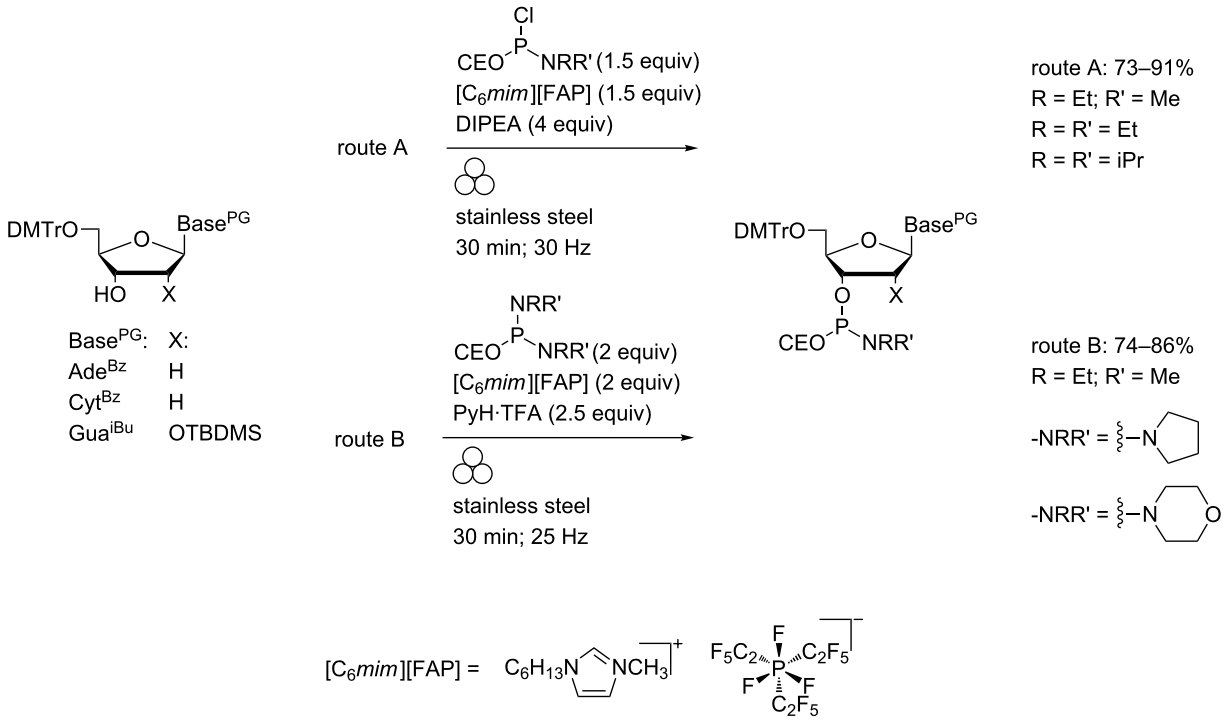
Scheme 10: Regioselective phosphorylation of nicotinamide riboside in a MBM.

phoromonochloridates in the absence of solvent has been reported [48].

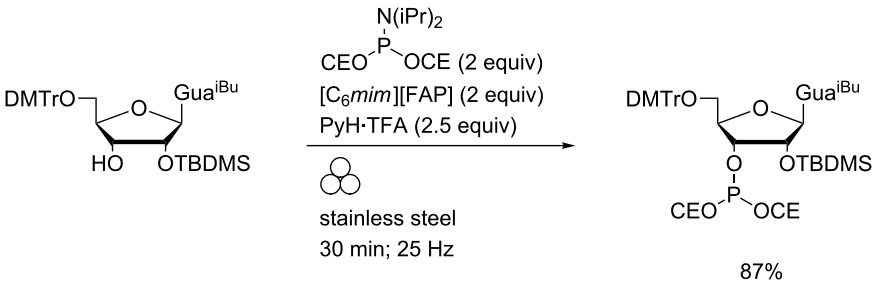
Migaud and co-workers prepared highly water-sensitive phosphitylating agents directly from PCl_3 in low viscosity ionic liquids derived from the tris(pentafluoroethyl)trifluorophosphate anion (e.g., [C_6*mim*][FAP]) and subsequently used the crude chlorophosphoramidite or phosphorodiamidite products to effect nucleoside phosphitylations using LAG (Scheme 11) [49,50].

In the absence of grinding, addition of a molecular cosolvent was required due to the low solubility of substrates in the ionic liquids (<10 mM) which rendered the phosphitylating agents prone to hydrolysis. Highly reactive phosphoramidite derivatives of low molecular weight amines could be isolated by this route (on 40–60 mg scales). Under the same conditions, coupling of bis(2-cyanoethyl)diisopropylaminophosphoramidite with a partially-protected guanosine derivative to the corresponding phosphite triester was also effected (Scheme 12).

Phosphate coupling using nucleoside phosphoromorpholidates is well established [51] but the reaction times are typically in the order of days. Recent developments in this field which yield pyrophosphate bonds more rapidly have been comprehensively reviewed by Peyrottes and co-workers [52] but in all cases, efficient coupling has been predicated on strictly controlling the water content of the reaction mixture. In contrast, LAG in the



Scheme 11: Preparation of nucleoside phosphoramidites in a MBM using ionic liquid-stabilised chlorophosphoramidites (route A) or phosphorodiamidites (route B).



Scheme 12: Preparation of a nucleoside phosphite triester using LAG in a MBM.

presence of water enabled the coupling of adenosine-5′-monophosphoromorpholidate with the sodium salts of 5′-phosphorylated nucleosides without any predrying and in the presence of acidic promoters and water gave complete reaction within 90 minutes (Scheme 13) [53].

In this original report, the preparation of nicotinamide adenine dinucleotide (NAD) and adenosine diphosphate ribose (ADPR) was also described. Subsequently, this methodology was applied to the preparation of a library of six ADPR carbonate derivatives in 23–68% yields (e.g., Scheme 14) and tested as sirtuin inhibitors [54].

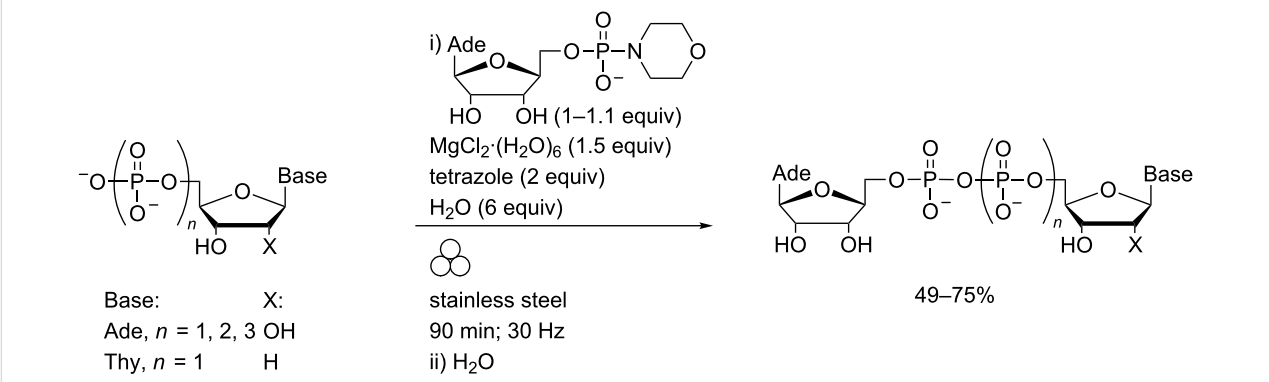
The efficiency of phosphate coupling under mechanochemical conditions was exploited to prepare pyrophosphorothiolate-linked dinucleoside cap analogues. Such materials had previously been inaccessible via this route due to the lability of intermediate phosphorothiolate monoesters under acidic conditions [55]. In contrast, the corresponding persilylated derivatives were found to be relatively stable under anhydrous conditions and could be readily prepared via Michaelis–Arbusov (M–A) chemistry (Scheme 15) [23,56].

Transfer of crude M–A reaction mixtures to a zirconia ball mill vessel and removal of volatiles enabled the concomitant partial hydrolytic desilylation of the monoester and phosphate coupling to AMP-morpholidate to be effected in one pot using LAG. Both 3′,5′- and 5′,5′- internucleoside linkages were prepared using this route.

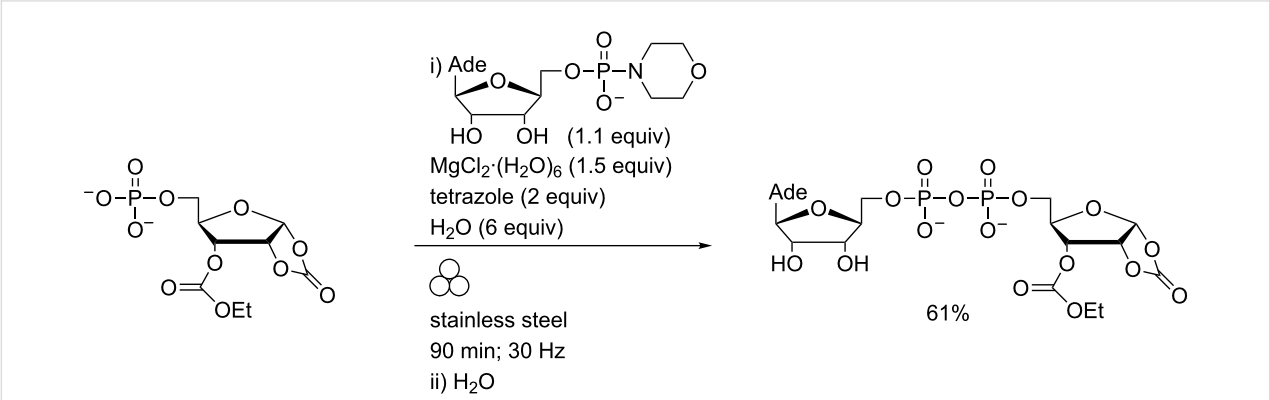
Mechanochemical transformations of nucleosides and related materials involving non-covalent bonds

Dissociative processes for DNA and RNA isolation

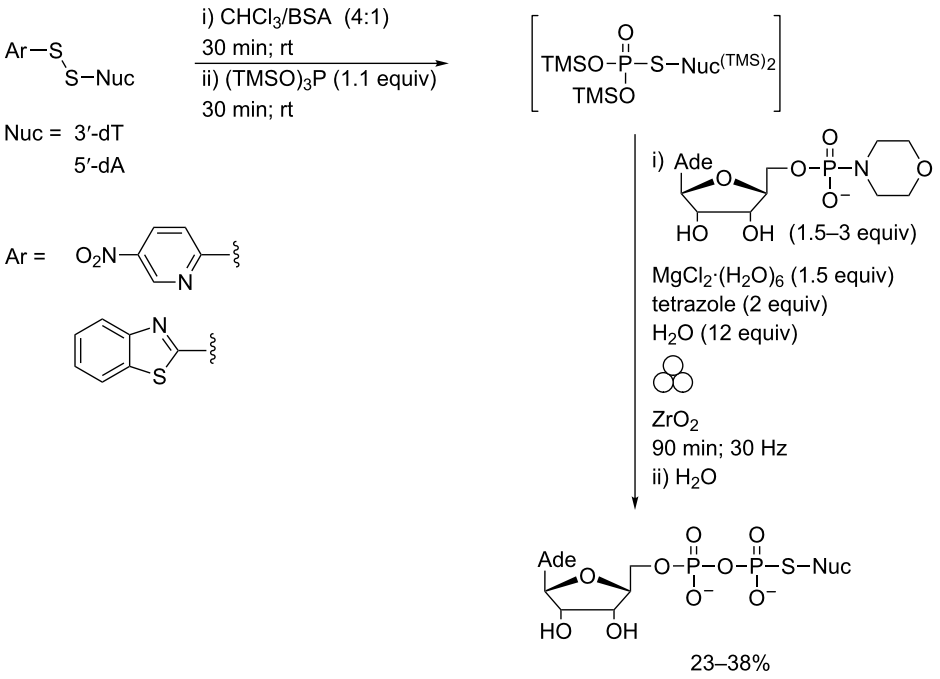
The lack of free-volume within double-stranded DNA at low hydration levels leads to limited ice formation even under cooling in liquid air [57]. In contrast, cold denaturation of globular proteins at such temperatures is almost ubiquitous [58]. Furthermore, large conformational reorientation of protein domains can be initiated at 30 pN compared with DNA which requires ca. 150 pN of highly directional force to bring about duplex melting [26]. Early recognition of these differences (even without a full understanding of their molecular origins) by pioneers in the field contributed to the development of DNA purification which featured mechanochemistry at low temperatures [59].



Scheme 13: Internucleoside phosphate coupling linkages in a MBM.



Scheme 14: Preparation of ADPR analogues using in a MBM.



Scheme 15: Synthesis of pyrophosphorothiolate-linked dinucleoside cap analogues in a MBM to effect hydrolytic desilylation and phosphate coupling.

Miescher reported grinding the solid residues from defatted salmon sperm heads with dilute HCl (0.5%) to effect such a separation and was able to report elemental analysis of nuclein with a phosphorus content (9.6%) close to that of the theoretical protein-free value [60,61].

Due to the accessibility of calf thymus and its high DNA content, much of the further work on large-scale extraction of "sodium nucleate" was performed using this tissue. Although details of the grinding actions employed were not always fully described, intensive mechanochemical processing (especially at low temperature) enabled the (frozen) fresh tissue to be powdered and in a subsequent step to bring about disaggregration of protein–DNA complexes at acidic pH. Early reports predominantly featured hand grinding in a mortar but could also include crushing with a glass rod and be supplemented by the use of a meat mincer and subsequently by an electrical blender [59,62-64]. The pure polymeric material isolated by these routes played a considerable role in revising Levene's tetranucleotide colloid hypothesis as he conceded in 1938 [65]. This culminated in Schwander and Signer isolating eight grams of pure material [63] from which high quality X-ray diffraction images (including "Photograph 51") [66] were obtained and its double-helical structure evinced [67].

Parallel to Miescher's work on salmon sperm, Kossel reported the isolation of ribonucleotide-derived material from yeast RNA in 1879 using mechanical disaggregation [68].

Altmann developed a more generalised method for isolating either RNA or DNA from a variety of tissues and organisms during which crude mixtures with protein were ground to a fine powder with 1:1 alcohol/6% HCl (aq) and subsequently triturated with pure alcohol and then ether [69]. Typically, RNA depolymerisation (especially when in contact with metal components) would be observed during these operations [70] although isolation of infectious viral RNA from frozen carcinoma tissue following grinding was reported in 1957. More recently, RNA was extracted from both Gram-negative and Gram-positive bacteria by hand grinding in a mortar with phenol under cooling with liquid nitrogen [71] .

The reproducibility of studies requiring arduous mechanochemical operations to be performed by hand lead to the rapid uptake of mechanisation for performing grinding actions by workers in this field. Behrens' procedure for isolating nuclei from calf heart included three separate mechanochemical operations first using a meat grinder, then a mortar and pestle and finally a mechanised ball mill in which a one litre flask containing 800 g of "glass pearls" was shaken at 3 Hz prior to subsequent trituration with benzene and carbon tetrachloride [72]. A subsequent development of this procedure included liquid-assisted grinding in a one litre porcelain jar which was rotated at 110 rpm for 24–48 hours with up to 1.4 kg of grinding stones (15–20 mm in diameter), 100 g of dried tissue powder and petroleum ether (200–450 mL) [73-75]. As early as 1903, low temperature grinding was applied to disrupting refractory mycobacteria

(using zirconia components under cooling in liquid air) [76] and subsequently mechanised using a steel ball mill (at −78 °C) [77] (e.g., Figure 4) [78].

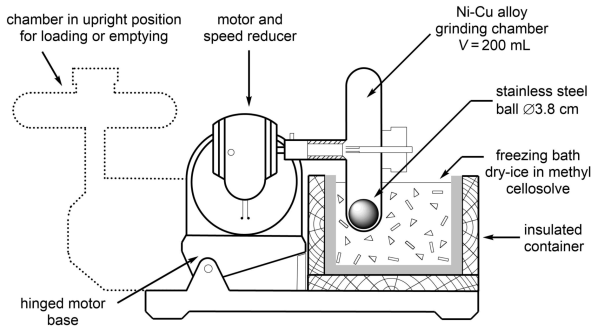


Figure 4: Early low temperature mechanised ball mill as described by Mudd et al. – adapted from reference [78].

Subsequently, this technology has been developed to allow larger scale ball milling of tissue samples (including at liquid nitrogen temperatures) and smaller scale "bead beating" in disposable plasticware. Depending upon the nature of the biological material and target sequence, grinding balls made of zirconia, garnet, glass or steel enable isolation and quantification of DNA or RNA from different sources [79]. Recently, a micro total analysis system was fabricated incorporating a 15 µL cell lysis chamber containing glass beads (30–50 µm) which were agitated using a membrane valve [80]. Within three minutes, almost complete disruption of Gram-positive bacteria was effected enabling downstream analysis by quantitative PCR.

Associative processes

Variable drug bioavailability associated with crystal and co-crystal polymorphism can be exacerbated if the solubility profiles of the API and coformer prevent solution-phase mixing. Under such circumstances, mechanochemistry can play a valuable role in improving both uniformity of dispersion and the screening rate of such polymorphs [10].

Etter and co-workers showed that solid-state grinding of equimolar quantities of 9-methyladenine and 1-methylthymine in an amalgam mill gave powder diffraction patterns consistent with the formation of Hoogsteen-type base-pairing (Scheme 16) [81].

No co-crystal formation was observed using 1-methylcytosine with 9-ethylguanine or other combinations which did not contain both adenine and thymine derivatives. The specificity of the Ade–Thy hydrogen bonding was not disrupted in the presence of non-interacting bases.

Tsiourvas and co-workers obtained a similar result after grinding an equimolar mixture of the hexadecylammonium salts of a succinylated acyclovir derivative (Figure 5) and its cytosine congener in an undefined "vibrator mill" at room temperature. However, above 80 °C a solid-state transition was observed in which base-pairing was inferred and upon further heating gave a smectic phase [82].

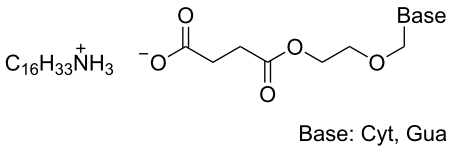
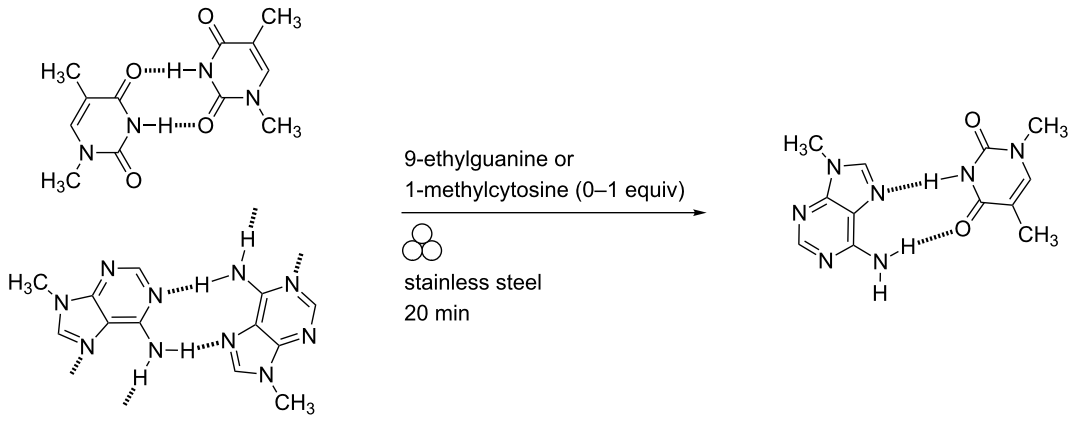


Figure 5: Materials used to prepare a smectic phase.

Using a mixer ball mill, Vogt and co-workers ground 5-fluorouracil (5FU, Figure 6) and thymine over a wide stoichiome-



Scheme 16: Co-crystal grinding of alkylated nucleobases in an amalgam mill (N.B. no frequency was recorded in the experimental description).

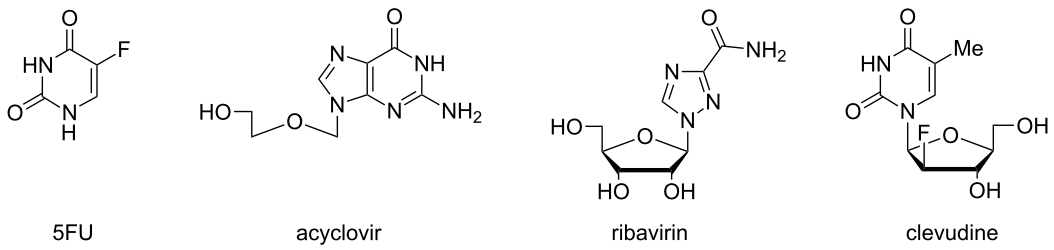


Figure 6: Structures of 5-fluorouracil (5FU) and nucleoside analogue prodrugs subject to mechanochemical co-crystal or polymorph transformation.

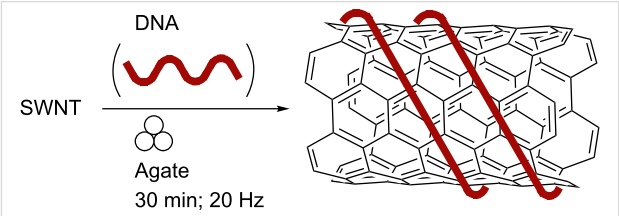
try range either under dry conditions or in the presence of a variety of organic solvents [83]. Liquid-assisted grinding of mixtures containing 50–90 mol % 5FU at 30 Hz for 30 min gave homogenised solid solutions using two drops of acetonitrile.

LAG of a 1:1 mixture of 5FU/4-hydroxybenzoic acid using a variety of liquids yielded co-crystals exhibiting polymorphism which was dependent upon the polarity of the added liquid [84]. Co-crystals of structurally-related carboxylic acids with 5FU prepared using LAG in a MBM in the presence of water exhibited enhanced membrane permeability compared with the pure API [85]. The preparation of co-crystals of 5FU with other API's (imatinib [86] and piperazine [87]) using LAG has also been reported.

Solid dispersions of acyclovir (20%) in neutral carriers (chitosan, hydroxypropylmethyl cellulose K100M® or Pluronic F68®) were prepared in a mixer ball mill over three hours [88]. All dispersions displayed antiviral activity and enhanced aqueous dissolution rates. The Pluoronic F68® dispersion displayed enhanced transport rates across a model intestinal cell monolayer.

The conversion of a stable ribavirin polymorph R-II into its metastable enantiotrope R-I has been investigated using mechanochemistry [89,90]. LAG in an improvised planetary mill using lead balls gave limited phase conversion [89] but dry milling R-II in a commercial mixer mill at 30 Hz gave 100% conversion within 15 minutes [90]. Three crystal polymorphs of the antiviral nucleoside prodrug clevudine were characterised and a large scale preparation of the most stable form from commercial material was performed using LAG in a mortar [91].

Geckeler and co-workers described the efficient preparation of both multi-walled and single-walled carbon nanotubes (CNTs) by grinding these materials in a mixer ball mill using agate components (Scheme 17) [92].



Scheme 17: Preparation of DNA-SWNT complex in a MBM.

In the absence of CNTs, DNA cleavage to a uniform size was found. Ball milling in the presence of monoribonucleotides has also been investigated as a method for solubilising single-walled carbon nanotubes [93]. In the presence of guanosine-5′-monophosphate, 78% of the SWNT (0.78 mg mL^{-1}) was dissolved but attempted removal of iron contamination from this material by treatment with acid gave a "viscous precipitate".

Formation of cyclodextrin–drug inclusion complexes can be accelerated using mechanochemistry [94] and Rajamohan and co-workers described using a mortar and pestle to effect LAG of β-cyclodextrin with either inosine [95] or cytidine [96] in the presence of water. Weak complex formation was inferred by powder XRD for cytidine.

Conclusion

Access to reliable and reproducible mechanised grinding has generated an upsurge in interest in mechanochemistry for a variety of chemical applications over the past decade. As a frontier science, theoretical models of reactivity under the action of mechanical forces are rapidly undergoing revision in the light of results available both from observations at a molecular scale [97] and from in situ monitoring of bulk-scale reactions [98]. The limited work relating to the chemical transformation of nucleoside and nucleotide substrates has been mainly focussed upon exploiting the (lack of) solvent requirements. This can allow access to unprecedented mechanochemical reaction pathways which would otherwise be unavailable through conventional solution-phase chemistry. This has included the preparation of pharmaceutical grade NR exclusively as the β-anomer and in the absence of bromide contamination [99] and in situ

hydrolytic unmasking of labile phosphorothiolate monoesters prior to rapid phosphate coupling. However, at the interface between biology and chemistry, the use of grinding to effect force-induced (mainly) dissociative reactions has a considerably longer heritage and it can be argued that Buchner's Nobel Prize in Chemistry was the first in this field. Although speaking from a more theoretical stand-point following observations on the infectivity of bacteriophages, Muller discussed the concept of genetic manipulation using (mechano)chemistry in 1922, commenting "perhaps we may be able to grind genes in a mortar and cook them in a beaker" [100].

Dubinskaya reviewed early investigations into the grinding and stretching of polypeptides and proteins which showed rapid loss of enzyme activity at 80 K [101]. In contrast, more recent reports, in which both native and immobilised enzymes were ground at higher temperatures, demonstrated efficient mechanoenzymatic transformations of amino acid [102-107], cellulose [108] or model lignin [109] substrates. Implicit within these studies is the resilience of chiral centres within both substrate and catalyst towards epimerisation during ball milling. This was also explicitly demonstrated by the Nagy lab in the context of developing models for the origin of non-racemic amino acid content within meteorites [110,111]. A role for mechanochemistry in understanding the origins of biochirogenesis is suggested by phase separation of co-crystals of the D- and L-enantiomers of malic acid in the presence of L-tartaric after grinding the racemate [112]. Heinicke briefly summarised early investigations into potential prebiotic α-amino acid preparation under "tribochemical stress" in the presence of transition metals [113] and more recently, Hernández and co-workers demonstrated that an efficient Strecker-type reaction could be effected in a ball mill using catalytic anhydrous ferricyanide in the presence of silica [114]. From a theoretical perspective, Hansma proposed that such mechanical energy could be supplied within moving mica sheets under high molecular crowding conditions [115].

In contrast, consideration of primordial nucleoside and nucleotide mechanochemistry has been much more limited with greater focus upon precursor mechanosynthesis under high-energy plasma conditions [116,117] or extra-terrestrial delivery of concentrated transition metals [118]. In this general context, although Orgel and co-workers describe transformation of adenine hydrochloride and D-ribose into the corresponding nucleoside α- (4%) or β- (3%) anomers following "thorough grinding" and heating of the mixture they do not distinguish between mechanochemistry and thermochemistry effects [119]. Considering recent investigations into low temperature ice eutectic phases as the incubators of early life [120] and separately, the effects of high hydrostatic pressures upon ribozyme activities [121,122], it is surprising how little consideration has been given to the role of nucleic acid mechanochemistry under prebiotic conditions. The capacity for stereoselective glycosidation, rapid phosphate coupling in the presence of water and also formation of specific base-pairing interactions have all been demonstrated in a ball mill and may facilitate understanding of the early appearance of life in the Hadean/Archean Eon.

Abbreviations

Table 1: List of abbreviations.

5FU	5-fluorouracil
Ade	N^9-adeninyl
Ade^{Bz}	N^6-benzoyl-N^9-adeninyl
ADPR	adenosine diphosphate ribose
AMP-morpholidate	adenosine 5′-monophosphoromorpholidate
Base	nucleobase (Ade, Cyt, Gua, Hyp, Thy or Ura)
Boc	*tert*-butyloxycarbonyl
BSA	*N,O*-bis(trimethylsilyl)acetamide
[C_6*mim*]	1-hexyl-3-methylimidazolium
CE	2-cyanoethyl
CNT	carbon nanotube
CuAAC	copper-assisted azide alkyne cycloaddition
Cyt	N^1-cytosinyl
Cyt^{Bz}	N^4-benzoyl-N^1-cytosinyl
dA	deoxyadenosinyl
DABCO	1,4-diazabicyclo[2.2.2]octane
DIPEA	*N*,*N*-diisopropylethylamine
DMAP	4-*N,N*-(dimethylamino)pyridine
DMTr	4,4′-dimethoxytrityl
dT	deoxythymidinyl
[FAP]	tris(pentafluoroethyl)trifluorophosphate
Gua	N^9-guaninyl
Gua^{iBu}	N^2-isobutyryl-N^9-guaninyl
HMDS	hexamethyldisilazane
Hyp	N^9-hypoxanthinyl
LAG	liquid-assisted grinding
M–A	Michaelis–Arbuzov
MBM	mixer ball mill
MMTr	4-methoxytrityl
NAD	nicotinamide adenine dinucleotide
NR	nicotinamide riboside
PCR	polymerase chain reaction
PG	protecting group
PTFE	polytetrafluoroethylene
Py	pyridine
SWNT	single-walled carbon nanotube
TBAB	tetra-*n*-butylammonium bromide
TBDMS	*tert*-butyldimethylsilyl
TFA	trifluoroacetate
Thy	N^1-thyminyl
Tr	trityl
Ura	N^1-uracilyl

Acknowledgements

Funding was provided by: the School of Chemistry and Chemical Engineering, QUB (PFC) and by the authors (OE, JSV, YL). We acknowledge Kerri Crossey for commentary upon patented work.

References

1. Takacs, L. *Chem. Soc. Rev.* **2013,** *42,* 7649–7659. doi:10.1039/c2cs35442j
2. McNaught, A. D.; Wilkinson, A., Eds. *IUPAC Compendium of Chemical Technology,* 2nd ed.; Blackwell Scientific Publications: Oxford, 1997.
3. Do, J.-L.; Friščić, T. *Synlett* **2017,** *28,* 2066–2092. doi:10.1055/s-0036-1590854
4. Do, J.-L.; Friščić, T. *ACS Cent. Sci.* **2017,** *3,* 13–19. doi:10.1021/acscentsci.6b00277
5. Achar, T. K.; Bose, A.; Mal, P. *Beilstein J. Org. Chem.* **2017,** *13,* 1907–1931. doi:10.3762/bjoc.13.186
6. Hernández, J. G.; Bolm, C. *J. Org. Chem.* **2017,** *82,* 4007–4019. doi:10.1021/acs.joc.6b02887
7. Tan, D.; Friščić, T. *Eur. J. Org. Chem.* **2018,** 18–33. doi:10.1002/ejoc.201700961
8. Leonardi, M.; Villacampa, M.; Menéndez, J. C. *Chem. Sci.* **2018,** *9,* 2042–2064. doi:10.1039/C7SC05370C
9. Sarkar, A.; Santra, S.; Kundu, S. K.; Hajra, A.; Zyryanov, G. V.; Chupakhin, O. N.; Charushin, V. N.; Majee, A. *Green Chem.* **2016,** *18,* 4475–4525. doi:10.1039/C6GC01279E
10. Tan, D.; Loots, L.; Friščić, T. *Chem. Commun.* **2016,** *52,* 7760–7781. doi:10.1039/C6CC02015A
11. Mottillo, C.; Friščić, T. *Molecules* **2017,** *22,* No. 144. doi:10.3390/molecules22010144
12. André, V.; Quaresma, S.; da Silva, J. L. F.; Duarte, M. T. *Beilstein J. Org. Chem.* **2017,** *13,* 2416–2427. doi:10.3762/bjoc.13.239
13. Hodgson, D. R. W. *Adv. Phys. Org. Chem.* **2017,** *51,* 187–219. doi:10.1016/bs.apoc.2017.09.002
14. Buchner, E. "Eduard Buchner - Nobel Lecture: Cell-Free Fermentation". 1907. https://www.nobelprize.org/nobel_prizes/chemistry/laureates/1907/buchner-lecture.pdf (accessed Dec 22, 2017).
15. Andersen, J.; Mack, J. *Green Chem.* **2018,** *20,* 1435–1443. doi:10.1039/C7GC03797J
16. Stolle, A.; Szuppa, T.; Leonhardt, S. E. S.; Ondruschka, B. *Chem. Soc. Rev.* **2011,** *40,* 2317–2329. doi:10.1039/c0cs00195c
17. Stolle, A. Technical Implications of Organic Syntheses in Ball Mills. In *Ball Milling Towards Green Synthesis: Applications, Projects, Challenges;* Stolle, A.; Ranu, B., Eds.; The Royal Society of Chemistry: Cambridge, 2015; pp 241–276. doi:10.1039/9781782621980-00241
18. Rightmire, N. R.; Hanusa, T. P. *Dalton Trans.* **2016,** *45,* 2352–2362. doi:10.1039/C5DT03866A
19. Sikchi, S. A.; Hultin, P. G. *J. Org. Chem.* **2006,** *71,* 5888–5891. doi:10.1021/jo060430t
20. Metro, T.-X.; Bonnamour, J.; Reidon, T.; Duprez, A.; Sarpoulet, J.; Martinez, J.; Lamaty, F. *Chem. – Eur. J.* **2015,** *21,* 12787–12796. doi:10.1002/chem.201501325
21. Štefanić, G.; Krehula, S.; Štefanić, I. *Chem. Commun.* **2013,** *49,* 9245–9247. doi:10.1039/c3cc44803g
22. Eguaogie, O.; Conlon, P. F.; Ravalico, F.; Sweet, J. S. T.; Elder, T. B.; Conway, L. P.; Lennon, M. E.; Hodgson, D. R. W.; Vyle, J. S. *Beilstein J. Org. Chem.* **2017,** *13,* 87–92. doi:10.3762/bjoc.13.11
23. Eguaogie, O.; Cooke, L. A.; Martin, P. M. L.; Ravalico, F.; Conway, L. P.; Hodgson, D. R. W.; Law, C. J.; Vyle, J. S. *Org. Biomol. Chem.* **2016,** *14,* 1201–1205. doi:10.1039/C5OB02061A
24. Makarov, D. E. *J. Chem. Phys.* **2016,** *144,* No. 030901. doi:10.1063/1.4939791
25. Stauch, T.; Dreuw, A. *Acc. Chem. Res.* **2017,** *50,* 1041–1048. doi:10.1021/acs.accounts.7b00038
26. Garcia-Manyes, S.; Beedle, A. E. M. *Nat. Rev. Chem.* **2017,** *1,* No. 0083. doi:10.1038/s41570-017-0083
27. Liang, J.; Fernández, J. M. *J. Am. Chem. Soc.* **2011,** *133,* 3528–3534. doi:10.1021/ja109684q
28. Schmidt, S. W.; Filippov, P.; Kersch, A.; Beyer, M. K.; Clausen-Schaumann, H. *ACS Nano* **2012,** *6,* 1314–1321. doi:10.1021/nn204111w
29. Fernández-Bertran, J. F. *Pure Appl. Chem.* **1999,** *71,* 581–586. doi:10.1351/pac199971040581
30. Rothenberg, G.; Downie, A. P.; Raston, C. L.; Scott, J. L. *J. Am. Chem. Soc.* **2001,** *123,* 8701–8708. doi:10.1021/ja0034388
31. Hutchings, B. P.; Crawford, D. E.; Gao, L.; Hu, P.; James, S. L. *Angew. Chem., Int. Ed.* **2017,** *56,* 15252–15256. doi:10.1002/anie.201706723
32. Stolle, A.; Schmidt, R.; Jacob, K. *Faraday Discuss.* **2014,** *170,* 267–286. doi:10.1039/C3FD00144J
33. Friščić, T.; Childs, S. L.; Rizvi, S. A. A.; Jones, W. *CrystEngComm* **2009,** *11,* 418–426. doi:10.1039/B815174A
34. Cintas, P.; Cravotto, G.; Barge, A.; Martina, K. Interplay Between Mechanochemistry and Sonochemistry. In *Polymer Mechanochemistry;* Boulatov, R., Ed.; Springer International Publishing: Cham, 2015; pp 239–284.
35. James, S. L.; Adams, C. J.; Bolm, C.; Braga, D.; Collier, P.; Friščić, T.; Grepioni, F.; Harris, K. D. M.; Hyett, G.; Jones, W.; Krebs, A.; Mack, J.; Maini, L.; Orpen, A. G.; Parkin, I. P.; Shearouse, W. C.; Steed, J. W.; Waddell, D. C. *Chem. Soc. Rev.* **2012,** *41,* 413–447. doi:10.1039/C1CS15171A
36. Thorwirth, R.; Bernhardt, F.; Stolle, A.; Ondruschka, B.; Asghari, J. *Chem. – Eur. J.* **2010,** *16,* 13236–13242. doi:10.1002/chem.201001702
37. Ciriminna, R.; Pagliaro, M. *Org. Process Res. Dev.* **2013,** *17,* 1479–1484. doi:10.1021/op400258a
38. Wallington, B. Environmental control. In *Active pharmaceutical ingredients - development, manufacturing and regulation,* 2nd ed.; Nusim, S. H., Ed.; CRC Press: Boca Raton, 2010; pp 203–236.
39. Khalafi-Nezhad, A.; Mokhtari, B. *Tetrahedron Lett.* **2004,** *45,* 6737–6739. doi:10.1016/j.tetlet.2004.07.054
40. Patil, P. R.; Kartha, K. P. R. *J. Carbohydr. Chem.* **2008,** *27,* 279–293. doi:10.1080/07328300802218713
41. Giri, N.; Bowen, C.; Vyle, J. S.; James, S. L. *Green Chem.* **2008,** *10,* 627–628. doi:10.1039/b801455h
42. Ravalico, F.; James, S. L.; Vyle, J. S. *Green Chem.* **2011,** *13,* 1778–1783. doi:10.1039/c1gc15131b
43. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002,** *41,* 2596–2599. doi:10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4

44. Cummings, A. J.; Ravalico, F.; McColgan-Bannon, K. I. S.; Eguaogie, O.; Elliott, P. A.; Shannon, M. R.; Bermejo, I. A.; Dwyer, A.; Maginty, A. B.; Mack, J.; Vyle, J. S. *Nucleosides, Nucleotides Nucleic Acids* **2015,** *34,* 361–370. doi:10.1080/15257770.2014.1001855
45. Crossey, K.; Cunningham, R. N.; Redpath, P.; Migaud, M. E. *RSC Adv.* **2015,** *5,* 58116–58119. doi:10.1039/C5RA12239B
46. Migaud, M.; Redpath, P.; Crossey, K.; Doherty, M. Methods of preparing nicotinamide riboside and derivatives thereof. WO patent WO2015014722A1, Feb 5, 2015.
47. Migaud, M. E.; Redpath, P.; Crossey, K.; Cunningham, R.; Dellinger, R.; Rhonemus, T.; Venkataraman, S.; Nettles, B. B-vitamin and amino acid conjugates of nicotinoyl ribosides and reduced nicotinoyl ribosides, derivatives thereof, and methods of preparation thereof. U.S. Patent US201702677099, Sept 21, 2017.
48. Migaud, M. E.; Redpath, P.; Crossey,, K.; Cunningham, R.; Rhonemus, T.; Venkataraman, S. Selective solvent free phosphorylation. U.S. Patent US20160355539, Dec 8, 2016.
49. Hardacre, C.; Huang, H.; James, S. L.; Migaud, M. E.; Norman, S. E.; Pitner, W. R. *Chem. Commun.* **2011,** *47,* 5846–5848. doi:10.1039/c1cc11025j
50. Crossey, K.; Hardacre, C.; Migaud, M. E. *Chem. Commun.* **2012,** *48,* 11969–11971. doi:10.1039/c2cc36367d
51. Moffatt, J. G.; Khorana, H. G. *J. Am. Chem. Soc.* **1961,** *83,* 649–658. doi:10.1021/ja01464a034
52. Roy, B.; Depaix, A.; Périgaud, C.; Peyrottes, S. *Chem. Rev.* **2016,** *116,* 7854–7897. doi:10.1021/acs.chemrev.6b00174
53. Ravalico, F.; Messina, I.; Berberian, M. V.; James, S. L.; Migaud, M. E.; Vyle, J. S. *Org. Biomol. Chem.* **2011,** *9,* 6496–6497. doi:10.1039/c1ob06041d
54. Dvorakova, M.; Nencka, R.; Dejmek, M.; Zbornikova, E.; Brezinova, A.; Pribylova, M.; Pohl, R.; Migaud, M. E.; Vanek, T. *Org. Biomol. Chem.* **2013,** *11,* 5702–5713. doi:10.1039/c3ob41016a
55. Brear, P.; Freeman, G. R.; Shankey, M. C.; Trmčić, M.; Hodgson, D. R. W. *Chem. Commun.* **2009,** 4980–4981. doi:10.1039/b908727c
56. Eguaogie, O.; Vyle, J. S. *Curr. Protoc. Nucleic Acid Chem.* **2017,** *70,* 1.41.1–1.41.12. doi:10.1002/cpnc.37
57. Warman, J. M.; Eldrup, M. *Biopolymers* **1986,** *25,* 1865–1874. doi:10.1002/bip.360251005
58. Privalov, P. L. *Crit. Rev. Biochem. Mol. Biol.* **1990,** *25,* 281–306. doi:10.3109/10409239009090612
59. Chargaff, E. Isolation and composition of the deoxypentose nucleic acids and the corresponding nucleoproteins. In *The Nucleic Acids,* 2nd ed.; Chargaff, E.; Davidson, J. N., Eds.; Academic Press: New York, 1955; Vol. 1, pp 307–372.
60. Miescher, F. *Die Histochemischen und Physiologischen Arbeiten von Friedrich Miescher;* F. C. W. Vogel: Leipzig, 1897; Vol. 2, pp 71–76.
61. Dahm, R. *Dev. Biol. (Amsterdam, Neth.)* **2005,** *278,* 274–288. doi:10.1016/j.ydbio.2004.11.028
62. Bang, I. *Beitr. Chem. Physiol. Pathol.* **1904,** *4,* 115–188.
63. Schwander, H.; Signer, R. *Helv. Chim. Acta* **1950,** *33,* 1521–1526. doi:10.1002/hlca.19500330618
64. Hammarsten, E.; Hammarsten, G. *Acta Med. Scand.* **1928,** *68,* 199–204. doi:10.1111/j.0954-6820.1928.tb12350.x
65. Schmidt, G.; Levene, P. A. *Science* **1938,** *88,* 172–173. doi:10.1126/science.88.2277.172
66. Franklin, R. E.; Gosling, R. G. *Nature* **1953,** *171,* 740–741. doi:10.1038/171740a0
67. Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. *Nature* **1953,** *171,* 738–740. doi:10.1038/171738a0
68. Kossel, A. *Hoppe-Seyler's Z. Physiol. Chem.* **1879,** *3,* 284–291.
69. Altmann, R. *Arch. Anat. Physiol.* **1889,** 524–536.
70. Magansanik, B. Isolation and composition of the pentose nucleic acids and of the corresponding nucleoproteins. In *The Nucleic Acids,* 2nd ed.; Chargaff, E.; Davidson, J. N., Eds.; Academic Press: New York, 1955; Vol. 1, pp 373–407.
71. Maes, M.; Messens, E. *Nucleic Acids Res.* **1992,** *20,* 4374. doi:10.1093/nar/20.16.4374
72. Behrens, M. *Hoppe-Seyler's Z. Physiol. Chem.* **1932,** *209,* 59–74. doi:10.1515/bchm2.1932.209.1-2.59
73. Dounce, A. L.; Tishkoff, G. H.; Barnett, S. R.; Freer, R. M. *J. Gen. Physiol.* **1950,** *33,* 629–642. doi:10.1085/jgp.33.5.629
74. Allfrey, V.; Stern, H.; Mirsky, A. E.; Saetren, H. *J. Gen. Physiol.* **1952,** *35,* 529–557. doi:10.1085/jgp.35.3.529
75. Dounce, A. L. Isolation and composition of cell nuclei and nucleoli. In *The Nucleic Acids,* 2nd ed.; Chargaff, E.; Davidson, J. N., Eds.; Academic Press: New York, 1955; Vol. 2, pp 93–153.
76. Macfayden, A.; Rowland, S. *Centr. Bakt. Orig.* **1903,** *34,* 765–771.
77. Barnard, J. E.; Hewlett, R. T. *Proc. R. Soc. London, Ser. B* **1911,** *84,* 57–66. doi:10.1098/rspb.1911.0046
78. Mudd, S.; Shaw, C. H.; Czarnetzky, E. J.; Flosdorf, E. W. *J. Immunol.* **1937,** *32,* 483–489.
79. Shehadul Islam, M.; Aryasomayajula, A.; Selvaganapathy, P. R. *Micromachines* **2017,** *8,* No. 83. doi:10.3390/mi8030083
80. Hwang, K.-Y.; Kwon, S. H.; Jung, S.-O.; Lim, H.-K.; Jung, W.-J.; Park, C.-S.; Kim, J.-H.; Suh, K.-Y.; Huh, N. *Lab Chip* **2011,** *11,* 3649–3655. doi:10.1039/c1lc20692c
81. Etter, M. C.; Reutzel, S. M.; Choo, C. G. *J. Am. Chem. Soc.* **1993,** *115,* 4411–4412. doi:10.1021/ja00063a089
82. Tsiourvas, D.; Mihou, A. P.; Couladouros, E. A.; Paleos, C. M. *Mol. Cryst. Liq. Cryst.* **2001,** *362,* 177–184. doi:10.1080/10587250108025768
83. Vogt, F. G.; Vena, J. A.; Chavda, M.; Clawson, J. S.; Strohmeier, M.; Barnett, M. E. *J. Mol. Struct.* **2009,** *932,* 16–30. doi:10.1016/j.molstruc.2009.05.035
84. Li, S.; Chen, J.-M.; Lu, T.-B. *CrystEngComm* **2014,** *16,* 6450–6458. doi:10.1039/C4CE00221K
85. Dai, X.-L.; Li, S.; Chen, J.-M.; Lu, T.-B. *Cryst. Growth Des.* **2016,** *16,* 4430–4438. doi:10.1021/acs.cgd.6b00552
86. Veverka, M.; Šimon, P.; Gallovič, J.; Jorík, V.; Veverková, E.; Dubaj, T. *Monatsh. Chem.* **2012,** *143,* 1405–1415. doi:10.1007/s00706-012-0788-3
87. Moisescu-Goia, C.; Muresan-Pop, M.; Simon, V. *J. Mol. Struct.* **2017,** *1150,* 37–43. doi:10.1016/j.molstruc.2017.08.076
88. Nart, V.; França, M. T.; Anzilaggo, D.; Riekes, M. K.; Kratz, J. M.; de Campos, C. E. M.; Simões, C. M. O.; Stulzer, H. K. *Mater. Sci. Eng., C* **2015,** *53,* 229–238. doi:10.1016/j.msec.2015.04.028
89. Tong, H. H. Y.; Shekunov, B. Y.; Chan, J. P.; Mok, C. K. F.; Hung, H. C. M.; Chow, A. H. L. *Int. J. Pharm.* **2005,** *295,* 191–199. doi:10.1016/j.ijpharm.2005.02.024
90. Vasa, D. M.; Wildfong, P. L. D. *Int. J. Pharm.* **2017,** *524,* 339–350. doi:10.1016/j.ijpharm.2017.04.002
91. Noonan, T. J.; Mzondo, B.; Bourne, S. A.; Caira, M. R. *CrystEngComm* **2016,** *18,* 8172–8181. doi:10.1039/C6CE01975G
92. Nepal, D.; Sohn, J.-I.; Aicher, W. K.; Lee, S.; Geckeler, K. E. *Biomacromolecules* **2005,** *6,* 2919–2922. doi:10.1021/bm050380m

93. Ikeda, A.; Hamano, T.; Hayashi, K.; Kikuchi, J.-i. *Org. Lett.* **2006,** *8,* 1153–1156. doi:10.1021/ol053089s
94. Lin, S.-Y.; Lee, C.-S. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1989,** *7,* 477–485. doi:10.1007/BF01080458
95. Prabu, S.; Sivakumar, K.; Swaminathan, M.; Rajamohan, R. *Spectrochim. Acta, Part A* **2015,** *147,* 151–157. doi:10.1016/j.saa.2015.03.056
96. Prabu, S.; Sivakumar, K.; Nayaki, S. K.; Rajamohan, R. *J. Mol. Liq.* **2016,** *219,* 967–974. doi:10.1016/j.molliq.2016.04.017
97. Wang, J.; Kouznetsova, T. B.; Niu, Z.; Ong, M. T.; Klukovich, H. M.; Rheingold, A. L.; Martinez, T. J.; Craig, S. L. *Nat. Chem.* **2015,** *7,* 323–327. doi:10.1038/nchem.2185
98. Užarević, K.; Halasz, I.; Friščić, T. *J. Phys. Chem. Lett.* **2015,** *6,* 4129–4140. doi:10.1021/acs.jpclett.5b01837
99. Trammell, S. A. J.; Schmidt, M. S.; Weidemann, B. J.; Redpath, P.; Jaksch, F.; Dellinger, R. W.; Li, Z.; Abel, E. D.; Migaud, M. E.; Brenner, C. *Nat. Commun.* **2016,** *7,* No. 12948. doi:10.1038/ncomms12948
100. Muller, J. H. *Am. Nat.* **1922,** *56,* 32–50. doi:10.1086/279846
101. Dubinskaya, A. M. *Russ. Chem. Rev.* **1999,** *68,* 637–652. doi:10.1070/RC1999v068n08ABEH000435
102. Hernández, J. G.; Frings, M.; Bolm, C. *ChemCatChem* **2016,** *8,* 1769–1772. doi:10.1002/cctc.201600455
103. Pérez-Venegas, M.; Reyes-Rangel, G.; Neri, A.; Escalante, J.; Juaristi, E. *Beilstein J. Org. Chem.* **2017,** *13,* 1728–1734. doi:10.3762/bjoc.13.167
104. Pérez-Venegas, M.; Reyes-Rangel, G.; Neri, A.; Escalante, J.; Juaristi, E. *Beilstein J. Org. Chem.* **2017,** *13,* 2128–2130. doi:10.3762/bjoc.13.210
105. Hernández, J. G.; Ardila-Fierro, K. J.; Crawford, D.; James, S. L.; Bolm, C. *Green Chem.* **2017,** *19,* 2620–2625. doi:10.1039/C7GC00615B
106. Ardila-Fierro, K. J.; Crawford, D. E.; Körner, A.; James, S. L.; Bolm, C.; Hernández, J. G. *Green Chem.* **2018,** *20,* 1262–1269. doi:10.1039/C7GC03205F
107. Bolm, C.; Hernández, J. G. *ChemSusChem* **2018,** in press. doi:10.1002/cssc.201800113.
108. Hammerer, F.; Loots, L.; Do, J.-L.; Therien, J. P. D.; Nickels, C. W.; Friščić, T.; Auclair, K. *Angew. Chem., Int. Ed.* **2018,** *57,* 2621–2624. doi:10.1002/anie.201711643
109. Weißbach, U.; Dabral, S.; Konnert, L.; Bolm, C.; Hernández, J. G. *Beilstein J. Org. Chem.* **2017,** *13,* 1788–1795. doi:10.3762/bjoc.13.173
110. Engel, M. H.; Nagy, B. *Nature* **1982,** *296,* 837–840. doi:10.1038/296837a0
111. Engel, M. H. Amino acids in ancient (Precambrian) rocks: their occurrence, abundance and degree of racemization. Ph.D. Thesis, University of Arizona, 1980.
112. Eddleston, M. D.; Arhangelskis, M.; Friščić, T.; Jones, W. *Chem. Commun.* **2012,** *48,* 11340–11342. doi:10.1039/c2cc36130b
113. Heinicke, G. *Tribochemistry;* Akademie-Verlag: Berlin, 1984; pp 474–480.
114. Bolm, C.; Mocci, R.; Schumacher, C.; Turberg, M.; Puccetti, F.; Hernández, J. G. *Angew. Chem., Int. Ed.* **2018,** *57,* 2423–2426. doi:10.1002/anie.201713109
115. Hansma, H. G. *J. Biomol. Struct. Dyn.* **2013,** *31,* 888–895. doi:10.1080/07391102.2012.718528 and references therein.
116. Kalson, N.-H.; Furman, D.; Zeiri, Y. *ACS Cent. Sci.* **2017,** *3,* 1041–1049. doi:10.1021/acscentsci.7b00325
117. Ferus, M.; Michalčíková, R.; Shestivská, V.; Šponer, J.; Šponer, J. E.; Civiš, S. *J. Phys. Chem. A* **2014,** *118,* 719–736. doi:10.1021/jp411415p
118. Sutherland, J. D. *Angew. Chem., Int. Ed.* **2016,** *55,* 104–121. doi:10.1002/anie.201506585
119. Fuller, W. D.; Sanchez, R. A.; Orgel, L. E. *J. Mol. Evol.* **1972,** *1,* 249–257. doi:10.1007/BF01660244
120. Feller, G. *Life* **2017,** *7,* No. 25. doi:10.3390/life7020025
121. Schuabb, C.; Kumar, N.; Pataraia, S.; Marx, D.; Winter, R. *Nat. Commun.* **2017,** *8,* No. 14661. doi:10.1038/ncomms14661
122. Tobé, S.; Heams, T.; Vergne, J.; Hervé, G.; Maurel, M.-C. *Nucleic Acids Res.* **2005,** *33,* 2557–2564. doi:10.1093/nar/gki552

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

An overview of recent advances in duplex DNA recognition by small molecules

Sayantan Bhaduri[1], Nihar Ranjan[2] and Dev P. Arya*[1,3]

Review

Open Access

Address:
[1]NUBAD, LLC, 900B West Faris Rd., Greenville 29605, SC, USA, [2]National Institute of Pharmaceutical Education and Research (NIPER), Raebareli 122003, India and [3]Clemson University, Hunter Laboratory, Clemson 29634, SC, USA

Email:
Dev P. Arya* - dparya@clemson.edu

* Corresponding author

Keywords:
alkylators; antibiotic; anticancer; antineoplastic; antiproliferative; DNA recognition; groove binders; hairpin polyamides; Hoechst 33258; intercalators

Beilstein J. Org. Chem. **2018,** *14,* 1051–1086.
doi:10.3762/bjoc.14.93

Received: 27 December 2017
Accepted: 06 April 2018
Published: 16 May 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

As the carrier of genetic information, the DNA double helix interacts with many natural ligands during the cell cycle, and is amenable to such intervention in diseases such as cancer biogenesis. Proteins bind DNA in a site-specific manner, not only distinguishing between the geometry of the major and minor grooves, but also by making close contacts with individual bases within the local helix architecture. Over the last four decades, much research has been reported on the development of small non-natural ligands as therapeutics to either block, or in some cases, mimic a DNA–protein interaction of interest. This review presents the latest findings in the pursuit of novel synthetic DNA binders. This article provides recent coverage of major strategies (such as groove recognition, intercalation and cross-linking) adopted in the duplex DNA recognition by small molecules, with an emphasis on major works of the past few years.

Review

1. Introduction

DNA is one of the central components of cellular machinery and storage unit of genetic information. It plays key roles in replication, transcription, protein-coding and cell integrity as well as in carrying the genetic blueprint for inheritance. The DNA–protein interactions involve high fidelity protein readout of the base edges exposed in the major and minor grooves of the DNA. Such interactions are also augmented by a series of electrostatic and van der Waals interactions including salt bridge formation with the phosphate backbone [1]. Although, the majority of proteins recognize DNA in the major groove due, in large part, to the potential and shape complementarity, several others also recognize the minor groove by sufficiently distort-

ing the DNA structures leading to the opening of the minor groove [2]. In addition to the conventional direct and indirect readout mechanism, proteins have also been proposed to recognize the DNA minor groove by sensing variations in the shape and electrostatics [3].

The coding regions of the human genomic DNA contain highly conserved sequences that express proteins, which are essential for the cell survival and maintenance. Over or under expression of proteins has been linked to several disease states including cancer [4]. Therefore, control of gene expression has been long perceived and successfully demonstrated as a means of therapeutic development. Since DNA–protein interactions involve significant contacts in the major and minor grooves of DNA for error-free readout, small molecules (natural and synthetic) that bind strongly in the grooves have been discovered and designed to competitively inhibit such interactions. Additionally, molecules that are capable of insertion between the DNA base pairs can also disfavor DNA–protein interactions directly or allosterically. Consequently, small molecule DNA binders have been in the limelight of drug-discovery programs due to their ability to act as gene expression inhibitors [5].

The recognition of DNA is both shape and sequence dependent as DNA polymorphism leads to significant changes in the groove structure. DNA is broadly categorized to possess three major forms: A, B and Z which differ from one another in several ways such as helical sense, pitch, groove width, base orientation and sugar pucker (Table 1). The major differences in the two generally encountered A- and B-forms of DNA is in the sugar pucker and their groove widths. In A-form DNA, the major groove is narrower but has a wide/shallow minor groove. In contrast, the minor groove of B-DNA is narrow and becomes even narrower in DNAs with contiguous AT stretches (termed as the B* form of the DNA) where the width of the narrow groove reduces to approximately 2.8 Å from a usually observed width of approximately 5.7 Å [6]. In contrast to the A- and B-form DNA, Z-DNA is a left handed structure formed by alternating G and C base pairs and contains some features of both A- and B-DNA such as the sugar pucker and a slightly bigger number of base pairs per turn [7].

The discovery of multistranded DNA structures such as G-quadruplexes [9], which uses eight Hoogsteen-paired hydrogen bonds to form a tetrad (Figure 1) has further enhanced our understanding of the diversity of DNA shapes and structures. In a parallel tetramolecular quadruplex d(TG_4T), the features of nucleotides at each base resemble that of the B-DNA (C2'-*endo* sugar pucker, *anti* orientation and ≈12 Å groove width). However, in quadruplex fold-back structures, unusual loop connectivity gives rise to extremely wide grooves in addition to narrow and medium grooves [10] in which the width of wide grooves goes up to approximately 18 Å, far exceeding the groove widths found in B-DNA structures. These variations in the groove widths and shapes shed light on the challenges in programmed DNA recognition in a sequence and shape selective manner.

DNA recognition by small molecules can be divided in two broad categories: covalent and non-covalent. Covalent binding (e.g., cis-platin binding to guanine bases) to DNA is irreversible and causes permanent stall of transcription leading to cell death. Non-covalent interaction between small molecules and DNA is usually reversible and can further be classified as minor groove binders, intercalators, backbone binders, and major groove binders. There are reports of natural and designed molecules that display multivalency in DNA recognition by binding at more than recognition sites (minor groove, major groove or base pair insertion) [11-13]. In synthetic multivalent ligands, which are made to enhance DNA affinity, tether length and composition play a significant role in target selectivity and specificity.

Several focused reviews on small molecule DNA binding agents have been published in recent years. A few have updated the progresses made in disease specific DNA binders [14,15] while others have included class specific or site-specific DNA binding agents [16-23]. A few others have covered nucleic acids

Table 1: A table showing the differences in the A-, B- and Z-form DNA [7,8].

	A-form	B-form	Z-form
helix sense	right-handed	right-handed	left-handed
base pairs/turn	11	10.4	12
pitch per turn of Helix	25.3 Å	35.4 Å	45.6 Å
glycosyl bond	*anti*	*anti*	alternating *anti* and *syn*
sugar pucker	C3'-*endo*	C2'-*endo*	C:C2'-*endo*, G:C3'-*endo*
major groove	narrow and very deep	wide and quite deep	flat
minor groove	very broad and shallow	narrow and quite deep	very narrow and deep

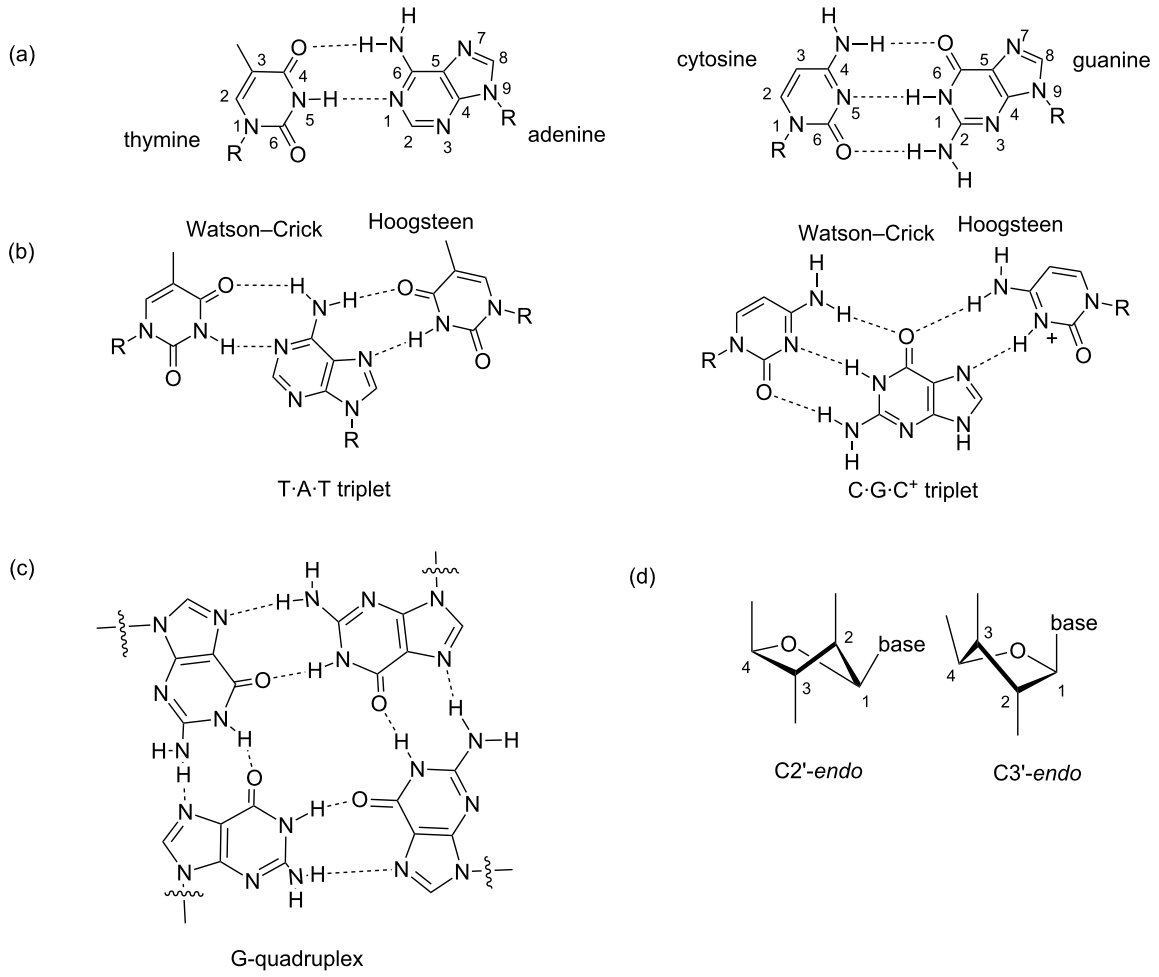


Figure 1: A figure showing the hydrogen bonding patterns observed in (a) duplex (b) triplex and (c) quadruplex DNA structures. (d) Conformations of sugar pucker in DNA.

binders in general [20] as well as an emerging therapeutic DNA target: the DNA G-quadruplex [24]. In this review, we provide a detailed overview of discoveries made in the search of duplex DNA recognition agents (groove binders, intercalators and alkylating agents), which includes both classical DNA binders and new advancements in the recent years (with emphasis on research advances reported in the last five years). For a focused work, we have excluded triplex and quadruplex DNA binders for this review. In particular, we cover the advances made in DNA minor groove recognition using new analogues and derivatives of classical minor groove binders such as distamycin, netropsin, polyamides, bisbenzimidazoles and organic cations. We have also included new intercalating agents as well as major groove binding ligands especially the multivalent ligands that can simultaneously recognize one or more sites on DNA leading to strong affinity for DNA. We finally shed light on new reports of DNA alkylating agents towards the end of this review. While it is impossible to absorb the vast expanse and comprehensiveness of reports on all DNA binding agents, this review article intends to provide a substantial coverage of new advancements made in the discovery of major leads in three most visited areas (groove recognition, intercalation and cross linking agents) of DNA recognition.

2. Minor groove binders (MGBs)

DNA groove binding small molecules comprise various heterocyclic and/or aromatic hydrocarbon rings with limited rotational freedom and torsion, allowing these drugs to fit into major/minor grooves of DNA by displacing water molecules from the spine of hydration as shown in the Figure 2 [25-27]. These molecules bind to the edges of the base pairs of the DNA duplex (usually G·C sites in the major groove, A·T sites in the minor groove) via reversible non-covalent interactions. These binding interactions reduce the conformational freedom of the small molecules and usually are opposed by an unfavorable entropic cost. However, these energetic costs are balanced and

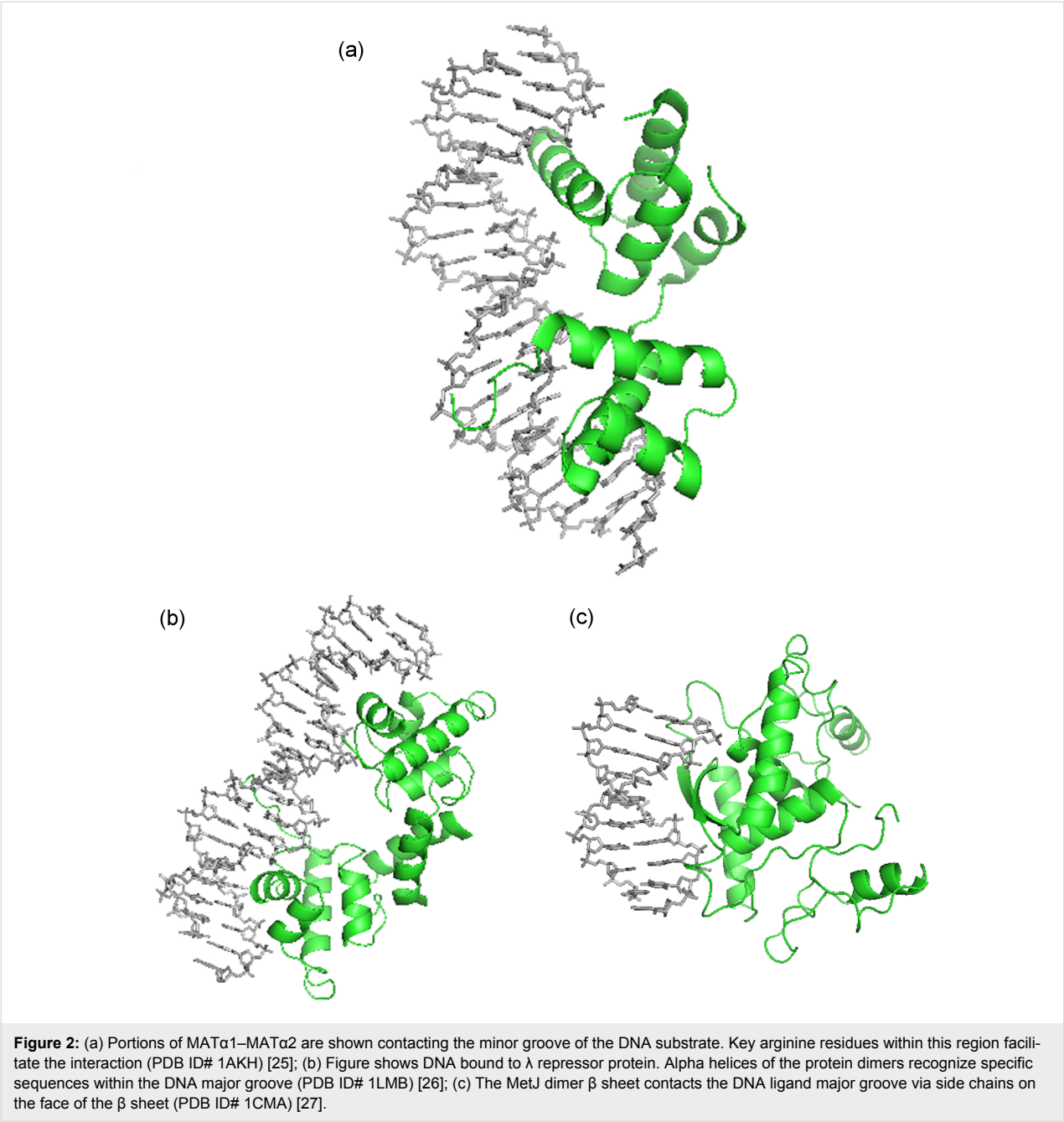


Figure 2: (a) Portions of MATα1–MATα2 are shown contacting the minor groove of the DNA substrate. Key arginine residues within this region facilitate the interaction (PDB ID# 1AKH) [25]; (b) Figure shows DNA bound to λ repressor protein. Alpha helices of the protein dimers recognize specific sequences within the DNA major groove (PDB ID# 1LMB) [26]; (c) The MetJ dimer β sheet contacts the DNA ligand major groove via side chains on the face of the β sheet (PDB ID# 1CMA) [27].

outweighed by favorable contributions from the hydrophobic transfer of drugs from solution to DNA-binding site [28,29]. Groove binding usually does not influence huge structural/conformational changes in the DNA duplex; this mode of binding may be considered similar to a standard lock and key recognition [30].

Minor groove binding drugs (MGBs) are usually isohelical, crescent-shaped molecules, which are compatible with the shape of the minor groove. Binding of MGBs and proteins occurs primarily via H-bonds, electrostatics, van der Waals and hydrophobic interactions (Figure 2). Figure 2a shows that the arginine side chain of the MATα2 N-terminal arm facilitates interaction between portions of the heterodimer MATα1–MATα2 with the minor groove of the DNA substrate by forming alternate H-bond interactions [25]. The main characteristic feature of MGBs is their preference for narrow A·T-rich regions compared to G·C regions because (i) they can form hydrogen bonds to N3 of adenine and O2 of thymine in the A·T region; (ii) less steric hindrance in the A·T region in comparison to the G·C region due to the presence of an extra protruding C2-amino group of the guanine base [19].

2.1. Polypyrroles and polyamides

The first two MGBs discovered were distamycin A and netropsin (Figure 3). These naturally occurring molecules are characterized by repeating *N*-methylpyrrole units with one or more positively charged nitrogen atoms at the end. Their concave-shaped aromatic framework fits perfectly in the convex-shaped minor groove of double-stranded DNA. Therefore, these drugs have been referred to as "shape-selective" binders [31]. They selectively interact with A·T-rich regions containing at least four A·T base pairs in the minor groove via hydrogen bonding interaction between the groove floor base pairs and the amides and electrostatic stabilizing interactions between the protonated amines under physiological pH and negatively charged phosphate backbone as reported by NMR and crystallographic studies [32-36]. These molecules were shown as inhibitors of Werner and Bloom syndrome helicases and dual topoisomerase I/II inhibitors [37,38].

In order to improve DNA binding affinity and sequence specificity with reduced side effects, a series of synthetic hybrid molecules derived from distamycin and netropsin was synthesized and their biological activities were thoroughly studied both in vitro and in vivo. One significant representative of this class is tallimustine (FCE 24517, TAM), which is a benzoyl nitrogen mustard derivative of distamycin characterized by an oligopeptidic pyrrolocarbamoyl framework ending with an amidino moiety [39,40]. The benzoyl nitrogen mustard (BAM) to the formyl end of the distamycin acts as an alkylating moiety whereas the distamycin framework acts as a DNA binding domain. Therefore, due to the installation of the alkylating moiety, TAM has higher cytotoxic activity in comparison to distamycin, and shows a broad spectrum of in vitro and in vivo antitumor activities. Tallimustine retains the preference for A·T-rich regions in the minor groove that alkylates N3 of adenine in a highly sequence specific manner, thereby inhibiting the

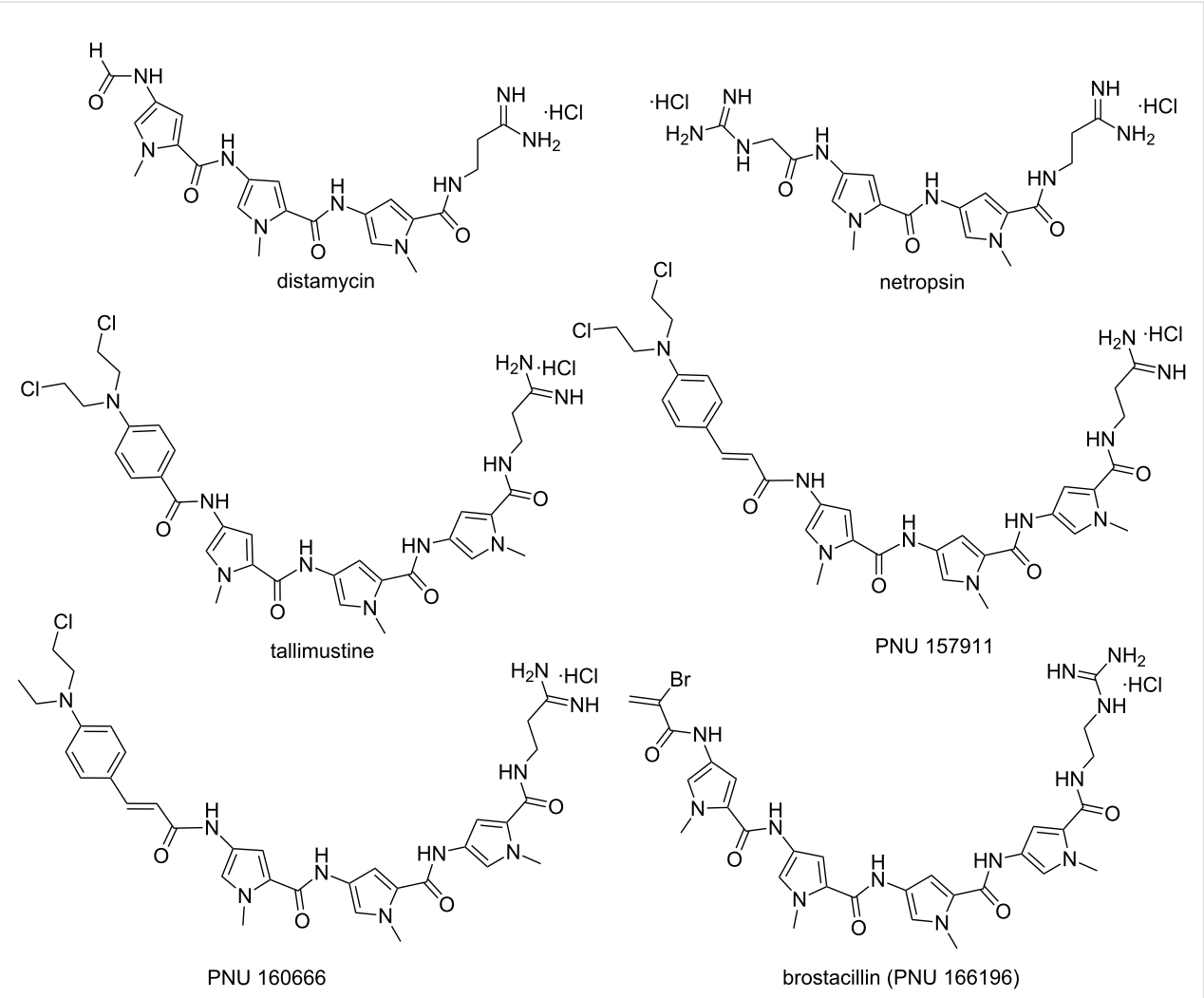


Figure 3: Chemical structures of naturally occurring and synthetic hybrid minor groove binders.

binding of transcription factors such as OTF-1 and NFE1 on specific AT-rich sequences [41,42]. However, clinical development of TAM was discontinued due to severe myelotoxicity.

With distamycin, netropsin and TAM as the lead compounds for novel anticancer drugs, a plethora of oligopyrrole derivatives were reported with the aim of increasing stability, greater DNA binding affinity, sequence specificity, more cytotoxicity and minimizing the unwanted physiological side effects [43]. It has been observed that drugs with high degree of sequence specific binding affinity and selective alkylation of DNA could inhibit the binding of the regulatory proteins to DNA. Several researchers have investigated the effect of adding alkylating groups [44] such as traditional nitrogen mustards [45] to α-halogenoacrylic [46] moieties by keeping the distamycin and netropsin frameworks intact. Cinnamic mustard (PNU 157911) and half-mustard (PNU 160666, Figure 3) derivatives of distamycin show excellent antileukemic activity and are found to be significantly less myelotoxic than TAM against murine and human hematopoietic progenitor cells [43]. The positively charged basic amidino side chain, responsible for electrostatic interaction with negatively charged DNA phosphate backbone, was also replaced by various amidine-like groups, such as cyanoamidine, *N*-methylamidine, *N*,*N*-dimethylamidine, and guanidino moieties either to increase the stability, cytotoxicity and enhance solubility at physiological pH. Comparable cytotoxicity was observed in these cases suggesting a general behavior of these classes of molecules including the amidine modification. In addition, a novel class of cytotoxic MGBs comprising of α-bromo or chloroacrylamide moieties linked to distamycin were identified. Among all different synthetic analogs, brostacillin (PNU-166196, Figure 3) was found to be a potent anticancer drug due to its improved cytotoxicity/myelotoxicity ratio [47,48]. Brostacillin acts as an effective DNA alkylator only in presence of high levels of cellular thiols such as glutathione [49]. Moreover, it was thirty-fold more active in comparison to TAM in inducing apoptosis in A2780 human ovarian carcinoma cells [43]. Khalaf et al. reported a new class of neutral, non-cationic minor groove binders derived from distamycin where the cationic tail group has been replaced by a neutral, polar variant including cyanoguanidine, nitroalkene, and trifluoroacetamide groups. These conjugates exhibit significant antibacterial activity against Gram-positive bacterial strains [50].

Several other distamycin analogs were synthesized by replacing one or more pyrrole rings with other heterocycles such as pyrazoles [51], benzofurans [52], thiazoles, thiophenes, imidazole and oxazoles [53] in order to establish a structure–activity relationship. It has been observed that the number and position of pyrrole rings are crucial for antileukemic activity. The presence of pyrrole rings close to the alkylating BAM moiety is responsible for better cytotoxic activity both in vitro and in vivo, whereas a pyrazole ring in close proximity to BAM drastically reduces the same as shown in the Figure 4 (**2** > **1** > **3**) [51]. Baraldi et al. designed and synthesized a series of novel compounds comprising different benzoheterocyclic rings, bearing a nitrogen mustard, a benzoyl nitrogen mustard or an α-bromoacryloyl group as alkylating moieties, tethered to a distamycin framework. Conjugate **4** (a 5-nitrogen mustard *N*-methylindole derivative) was found to exhibit excellent antileukemic activity with a very long survival time in comparison to tallimustine [52]. Khalaf et al. reported several heterocyclic trimeric distamycin analogs with enhanced lipophilicity [53]. These structural analogs comprise of branched *N*-alkyl- and *N*-cycloalkylpyrroles to test the conformational flexibility towards DNA binding. Hydrophobic N-terminal amides and substituted thiazole replacing pyrrole were installed in order to impart more lipophilicity.

All these compounds were shown to bind A·T-rich regions preferentially. The compounds containing branched *N*-alkylpyrrole, hydrophobic N-terminal amide, and especially *C*-isopropylthiazole (thiazotropsin A as shown in the Figure 4) showed significant antimicrobial activity against MRSA and *Candida albicans* strains. Thiazotropsin A has shown much higher affinity than parent distamycin A (preferential selectivity towards G·C sites) due to the presence of an isopropyl-substituted thiazole ring, which makes the molecule more hydrophobic [54]. Recently, a small set of analogs of thiazotropsin was designed and synthesized to study their solution-phase self-association characteristics and DNA molecular-recognition properties [17]. The authors showed a measurable difference in solution-phase self-assembly character with enhanced DNA association characteristics by replacing the formamide head group in thiazotropsin A with nicotinamide as shown in the Figure 4 (conjugate **5**). Suckling et al. further demonstrated another structural analog of thiazotropsin conjugate **6**, a heterocylic triamide containing thiazole carboxylic acid, which showed significant activity (MIC = 63 nM) against *Trypanosoma brucei* [55]. However, the authors reported other conjugates with two thiazoles directly linked via an amide bond, which retained activity to a lesser extent. Baraldi et al. designed and synthesized a novel conjugate **7** by combining naturally occurring antitumor agent distamycin A with the pyrrolo[2,1-*c*][1,4]benzodiazepine moiety (PBD), related to the naturally occurring anthramycin for investigating its antitumor activity [56]. Conjugate **7** demonstrated much better activity compared to distamycin in vitro by inhibiting cell growth of neoplastic cell lines and preferentially binding to G·C-rich sequences in the minor groove. In similar fashion, they further reported a series of novel hybrids by tethering distamycin A with the antineoplastic agent uramustine via

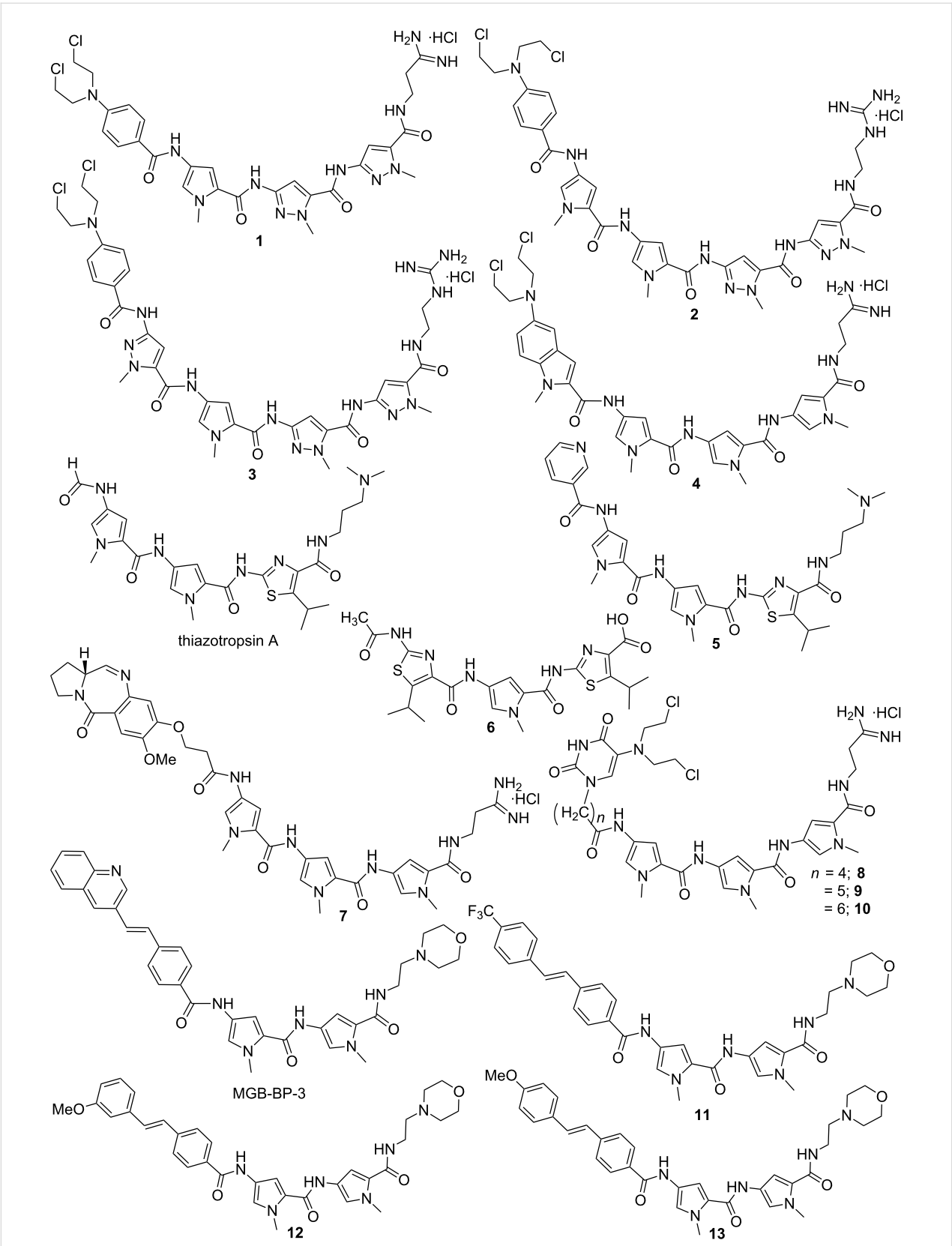


Figure 4: Synthetic structural analogs of distamycin A by replacing one or more pyrrole rings with other heterocycles or by tethering with known anti-tumor agents.

a flexible polymethylene chain of variable length (n = 1 to 6) in order to test their DNA binding affinity and cytotoxicity [57].

It has been observed that hybrid conjugates **8**, **9** and **10** with longer linkers exhibit relatively higher cytotoxicity in comparison to both distamycin and uramustine. The distamycin fragment directs binding to the A·T-rich sequences in the minor groove, and higher flexibility due to the longer linker allows optimal positioning of the mustard for DNA alkylation. In addition, longer linker imparts more lipophilicity, which in turn, favors better transportation of these compounds into the cells. Anthony et al. reported a series of short MGBs based on the lead compounds distamycin and thiazotropsins with the installation of hydrophobic aromatic head groups, including quinolyl and benzoyl derivatives, and alkenes as linkers in order to investigate their antimicrobial properties [58,59]. One of these structural analogs, MGB-BP-3 (Figure 4), containing a stilbene like fragment as head group and two *N*-methylpyrroles attached to an aminoethylmorpholine as tail group, was found to be extremely potent (MIC values in the range of 0.5–13 μg mL^{-1}) against several strains of *S. aureus*, both methicillin-sensitive and resistant strains. High antimicrobial activity, shown by this drug, was due to the presence of a hydrophobic head group with a hydrogen-bonding substituent (3-quinolinyl nitrogen forming a hydrogen bond with a guanine amino group at the base of the minor groove) and a low pK_a tail group. This drug was further selected for the treatment of Gram-positive bacteria *Clostridium difficile* infections and is currently in the phase II clinical trials. Szerszenowicz et al. developed a new set of potential minor groove binders derived from netropsin and bis-netropsin analogs by replacing *N*-methylpyrrole rings with other heterocyclic rings and their antiproliferative activity was tested on MCF-7 breast cancer cells [60]. Suckling et al. recently designed and synthesized a series of structurally diverse MGBs, derived from distamycin, in order to test their lung cancer inhibition activity against the melanoma cancer cell line B16-F10 [14]. Conjugate **11** was found to be extremely potent and exhibits 70-fold activity in comparison to the standard therapy, gemcitabine. Thus, the conjugate **11** was chosen for further development as an anti-lung cancer therapeutic. In the similar fashion, the same group investigated the correlation between DNA binding and antibacterial activity shown by these novel distamycin alkene-containing analogs (MGB-BP-3, **12** and **13,** Figure 4). This has been attributed to strong self-association (dimerization) in an antiparallel, head-to-tail orientation in aqueous solution during complex formation with duplex DNA oligomers verified via NOE experiments [61]. They further reported several structurally diverse MGBs, derived from distamycin, in order to probe their antifungal and antimycobacterial activity; several of these novel conjugates showed promising activity against the fungus *C. neoformans* (MIC_{80}s ranging from 0.25–4 μg/mL) and the mycobacterium *M. tuberculosis* (MIC_{99}s 3.1 μM) [62].

Since the last few decades, a plethora of synthetic structural analogs of distamycin, netropsin and thiazotropsins were developed to test their DNA binding affinity, sequence specificity and cytotoxicity, thereby eventually developing a general approach for the regulation of gene expression by DNA binding small molecules. However, all these analogs do not possess the ideal crescent shape required to wrap around the minor groove of DNA, which limit their efficacy to recognize longer stretches of DNA sequence. In order to achieve better sequence specificity, a series of oligomeric "hairpin (HP)" polyamides containing pyrrole and imidazole ring systems (Py/Im) were designed and synthesized by Dervan et al. and followed by other groups. It was observed that pyrrole/imidazole polyamides were able to bind side-by-side in the minor groove of DNA with high affinity and in a sequence-specific manner. Crystal structure studies confirmed the existence of a hydrogen bond between the Im nitrogen and the exocyclic amine of guanine. Dervan et al. have further developed rules for base pairing recognition of minor groove binding polyamides where antiparallel side-by-side pairings of pyrrole (py) and imidazole (Im) amino acids successfully distinguish G·C from C·G base pairs, and both of these from A·T/T·A base pairs as depicted in Figure 5 [63]. Again, a Py/Py pair specifies A·T from G·C but does not distinguish A·T from T·A. Thus, in order to break this degeneracy, Dervan et al. successfully introduced another aromatic amino acid, 3-hydroxypyrrole (Hp). With this subtle change by replacing a single hydrogen atom with a hydroxy group, hydroxypyrrole–imidazole–pyrrole polyamides form four ring pairings (Im/Py, Py/Im, Hp/Py and Py/Hp) and are able to distinguish all four Watson–Crick base pairs in the minor groove of DNA [64-66]. These polyamides are a successful class of synthetic DNA (minor groove) binders that can be designed to bind chosen DNA sequences via directed H-bonds, shape complementarity, and can compete with specific protein–DNA binding interactions in the minor or major grooves [67,68].

A variety of sequence-specific Py/Im polyamides were designed and synthesized in order to interfere with transcription factor binding and to regulate gene expression, both in vitro and in vivo. These polyamides are shown to bind DNA with comparable and/or even higher affinities than those of natural DNA-binding transcription factors. Dickinson et al. designed novel polyamides, which were able to bind adjacent to the recognition sites of a broad-range of transcription factors TBP, Ets-1, LEF-1 and NF-κB [69], thereby inhibiting binding of these transcription factors to DNA and ternary complex formation [70]. Dervan et al. has further introduced a novel Py/Im polyamide

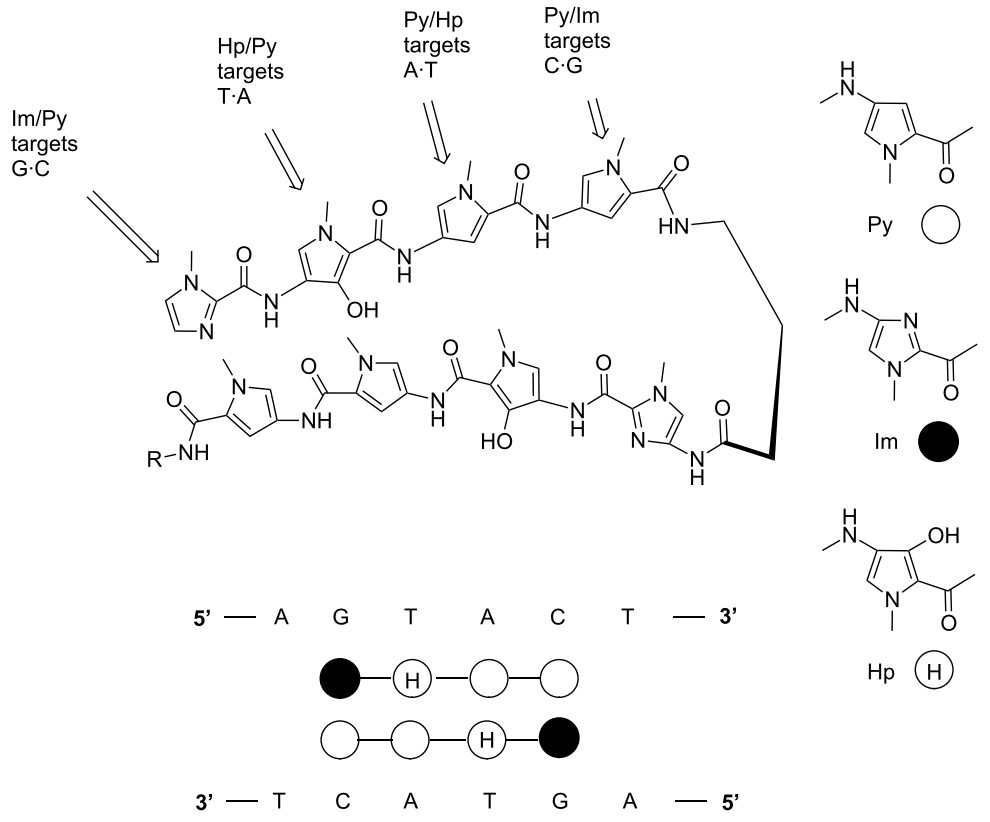


Figure 5: Pictorial representation of the binding model of pyrrole–imidazole (Py/Im) polyamides based on the base pairing recognition rule in the minor groove.

14 (Figure 6) that was able to bind preferentially the sequences 5′-WGGWWW-3′ and 5′ GGGWWW-3′ in the Nuclear factor κB sites, thereby reducing the expression of various NF-κB-driven genes including IL6 and IL8 [71]. Another structural analog of conjugate **11**, conjugate **15** was developed to interrogate its effect on the activity of RNA polymerase II [72]. Lenzmeier et al. provided strong evidence for inhibition of Tax protein–DNA minor groove interaction via synthetic Py/Im polyamides, which is believed to be essential for treating and/or preventing HTLV-I-associated diseases [73]. Gottesfeld et al. synthesized a series of Py/Im HP-polyamide–DNA alkylator (chlorambucil) (HP-Chl) conjugates in order to bind and alkylate within the HIV-1 promoter region, thereby blocking HIV-1 replication and screened them against human colon carcinoma cell lines [74,75]. It has been observed that conjugate **16** showed significant changes in cellular morphology and causes cells to arrest in the G2/M stage of the cell cycle. The authors further confirmed via microarray analysis that the histone H4c gene is significantly downregulated by the conjugate **16** which was assumed to be bound to and alkylate a site in the H4c promoter in treated cells, thereby inhibiting tumor growth in mice. Chenoweth and Dervan showed DNA structural distortion induced by an 8-ring cyclic Py/Im polyamide (conjugate **17**) bound to the central 6 bp of the sequence d(5′-CCAGGC-CTGG-3′)$_2$ by using a high resolution X-ray crystal structure as shown in Figure 7a [76]. This allosteric perturbation of the DNA helix by small molecules through binding at distinct locations on promoter DNA provides a clear understanding of how transcription factor activity could be disrupted and gene expressions could also be regulated. In order to target the inverted CCAAT box (ICB) of the human multidrug resistance 1 gene (MDR1) promoter and to distinguish between different promoter ICB sites, several ICB-containing DNA hairpin polyamides were designed with different flanking base pairs. It was confirmed via thermal-denaturation studies and DNase I-footprinting assays that one of these conjugates containing a 3-methylpicolinate moiety (ZT65B, compound **18**) binds in the minor groove and effectively targeted ICBa and ICBb, similar to the 3′-ICB site of MDR1 (TGGCT) [77].

Lai et al. synthesized the novel hairpin Py/Im polyamide conjugate **19** and a mismatch conjugate in order to target -545 to -539 base pairs of human transforming growth factor-beta1 (hTGF-beta1) promoter and diminish the gene and protein expression

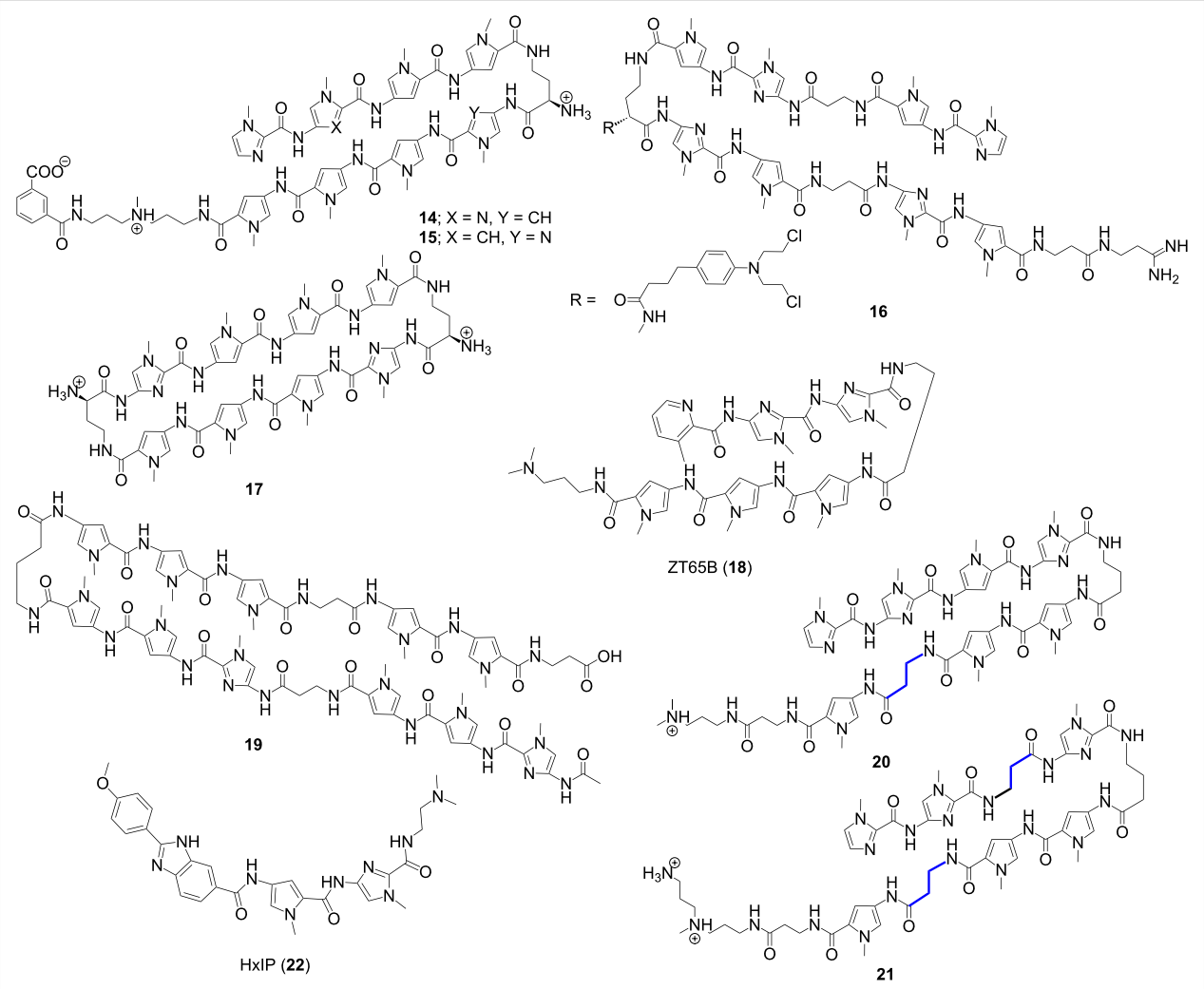


Figure 6: Chemical structures of synthetic “hairpin” pyrrole–imidazole (Py/Im) conjugates.

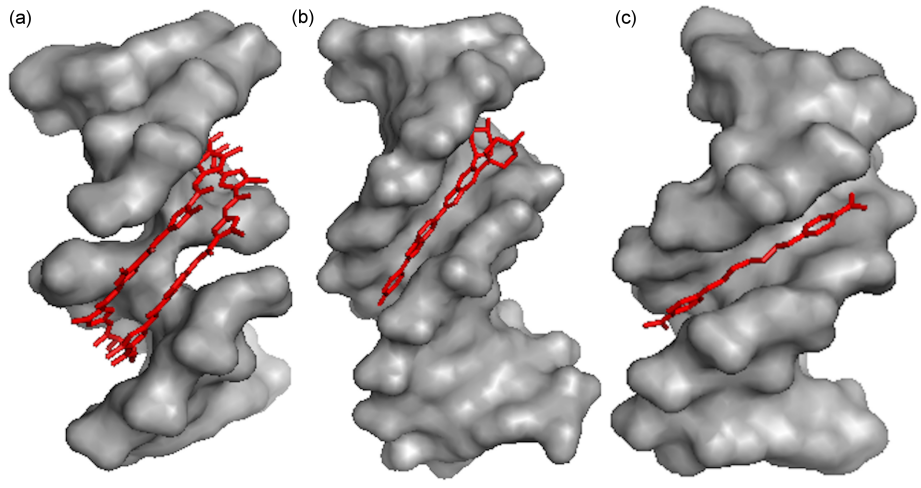


Figure 7: (a) Minor groove complex formation between DNA duplex and 8-ring cyclic Py/Im polyamide (conjugate **17,** PDB ID# 3I5L) [76]; (b) Complex between DNA duplex and Hoechst 33258 (PDB ID# 8BNA) [82]; (c) Crystal complex between DNA duplex and pentamidine (PDB ID# 3EY0) [83].

[78]. The authors went on to confirm that conjugate **19** binds its corresponding target sequence whereas the mismatch conjugate fails to recognize the sequence by using a gel mobility shift assay. Additionally, conjugate **19** drastically inhibited the promoter activity of hTGF-beta1 as well as gene and protein expression as determined via in vitro transcription experiments and luciferase assay. This research paved the way for a novel gene therapy for the treatment of TGF-beta-related diseases. Several researchers have installed a flexible β-alanine fragment on Py/Im HP polyamides for better recognition of DNA by reducing molecular rigidity. However, in order to get a better understanding of how β-substitution diversely affects the HP–DNA binding affinity, selectivity, and especially kinetics, Wilson and co-workers conducted a thorough study by synthesizing eight heterocyclic HP polyamides having single and double β-substituted derivatives with their cognate and mutant sequences [79]; two of the representative conjugates **20** and **21** are shown in the Figure 6. In conclusion, the authors reported that β-substituted polyamides weakens the binding affinity of these conjugates with cognate DNA and drastically influence the binding kinetics such as association and dissociation rates in a position- and number-dependent manner. The authors, in addition, replaced the monocationic Dp group [3-(dimethylamino)propylamine] in conjugate **20** with a dicationic Ta group (3,3′-diamino-*N*-methyldipropylamine) in conjugate **21** to minimize the frequently observed polyamide aggregation. This subtle modification retains the polyamide-DNA binding mode and affinity by reducing aggregation and also helps to conduct a detailed thermodynamic study for the 8-ring HP polyamides for the very first time. Recently, Hartley et al. designed and synthesized a hybrid fluorescent HP polyamide conjugate **22** (Figure 6) by attaching the A·T recognizing fluorophore, *p*-anisylbenzimidazolecarboxamido (Hx) in order to target the inverted CCAAT box 2 (ICB2) of the topoisomerase IIα (topo IIα) promoter and to monitor the cellar uptake of the conjugate [80,81]. Gratifyingly, conjugate **22** targets the 5'-TACGAT-3' sequence of the 5' flank of ICB2 with high affinity and sequence specificity, thereby disrupting the NF-Y-ICB2 interaction. In addition, cellular uptake and nuclear localization of conjugate **22** could be easily monitored as a result of its inherent fluorescence property.

Despite myriad important biological roles of hairpin and cyclic Py/Im polyamides in regulating natural gene expression via sequence-specific DNA binding, the lack of viable strategies for facile synthesis of library of structural variants of these classes of conjugates remains a huge challenge for the researchers. In order to resolve this issue, Dervan et al. recently published a modular microwave-assisted Fmoc-based solid phase synthetic approach for the syntheses of cyclic Py/Im polyamides [84]. This group previously optimized and reported a machine-assisted Fmoc solid phase synthesis of simpler polyamides to afford high step-wise coupling yield [85]. A seven-member library of cyclic polyamides targeting androgen response element (ARE) and the estrogen response element (ERE) was synthesized in 12–17% overall yield. Selective modifications could also be done on the GABA turn units, which showed improved cellular uptake properties.

Sugiyama et al. designed and synthesized a series of telomere-targeting synthetically challenging tandem hairpin Py/Im polyamides which could recognize >10 base pairs with flexible linker conjugated with a fluorescent dye (either Texas Red (TR) or Cyanine 3 (Cy3)) using a Fmoc-based solid phase synthetic approach; two of the representative conjugates **23** and **24** are shown in the Figure 8 [86,87]. The authors investigated the binding affinity and sequence specificity of these conjugates for the human telomeric repeat TTAGGG in mouse MC12 and human HeLa cells. In mouse and human cells, TR-conjugated polyamides **23** and **24** successfully targeted to the corresponding telomeres and highlighted the telomere foci clearly because of their fluorescent nature. Later on, the authors successfully designed tandem tetramer Py–Im polyamides with 4 hairpins and 3 hinges targeting 24 bp of the human telomere sequences [88]. Thus, the authors set the new record for the longest binding site of synthetic, non-nucleic-acid-based, sequence-specific DNA-binding molecules. These conjugates could bind to four telomeric repeats with nanomolar dissociation constants, confirmed via SPR analysis. In the similar fashion, Nozeret et al. reported a series of nine fluorescent hairpin polyamides by attaching cyanine and fluorescein dyes to target mouse major satellite DNA using thermal denaturation, gel-shift electrophoresis, circular dichroism and fluorescence spectroscopy [89,90]. Some of these fluorescent probes were found to detect target sequences in mouse living cell lines and the nuclear substructures formed by repeated DNA sequences in living cells were nicely visualized. Choice of fluorophores attached to the N-terminus of the polyamides remains extremely crucial, as they seem to affect DNA minor groove binding significantly. In order to design a novel DNA cleaving agent, a bis(guanidinium)alcohol tethered with Dervan hairpin polyamide was synthesized. The resulting conjugate **25** binds A·T-rich DNA duplexes with comparable affinity to that of the parent polyamide and breaks one strand of double-stranded plasmid DNA by interacting with anionic phosphodiesters in a fast transphosphorylation step as contact ion pairs at micromolar to high nanomolar concentration range [91]. Richert et al. designed a novel set of three-pronged probes (TPPs) comprising of cap, β-alanines and oligopyrrolamides in order to bind A·T-rich target strands from three sides (Watson–Crick face, terminus, and minor groove) resulting in exceptionally stable duplexes (ΔT_m = +44.8 °C) and high selectivity [92].

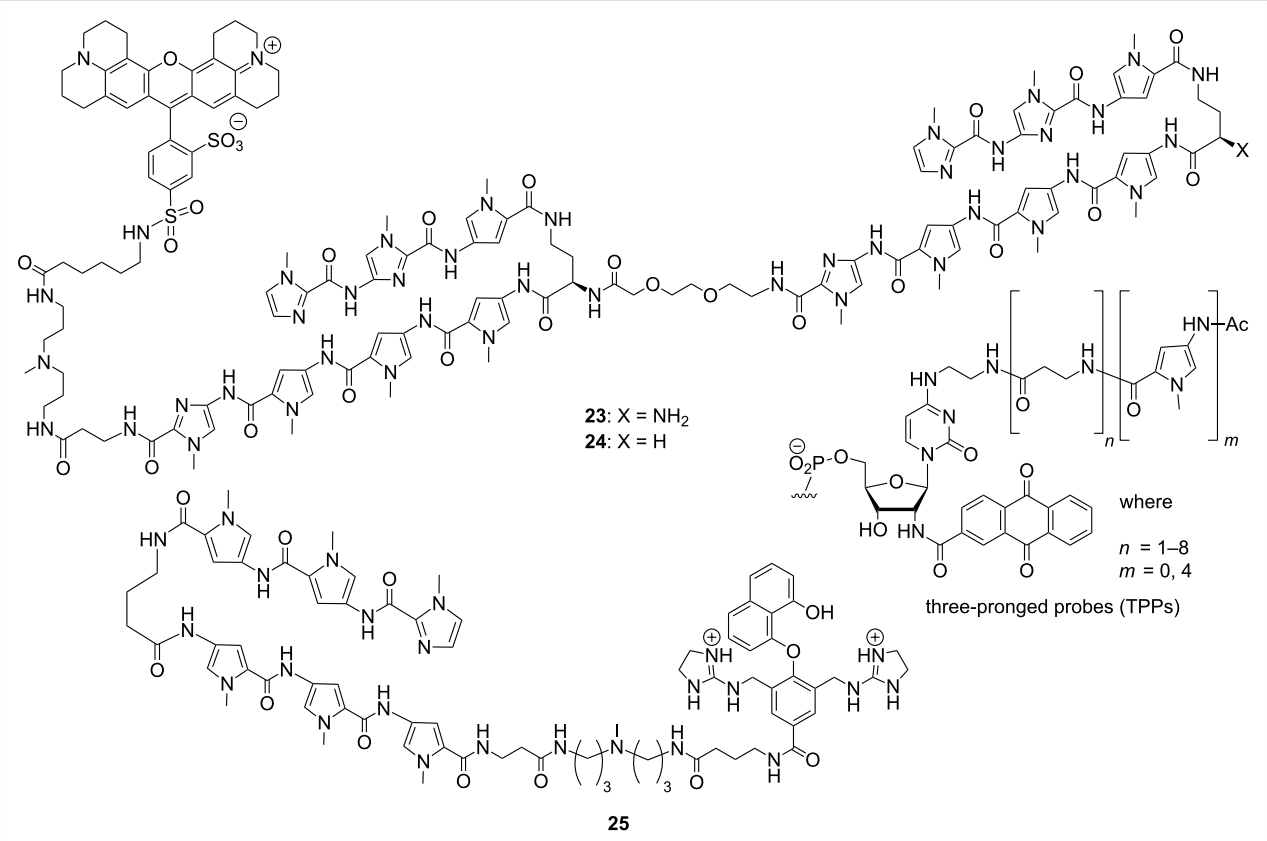


Figure 8: Telomere-targeting tandem hairpin Py/Im polyamides **23** and **24** capable of recognizing >10 base pairs; **25**: a novel DNA-cleaving agent comprising of bis(guanidinium)alcohol tethered with Dervan hairpin polyamide; three-pronged probes (TPPs) to target A·T-rich sequences selectively.

Six novel 4-aminoantipyrine derived Schiff bases and their metal complexes with Cu(II), Ni(II), Zn(II) ions (conjugates **26**–**31**) were synthesized and characterized and binding of these complexes with ct-DNA were analyzed by electronic absorption spectroscopy, viscosity measurement, cyclic voltammetry and molecular modeling (Figure 9) [93]. Docking results confirmed that these complexes have the ability to interact with the minor groove of the ct-DNA. In addition, the authors confirmed that in presence of ascorbic acid, these complexes could facilitate DNA cleavage. Moreover, these complexes showed improved biocidal activity than the free ligands against various bacterial strains such as *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Staphylococcus epidermidis*, and *Klebsiella pneumonia*. Nair et al. synthesized and characterized three mononuclear copper(II) complexes, $[Cu(tpy)Cl_2]$, $[Cu(tpy)(NO_3)_2(H_2O)]$ and $[Cu(Ptpy)Cl_2]·H_2O·HCl$ and investigated their cytotoxicity and primary mode of DNA binding mechanism [94]. Molecular modeling as well as DNA cleavage studies have revealed that the first two complexes are DNA minor groove binders, whereas the third complex prefers an intercalative mode of binding to DNA. All these complexes show nuclease activity in the presence of hydrogen peroxide and induce apoptosis to human A549 lung adenocarcinoma cells. A series of novel glyco-oligoamides (Figure 9) has been designed and synthesized in order to investigate the molecular basis of carbohydrate–minor groove DNA interactions by Vicent et al. [95].

NMR spectroscopy and molecular modeling studies further confirmed the existence of directional intramolecular hydrogen bonds and CH–π interactions, which results in stabilizing these conjugates in the minor groove by marinating a stable hairpin structure [96]. The authors tethered various monosaccharides such as β-xylose, α-xylose, β-galactose, β-glucose and β-L-fucose to a minor groove binding residue, Py-γ-Py-Ind, structurally analogous to distamycin and netropsin. A new set of novel anthraquinone–chalcone hybrids were synthesized using Claisen–Schmidt reaction in order to test their anticancer potential against human cancer cell lines and DNA binding affinity and specificity. It has been observed that three conjugates **32**–**34** exhibited significant cytotoxicity against LS174 and HeLa cancer cell lines by interacting non-covalently with the minor groove of the double helical ct-DNA [97]. Barker et al. have designed a series of novel di- and triaryl benzamide MGBs differing in the polar side chain, bonding and substitution patterns and functionalization of benzylic substituents and eval-

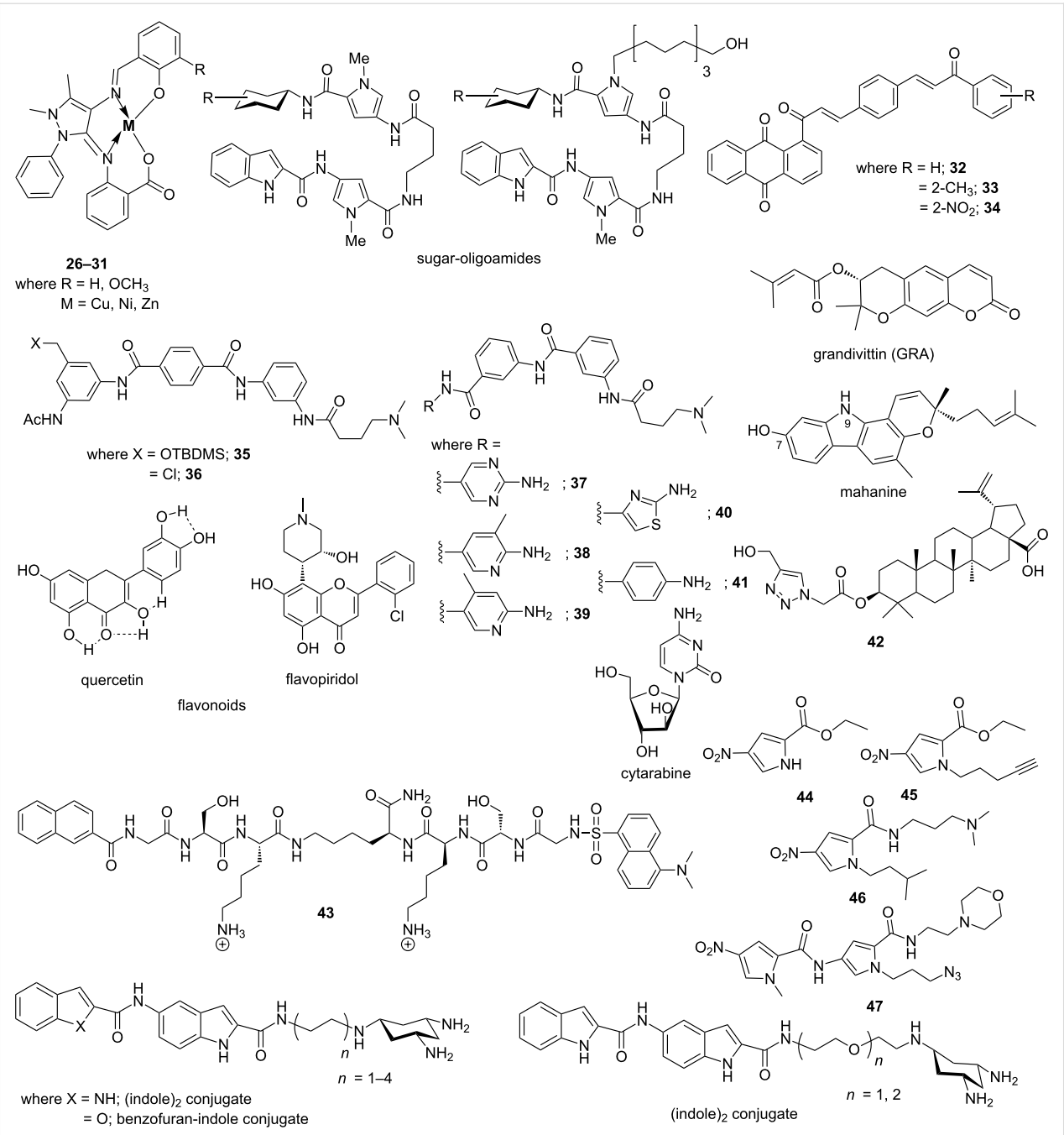


Figure 9: Representative examples of recently developed DNA minor groove binders.

uated their antiproliferative activity as well as their DNA binding affinity [98]. It has been confirmed that the most active conjugates are unsymmetrical triaryl benzamides **35** and **36** comprising of a bulky and alkylating chlorobenzylic substituent, respectively, and a polar amino side chain. Conjugate **35** with a bulky OTBDMS benzylic substituent was found to be the most active agent with (IC_{50} 5.0 μM) followed by conjugate **36** with a chloro substituent (IC_{50} 9.9 μM). Drozdowska et al. reported a series of distamycin analogues **37**–**41** (Figure 9) as potential minor groove binders and their minor groove DNA binding affinity as well as antiproliferative effects on human MCF-7 breast cancer cells were evaluated [99]. These conjugates bind within the minor groove of B-DNA. They inhibited catalytic action of endonucleases in A·A, A·T, T·T and A·G restriction sites but failed to block G·C-rich sequences. In addition, they act as potent topoisomerase II inhibitor at the concentration 10 μM and show antiproliferative and cytotoxic activities in breast cancer cell line in the range of 81.70 μM and 200.00 μM.

Conjugate **41** with a 6-aminophenyl moiety appeared to be the most effective among others. Suckling et al. designed a set of 31 Strathclyde minor groove binders (S-MGBs), derived from distamycin, by varying the head groups (amidine, amide, or alkene), heterocyclic building blocks and their alkyl substituents and the basicity of the C-terminal tail group in order to investigate their antimalarial activity against a chloroquine sensitive (3D7) and resistant (Dd2) strain of *Plasmodium falciparum* [100]. Conjugates with an alkene link between the two N-terminal building blocks and a *C*-alkylthiazole moiety appeared to the most active among others with IC_{50} values in the range of 30–500 nM. The same group further demonstrated that the head group plays a crucial role in determining the activity against *Trypanosoma brucei* with another set of novel S-MGBs, structurally analogous to distamycin [101]. Coumarins are a group of phenolic compounds with excellent cytotoxic and antiviral properties. Again, dihydrofuranocoumarins are another class of coumarins possessing anticancer activities. Recently, Ahmadi et al. identified several dihydrofuranocoumarins, especially grandivittin (GRA), from *Ferulago macrocarpa* (Fenzl) Boiss., and their mechanism of minor groove DNA binding and antibacterial, cytotoxic and antioxidant activities were evaluated [102]. A molecular docking study has revealed that GRA interacts with ct-DNAs via hydrogen bonding interactions between the oxygen atoms of GRA and adenine bases of DNA and van der Waals interactions. Moreover, GRA significantly reduces the polymerization activity of DNA polymerase as a result of binding to minor groove DNA. Samanta et al. investigated a thorough structure–activity correlation between mahanine, an anticancer carbazole alkaloid, and its chemically modified analogs to test the role of various functional groups on its antiproliferative activity against 19 cancer cell lines [103]. It has been shown that the C-7 hydroxy and the 9-NH group showed significant contribution towards its DNA minor groove binding ability via strong association with the phosphate backbone. In addition, the presence of these functional groups could enhance antiproliferative activity of cancer cells towards apoptosis through the mitochondrial pathway. Mitrasinovic has reported sequence-dependent binding of various structurally different flavonoids (quercetin (QUE) and flavopiridol (FLP)), a family of prospective anticancer agents, to duplex DNAs [104]. The five hydroxy groups in QUE involve in the intramolecular hydrogen bonding which is attributed to its planar orientation whereas the chlorophenyl moiety, the heterocyclic fragment with the C5 and C7 hydroxy groups and C8 piperidinyl substituent in FLP favor non-planar binding geometry. The author examined their sequence-specific binding affinity using sophisticated molecular dynamics approach with eight different nucleotides having variety of sequences. It has been observed that QUE appears to be a minor groove binder, whereas FLP involves in combined mode of interaction such as minor groove binding and intercalation. A set of betulinic acid analogs were synthesized by using azide–alkyne click reaction and their anticancer activities against different cancer cell lines and normal human PBMC cell line were evaluated by MTT assay. Conjugate **42** was found to be extremely potent against HT-29 cell line with an IC_{50} value of 14.9 μM and its cytotoxicity was attributed to DNA minor groove binding ability [105]. Recently, Schmuck et al. have developed a first prototype of cationic oligopeptide-based molecular beacon (conjugate **43**) coupled with a FRET pair, a naphthalene donor and a dansyl acceptor, for ratiometric detection of ds-DNA by fluorescence microscopy with preference for A·T-rich sequences [106]. Two positively charged lysine residues are expected to interact with ds-DNA electrostatically. Upon binding to the minor groove of ds-DNA, the conformation of conjugate **43** was changed from an extended to a folded form, thereby changing the efficiency of the FRET process between the two fluorophores and exhibiting a significant red shift in the emission spectrum. Moreover, the conjugate **43** could be used as an attractive tool for imaging of nuclear DNA in the cells due to its low cytotoxicity. A series of water-soluble peptidocalix[4]arenes with arginine-rich short narrow groove binding residues on the lower rim of the calix[4]arene scaffold were reported by Soltani et al. in order to study the binding between well-matched and mismatched DNA duplexes [107]. Fluorescent titrations, ethidium bromide (EB) displacement assays, DNA-melting experiments, and circular dichroism (CD) analysis revealed these conjugates are high affinity sequence specific DNA groove binders and could successfully recognize a C·C mismatch in a DNA duplex. Recently, the binding mechanism of the anticancer drug cytarabine with calf thymus DNA (ct-DNA) was investigated in vitro by Shahabadi et al. by multispectroscopic techniques and molecular modeling study [108]. It has been shown that cytarabine acts in a groove-binding mode, which was confirmed by fluorescence experimental results of Hoechst 33258 displacement by the drug. Hydrophobic interactions play a crucial role in its binding to DNA groove. Similarly, the same group recently reported a macrocyclic copper(II) complex, ($[CuL(ClO_4)_2]$ where L is 1,3,6,10,12,15-hexaazatricyclo[13.3.1.1^{6,10}]eicosane) and studied its interaction with calf thymus DNA (ct-DNA). It was confirmed that the Cu(II) complex could displace the ct-DNA-bound Hoechst33258 suggesting it binds to the minor groove of ct-DNA via groove binding mechanism [109]. Suckling et al. have recently reported four nitro-pyrrole-based compounds (conjugates **44**–**47**, Figure 9) as building blocks for the synthesis of novel minor groove binders [110]. Crystal structure data revealed that nitro groups and ester moieties in conjugates **44** and **45** are coplanar with the pyrrole ring, whereas the isopropyl fragment in conjugate **46** lies out of the pyrrole plane. Coplanarity extends to the second pyrrole ring in case of conjugate **47** and all these conjugates form layer-like structures

during crystal formation via multiple hydrogen bonding interactions. This structural information indeed helps to design novel MGBs with much better binding affinity and specificity. A new family of conjugates between a Zn(II)-tach complex and $(indole)_2$ or benzofuran–indole amide minor groove binders connected through alkyl or ethoxyethyl linkers were developed by Tecilla et al. [111]. The authors confirmed that these conjugates with tach units, either free or Zn(II)-complexed forms, bind strongly to the minor groove through electrostatic interactions with the phosphate backbone and the binding affinity strongly depends upon the nature and length of the linkers.

2.2. Bisbenzimidazoles

Bisbenzimidazoles are one of most extensively studied DNA minor groove binding compounds; Hoechst 33258 and 33342 are representatives of this class of compounds as shown in Figure 10. Minor groove complex formation between DNA duplex and Hoechst 33258 is shown in Figure 7b [82]. X-ray crystallographic and NMR studies confirmed that Hoechst

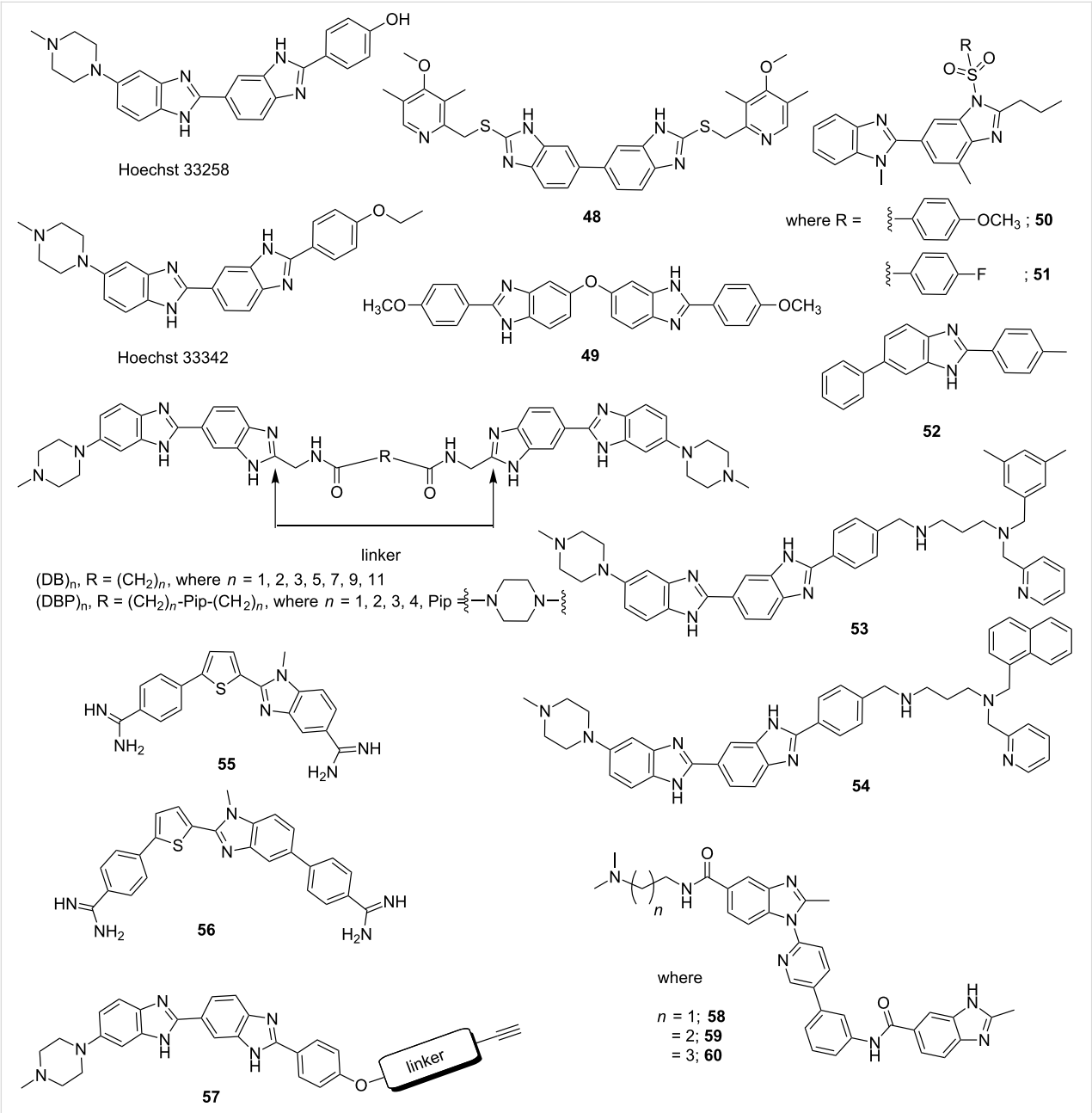


Figure 10: Chemical structures of bisbenzamidazoles Hoechst 33258 and 33342 and their synthetic structural analogs by varying substitutions and linkers.

33258 binds to the A·T-rich sequences in minor groove with the planar benzimidazole groups are oriented parallel to the direction of the groove. Hoechst 33258 primarily acts as human topoisomerase I poison [112] and initially showed cytotoxicity against L1210 murine leukemia; however, after passing human phase I clinical trials for pancreatic cancer, it failed to produce any effective result in phase II trials [113]. However, due to its high binding affinity to B-DNA duplexes, several groups have designed various structural analogs of Hoechst 33258 in order to achieve a better sequence-specific DNA binder with reduced toxicity [114]. Yang et al. reported a series of novel symmetrical bisbenzimidazoles as DNA minor groove binders. A molecular modeling study confirmed that conjugate **48** could dock into the minor groove of DNA. These conjugates exhibited cytotoxic activities on SKOV-3, HeLa, and BGC-823 cell lines in vitro in the single-digit micromolar range [115]. Another set of bisbenzimidazoles was synthesized by varying substitutions on the phenyl ring where the two benzimidazoles were linked via an oxygen atom. Most of these conjugates showed significant antitumor activity in vitro compared to Hoechst 33258. Amongst them, conjugate **49** (Figure 10) was found to be most potent with IC_{50} values of 0.56 μM for HL60 (Human promyelocytic leukemia cells) tumor cell line and 0.58 μM for U937 (Human leukemic monocyte lymphoma cells) tumor cell line with reduced toxicity in comparison to paclitaxel and 5-FU [116]. Ivanov et al. reported two different sets of strong minor groove binders, derived from well-known DNA minor groove binder Hoechst 33258. These conjugates are fluorescent dimeric bisbenzimidazoles [$(DB)_n$ and $(DBP)_n$] tethered by oligomethylene linkers of varied lengths with or without a central 1,4-piperazine residue [117]. The low solubility of $(DB)_n$ in aqueous solution due to aggregation has forced the authors to introduce a 1,4-piperazine residue in the oligomethylene linkers $(DBP)_n$, making them tetracations instead of dications for $(DB)_n$ at neutral pH. By the virtue of their higher solubility in aqueous media, $(DBP)_n$ could easily penetrate cell and nuclear membranes of living cells and inhibit in vitro eukaryotic DNA topoisomerase I and prokaryotic DNA methyltransferase (MTase) at micromolar concentrations. Rangappa et al. recently reported the synthesis of a series of novel bisbenzimidazole derivatives and evaluated their antiproliferative and antiangiogenic activity properties. Conjugates **50** and **51** were found to be not only potent antiproliferative agent against HeLa, HCT116 and A549 cells, but also did not exhibit cytotoxicity towards non-diseased (Vero) cells [118]. In addition, the authors tested the efficacy of these two lead conjugates **50** and **51** against Ehrlich ascites tumor (EAT) bearing mice for its antitumor and antiangiogenic properties and concluded that these conjugates drastically reduced the cell viability, body weight, ascites volume and downregulated the formation of neovasculature and production of Vascular Endothelial Growth Factor (VEGF).

They further reported another novel benzimidazole derivative conjugate **52** which could inhibit topoisomerase II activity and in vitro transcription by binding to the DNA minor groove [119]. Conjugate **52** could successfully exhibit cytotoxicity in leukemic cells by inducing apoptosis. Amirbekyan et al. reported a novel groove binding anchoring strategy for DNA-based asymmetric catalysis by synthesizing various structural analogs of Hoechst 33258. It has been observed that amine analogs (conjugate **53** and **54**) showed higher affinity towards ct-DNA and poly[$d(A{\cdot}T)_2$] in comparison to alkyne analogs with reduced flexibility and one less charged nitrogen atom, thereby reducing strength of electrostatic interactions between the ligands with DNA phosphate backbone [120]. Wilson et al. rationally designed benzimidazole derivatives by keeping pre-organized *N*-methylbenzimidazole (*N*-MeBI)-thiophene as central fragment (conjugates **55** and **56,** Figure 10) in order to selectively bind mixed G·C and A·T sequences of DNA. They hypothesized that thiophene (positive electrostatic potential) and the electron-donor nitrogen of N-MeBI should pre-organize the conformation for accepting hydrogen bond from $G\text{-}NH_2$, which was validated by replacing the thiophene moiety with other heterocycles, resulting in lowering the binding affinity and specificity [121]. Arya et al. reported a series of Hoechst 33258 based mono- and bisbenzimidazole derivatives and their *E. coli* DNA topoisomerase I inhibition, binding to B-DNA duplex, and antibacterial activity has been evaluated [122]. It has been observed that the conjugates with alkynyl side chains show excellent *E. coli* DNA topoisomerase I inhibition properties with IC_{50} values of <5.0 μM, which was attributed to critical interactions between the inhibitor side chain and amino acids of the active site of DNA topoisomerase I, as suggested by the modeling study. In general, bisbenzimidazole derivatives (conjugate **57**) exhibit much better antibacterial activity than mono-benzimidazoles for Gram-positive strains. More importantly, the linker lengths and composition have dramatic influence on DNA binding and cell uptake, suggesting that the roles of the linkers should be carefully investigated when combining fragments in drug discovery applications [123]. Recently, Picconi et al. reported a series of nontoxic triaryl benzimidazole conjugates derived from existing classes of MGBs, to probe their antibacterial activity against multidrug resistant (MDR) Gram-positive and Gram-negative species; conjugates **58**–**60** (Figure 10) showed excellent antibacterial activity with MICs ranging from 0.5–4 μg/mL for Gram-positive strains and MICs ranging from 16–32 μg/mL for Gram-negative strains [124]. However, molecular modeling revealed that these conjugates could not bind into the minor groove due to change in their conformation, thereby showing negligible DNA binding. Thus, their antibacterial activity is not attributed to DNA binding affinity due to lack of DNA stabilization by these conjugates.

2.3. Bisamidines

One of the oldest known clinically relevant small molecule MGBs with immense biological applications is the aryl bisamidine class related to diminazene, DAPI and pentamidine as shown in Figure 11. Minor groove complex formation between DNA duplex and pentamidine is shown in Figure 7c [83]. These small molecules are known to bind A·T-rich sequences preferentially. Moreno et al. reported a coiled-coil structure formed by the complex of the DNA duplex with pentamidine. The authors showed that the central part of the pentamidine binds to the minor groove, whereas the charged terminal amidine groups interact electrostatically with negatively charged phosphates, thereby stabilizing the complex through the formation of cross-links between neighboring duplexes [83]. However, due to intrinsic toxicity, various structural analogs of pentamidine were designed over the years by replacing the ether linkage with bis-amide **61** [125], introducing heterocyclic rings such as furan **62** and **63** [126], thiophene **64** [127] and pyridine **65** (Figure 11) [128].

These conjugates exhibit potent antibacterial and antiprotozoal activity with much reduced toxicity. It was further concluded that π-stacking, H-bonding with the floor of the minor groove along with appropriate curvature helps them to bind with specific DNA sequence [129]. A series of arylimidamide analogues were synthesized and their binding affinities towards DNA minor groove was studied by Wilson et al. via fluorescence displacement titration, circular dichroism, DNase I footprinting, biosensor surface plasmon resonance, X-ray crystallography and molecular modeling [130]. These experiments revealed that these novel conjugates form 1:1 complexes with A·T sequences in the DNA minor groove, and the binding strength depends upon substituent size, charge and polarity. In addition, they have also exhibited improved uptake properties in *Leishmania*

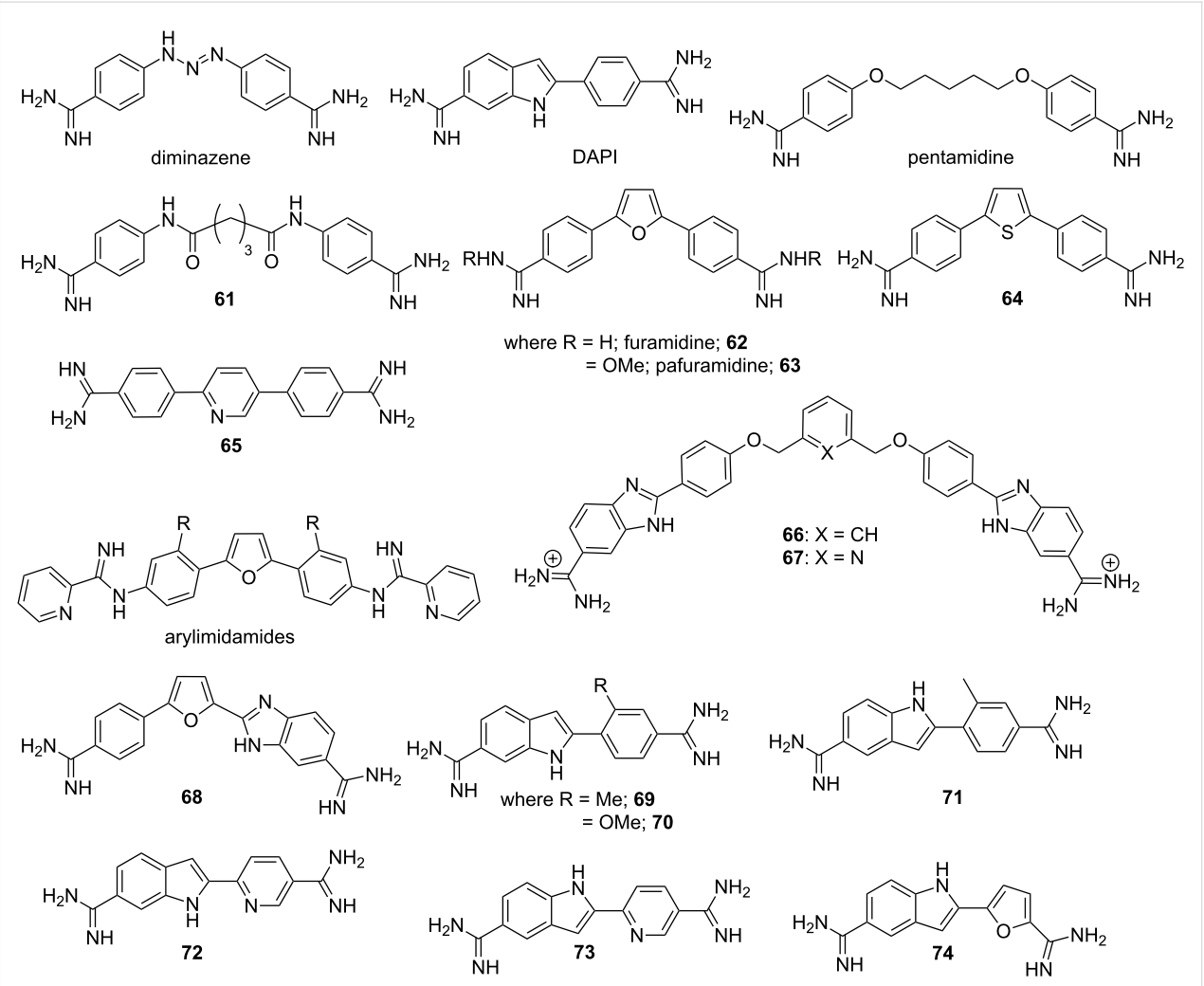


Figure 11: Chemical structures of bisamidines such as diminazene, DAPI, pentamidine and their synthetic structural analogs by varying substitutions, linkers and introducing heterocycles.

and *Trypanosoma cruizi* than existing heterocyclic diamidines. With this success, this group further rationally designed several other minor groove binders in order to achieve even better specificity, which could bind to two A·T sites separated by G·C base pairs. Molecular modeling and other biophysical studies confirmed that the conjugate **67**, pyridyl analog of conjugate **66**, could successfully recognize a single G·C base pair flanked by A·T sequences via several van der Waals and hydrogen bonding interactions [131]. Wilson et al. further designed a novel dicationic diamidine (conjugate **68**) to recognize a mixed base pair site for the first time. It has been confirmed via ESIMS that the conjugate **68** binds in the minor groove of ATGA sequences as a dimer with positive cooperativity [132]. Recently, they reported a series of structural analogs of DAPI by replacing the phenyl ring with substituted phenyl and heterocyclic rings as shown in the Figure 11. Amongst them, conjugates **69**–**74** are found to bind in the minor groove with improved affinity. Additionally, these conjugates exhibit superior in vitro antitrypanosomal activity in comparison to DAPI itself [133].

Rozas et al. designed and synthesized a new family of asymmetric peptide-linked diaromatic dications with a linear core as potent DNA minor groove binders (Figure 12) [134]. Various biophysical experiments such as surface plasmon resonance and circular dichroism revealed that due to the presence of a planar amide linker between the phenyl rings, these newly synthesized bis-cationic ligands (conjugates **75**–**77**) showed a much improved preferential minor groove binding ability towards A·T-rich regions in comparison to other guanidinium-like derivatives with curved cores. Dardonville reported a series of high affinity DNA minor groove binders N-substituted bisimidazo-

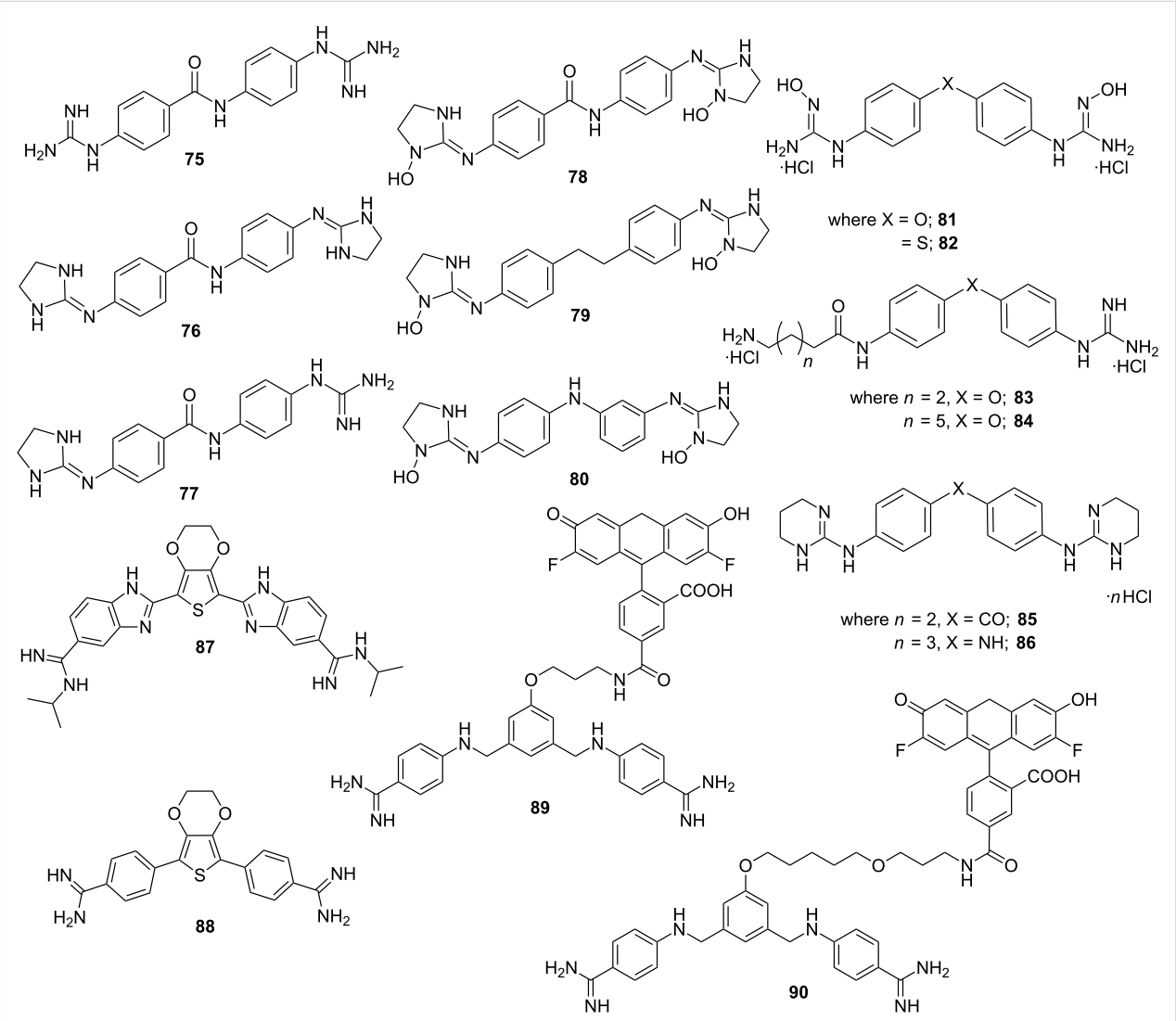


Figure 12: Representative examples of recently developed bisamidine derivatives.

line arylamides to test the effect of imidazoline ring N-substitution on preferentially binding at A·T sites over G·C sites [135]. The authors demonstrated N1 hydroxylation could enhance DNA binding affinity and selectivity towards AATT sites over $(A{\cdot}T)_4$ sequences (conjugates **78**–**80**). Rozas et al. further reported the syntheses of a new family of hydroxyguanidinium aromatic derivatives as potential minor groove binders and cytotoxic agents; two of the representative structures **81** and **82** are shown in the Figure 12. These conjugates showed antiproliferative effects in human promyelocytic HL-60, breast carcinoma MCF-7, and neuro-blastoma cell lines, although no direct correlation between their cytotoxicity and DNA binding affinity was established yet [136]. With the initial success, they reported DNA minor groove binding aminoalkyl derivatives of diaromatic guanidines **83** and **84**, which exhibit significant antiprotozoal activity in vitro against *P. falciparum* and *T. b. rhodesiense* strains [137]. Moreover, the authors further developed a new family of dicationic bis-2-amino-1,4,5,6-tetrahydropyrimidines with more suitable size and lipophilicity to bind in the minor groove than the previously reported conjugates [138]. Thermal denaturation experiments and DFT calculations revealed that conjugates **85** and **86** appeared to be much better binders than bis-guanidiniums, but weaker in comparison to bis-2-aminoimidazolinium derivatives as reported earlier [139]. Recently, a series of novel amidine derivatives of 3,4-ethylenedioxythiophene with excellent antibacterial activities against Gram-positive (including resistant MRSA, MRSE, VRE strains) and Gram-negative bacterial strains has been reported [140]. The bisbenzimidazole derivatives (conjugate **87**) exhibited the widest spectrum of activities whereas bis-phenyl derivatives were the most potent ones (conjugate **88**). In addition, these conjugates demonstrated excellent DNA binding ability (ΔT_m = 15.4 °C) through various electrostatic and hydrogen bonding interactions. Bordello et al. designed two fluorescence-labeled bisbenzamidine (BBA) derivatives (conjugates **89** and **90,** Figure 12) tethered with the dye Oregon Green (OG) separated via linkers of various lengths in order to develop highly sensitive sequence-specific DNA binders [141]. Detailed photophysical analysis revealed that these conjugates enforce a significant fluorescence enhancement upon binding to the minor groove of ds-DNA with excellent sequence specificity and reduced affinity constants in comparison to the parent BBA without the dye. Recent work from the Poon and Wilson groups has also shown how these designed amidines can be used to target TF activity [142].

2.4. Aureolic acid group of anticancer drugs

The antineoplastic and antibiotic natural products mithramycin (MTM) and chromomycin act as minor groove binder with the preference for G·C-rich sequences and represent aureolic acid group of anticancer drugs (Figure 13) [114]. Aich and Dasgupta established two different types of mithramycin-Mg^{2+} complex formation by which MTM exhibits its cytotoxic effect by interacting with DNA minor groove as a divalent metal coordinated dimer, thereby regulating gene expression [143].

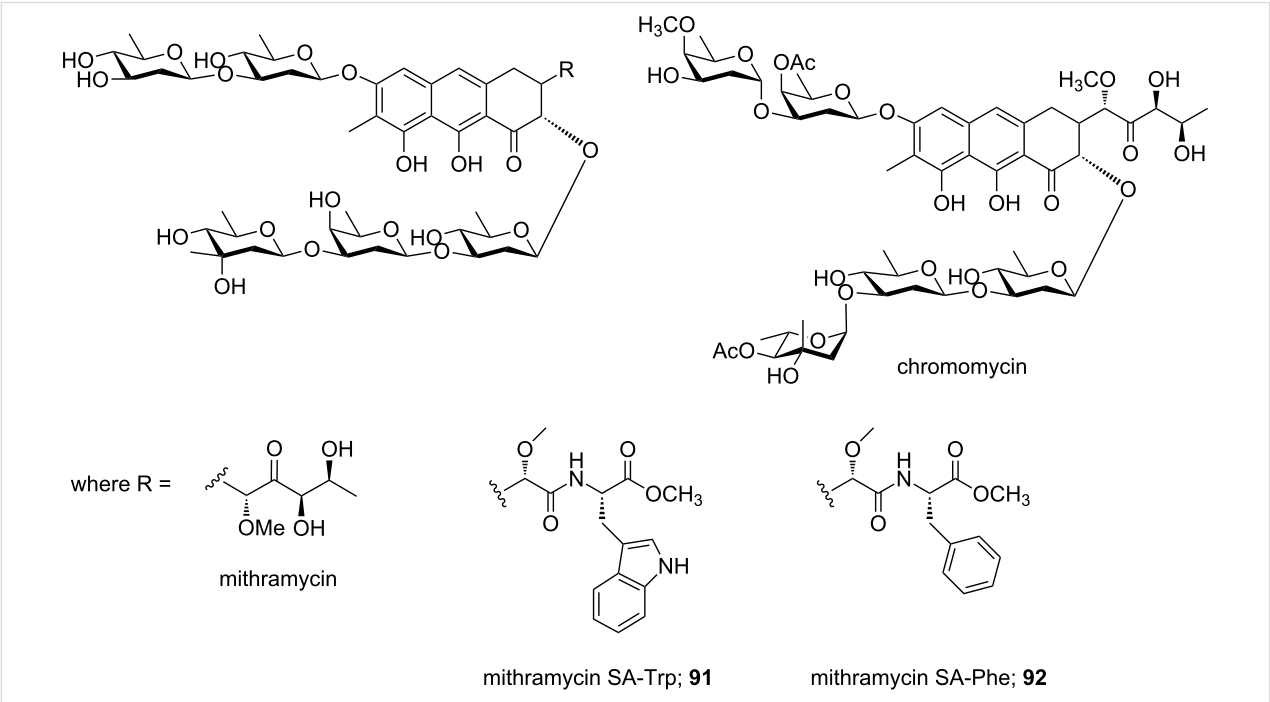


Figure 13: Chemical structures of chromomycin, mithramycin and their synthetic structural analogs **91** and **92**.

Recently, mithramycin was identified as a selective inhibitor of abnormal oncogenic transcription factor EWS–FLI1 in Ewing sarcoma. Hou et al. designed two different mithramycin analogs **91** and **92** in order to probe the mechanism for MTM recognition of DNA to understand how MTM interferes with EWS–FLI1 [144,145]. The authors reported crystal structures of conjugates **91** and **92** bound to DNA sequence specifically and also confirmed a ternary complex formation in the minor groove between FLI1–DNA–MTM on a single GGAA FLI1/MTM binding site. This research introduces a new approach to selectively target EWS–FLI1 or other oncogenic transcription factors to develop anticancer therapeutics.

3. Intercalators

Another mode of non-covalent reversible interaction between DNA and small molecules is intercalation. In general, DNA intercalators consist of planar aromatic or heteroaromatic groups capable of stacking between the adjacent DNA base pairs. These complexes are stabilized by π–π stacking interactions, van der Waals forces, hydrophobic interactions and/or charge transfer forces [29,146].

DNA intercalation induces local structural perturbations in the DNA helix; mainly decrease in the helical twist, which results in lengthening of the DNA [147]. These structural modifications lead to the interruption of DNA replication, transcription and DNA repair processes by interfering with the function of DNA-associated proteins such as polymerases, transcription factors and topoisomerases [19]. Therefore, DNA intercalators are often used as chemotherapeutic agents. Several DNA intercalating drugs have been identified over the years, which include daunomycin (trade name Cerubidine), doxorubicin (trade name Adriamycin), epirubicin (anthracycline family), dactinomycin (trade name Cosmegen), ditercalinium, bleomycin, elsamicin A, m-AMSA, mitoxantrone, acridines, ethidium bromide and so on (Figure 14) [30,148-151]. Anthracyclines are a class of antitumor antibiotics, isolated from *Streptomyces* species, mostly used in various cancer chemotherapy such as acute leukemia, Hodgkin's and non-Hodgkin's lymphoma, breast and ovarian cancer, lung cancer, gastric (stomach) cancer, testicular cancer, bladder cancer and soft tissue sarcoma etc. In addition, they act as topoisomerase II inhibitors [152]. Daunomycin and doxorubicin both possess a

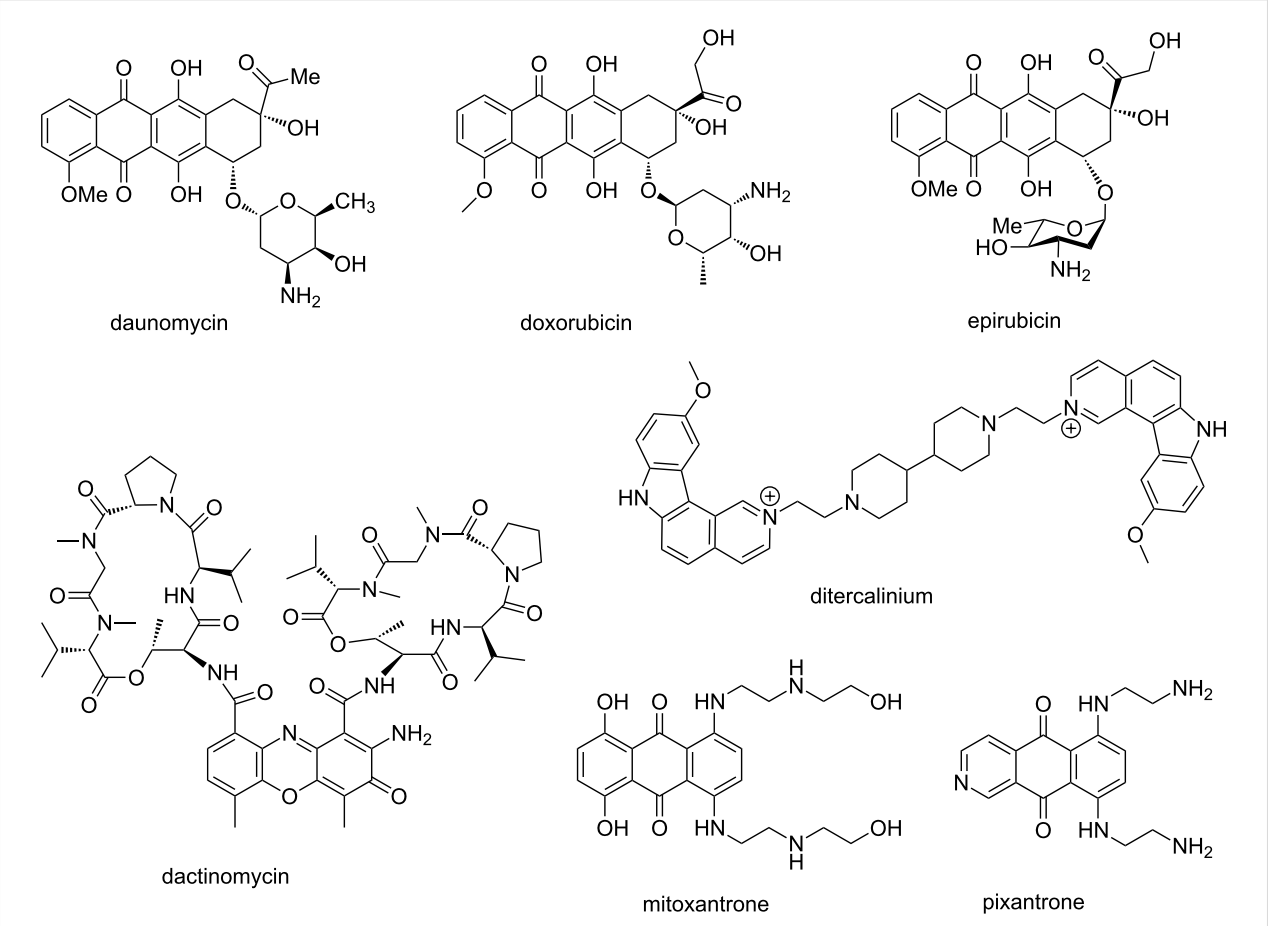


Figure 14: Chemical structures of well-known naturally occurring DNA binding intercalators.

planar ring, a fused cyclohexane ring system and an amino sugar moiety. The ionic interaction between the protonated amine group on the carbohydrate residue and the negatively charged DNA phosphate backbone hold these drugs within the DNA groove, thereby allowing the planar aromatic ring system to intercalate within the G·C steps of the double helix [153,154]. Epirubicin, another drug in the anthracycline family, is the 4′-epimer of doxorubicin. It has been used as a chemotherapy treatment either alone or in combination with other cytotoxic agents. Epirubicin is favored over doxorubicin due to lesser side effects such as reduced myelosuppression and cardiotoxicity. Similar to the other anthracycline drugs, it also acts via intercalating into DNA strands, which eventually inhibits DNA and RNA synthesis leading to cell death [155]. Dactinomycin, also known as actinomycin D, a member of the polypeptide family, is known to inhibit DNA transcription by blocking the chain elongation. This antibiotic has a clear preference for G·C base pairs and interacts with the 2-amino group of guanine. The pentapeptide moiety interacts with the DNA minor groove by hydrogen bonding and hydrophobic interactions, whereas the phenoxazone ring slides into the G·C base pairs for intercalating. Another antitumor drug, ditercalinium, used for treatment of cancer, is an example of non-covalent DNA-binding ligand via bis-intercalation [156]. This drug is a 7*H*-pyridocarbazole dimer, which intercalates into two G·C steps in the major groove. Moreover, the positively charged bis(ethylpiperidinium) moiety interacts with the major grove via charge interaction and induce DNA repair in eukaryotic or prokaryotic cells [157-159]. These dual binding mechanisms (intercalation and minor groove binding) help to form a steady complex between these above mentioned small molecule drugs and DNA duplex. Mitoxantrone is a tricyclic planar anthraquinone derivative with two basic side chains which acts as anticancer chemotherapeutic agent via inducing DNA damage by breaking single and double strands. It is a type II topoisomerase inhibitor [160]. With reduced cardiotoxicity and functionally similar to doxorubicin, it disrupts DNA synthesis and DNA repair via intercalating between the bases in DNA duplex [161]. It has been observed that intercalating anthraquinone chromophore in a pyrimidine (3′-5′) purine sequence remains perpendicular to the direction of inter-base hydrogen bonds, whereas positively charged N-containing basic side chains project outward from the drug [162]. It shows significant activity against acute myeloid leukemia, advanced breast cancer and non-Hodgkins lymphoma [163]. Recently, Konda et al. demonstrated a binding mechanism of another anticancer drug pixantrone to three different oligonucleotide sequences by using NMR and molecular modeling. The upfield shift of pixantrone aromatic protons observed after preferential binding to symmetric CpA dinucleotide sequences supported the intercalative mode of the binding mechanism [164].

Indolocarbazoles represent a family of alkaloids containing bisindoles, which are mostly used as anticancer drugs. The natural antibiotic, rebeccamycin, isolated from *Saccharothrix aerocoloniegenes*, is a representative of this class of molecules as shown in Figure 15. This is a well-known DNA-binding agent and acts as inhibitor of topoisomerase I. The glycoside residue attached with the DNA intercalating domain plays a major role in binding of the drug to the DNA double helix, similar to daunomycin and doxorubicin. It was shown that by replacing the glucose moiety with a 2’-aminoglucose residue, DNA-binding affinity and sequence specificity of compound **93** was enhanced [165]. Another series of structural analogs were synthesized in order to develop novel tumor-active rebeccamycin derivatives. DNA binding affinity of a cationic derivative **96** containing a *N,N*-diethylaminoethyl side chain and **95** bearing an aminoglycoside moiety were compared with an uncharged analog **94**. It was observed that the cytotoxic potential of cationic **95** and **96** was higher in comparison to

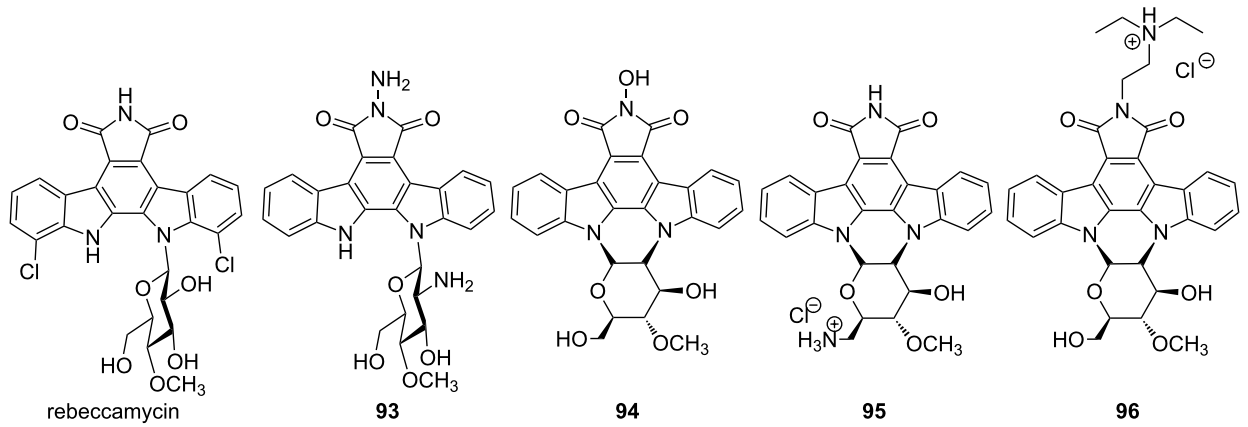


Figure 15: Naturally occurring indolocarbazole rebeccamycin and its synthetic analogs.

uncharged **94**, which is mainly attributed to the enhanced DNA binding affinity and sequence specificity. Installation of the cationic moiety on either the indolocarbazole domain or on the carbohydrate residue greatly reinforces the binding of these drugs to DNA. These molecules preferentially recognize sequences GpT·ApC and TpG·CpA steps [166].

MLN944 (XR5944) is a novel bisphenazine derivative showing excellent cytotoxic activity against various in vitro and in vivo human and murine tumor models (Figure 16) [167,168]. Sappal et al. suggested the primary mechanism of action of this drug involves DNA major groove binding via bis-intercalation and is not involved in the catalytic activity of topoisomerase I or II [169]. When applied in combination with carboplatin or doxorubicin in non-small-cell lung carcinoma [170], or in combination with 5-fluorouracil and irinotecan in colon cancer cell lines [171], MLN944 exhibited synergistic effect in vitro and in vivo. Another DNA intercalating drug TAS-103 (BMS-247615), novel quinolone derivative, is a dual inhibitor of topoisomerases I and II and shows potent cytotoxic effects in vitro and in vivo against subcutaneously-implanted murine and human tumors in vivo, as well as various lung-metastatic murine tumors [172,173]. When this drug was applied with the approved antitumor drug cis-platin, a synergistic effect was observed which could be helpful for the treatment of small-cell lung cancer.

Similar to TAS-103, DACA (*N*-[2-(dimethylamino)ethyl]acridine-4-carboxamide) is another DNA intercalating anticancer drug capable of inhibiting both topoisomerases I and II and currently in clinical trial phase II (Figure 16). It has been observed that the acridine ring intercalates between the DNA base pairs and the 4-carboxamide side chain assists in the major grove binding with its protonated *N,N*-dimethylamino group forming hydrogen-bonding interactions with guanine [174]. The prodiginine family of bacterial alkaloids, isolated from *Serratia* species, represents a varied set of heterocyclic red-pigmented natural products with potent immunosuppressive, antimicrobial and proapoptotic anticancer attributes. These 4-methoxypyrrolic natural products are structurally characterized by the presence of 4-methoxy-2,2'-bipyrrole skeleton [175]. They bind to DNA in the intercalative fashion with the preference for A·T sites. It was further confirmed that they intercalate from the minor groove, as minor groove binding drug distamycin was able to

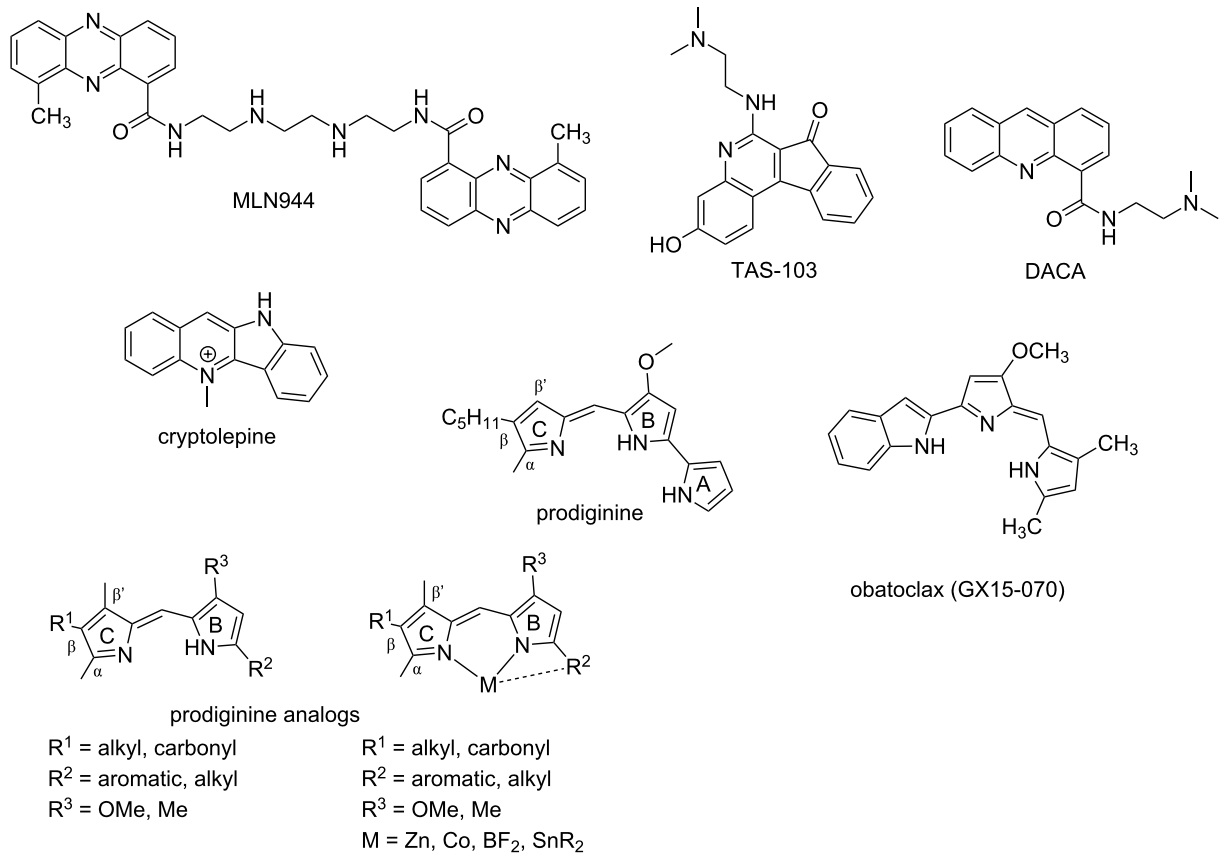


Figure 16: Representative examples of naturally occurring and synthetic derivatives of DNA intercalating agents.

displace them from the DNA double helix. Prodiginine act as a dual topoisomerase I/II inhibitor and has been tested against more than 60 cancer cell lines including breast, lung, stomach, liver, spleen, colon, blood, and chronic myeloid leukemia with an average inhibitory concentration of 2.1 μM [176]. Obatoclax (GX15-070) is a synthetic derivative of natural prodiginines and currently under phase I and phase II clinical trials for the treatment of various types of cancer cell lines [177]. Combination therapies with other chemotherapeutic agents are also currently being tested with obatoclax. Due to their immense biological activities, numerous chemical, chemoenzymatic and biosynthetic strategies were reported to afford several structural analogs of this class of natural products [178,179]. Recently, Marchal et al. reported several structural analogs of natural prodiginines and their complexes with tin, cobalt, boron, and zinc salts with modifications at rings A and C and their antimalarial activities were evaluated in vitro using the 3D7 *Plasmodium falciparum* strain [18]. The authors went on to confirm that the presence of the nitrogen atom in the A-ring is mandatory to show antimalarial activity whereas on the contrary, the presence of an alkyl group at the β′-position of the C-ring is not essential, in fact at times detrimental. Moreover, dibutyltin complexes could also enhance the inhibitory effect in comparison to natural prodiginines, exhibiting IC_{50} values in the nanomolar range. Cryptolepine, isolated from the roots of *Cryptolepis sanguinolenta*, is an indoloquinoline alkaloid with antibacterial, antiviral, and antimalarial properties [180]. Its mode of binding to DNA was tested via absorption, fluorescence, circular and linear dichroism, as well as by a relaxation assay using DNA topoisomerases [181]. It has been observed that this alkaloid binds tightly to DNA and its primary mode of action is intercalation. Cryptolepine has a clear preference for G·C-rich sequences containing non-alternating G·C sites as demonstrated via competition dialysis assays. Besides, the positively charged nitrogen helps to maintain the stability of the DNA–ligand complex via charge interaction. Moreover, it was confirmed that this alkaloid is a potent topoisomerase II inhibitor and a promising antitumor agent [182].

Dar et al. designed and reported a series of novel steroidal imidazo[1',2'-*a*]pyridine derivatives (conjugates **97–99**) via an one-pot three-component tandem approach by reacting different steroidal ketones, 2-aminopyridine and isocyanides and simultaneously investigated their DNA binding affinity and in vitro cytotoxicity (Figure 17) [183]. UV–vis, fluorescence, gel electrophoresis and molecular docking studies revealed that the primary mode of binding of these conjugates with the minor groove of the DNA is intercalation, although the van der Waals and other types of electrostatic and hydrophobic interactions could also play important roles. Significant antiproliferative activity of these conjugates against different cancer cells were observed from MTT assays. These steroidal imidazopyridines induced an apoptosis in A549 cells resulting in cell cycle arrest at low concentration, respectively, confirmed via western blotting and FACS analysis. A series of novel benzo[*k*,*l*]xanthene lignans were designed and synthesized by biomimetic, Mn-mediated oxidative coupling of caffeic esters and amides by Tringali et al. and their DNA binding mechanism was thoroughly studied via DF-STD NMR analysis and molecular docking [184]. These experiments revealed their dual mode of binding mechanism; the planar core intercalates between the minor groove base pairs and the flexible protruding moieties act as minor groove binders. Moreover, conjugates **100** and **101** comprising of lipophilic esters showed significant antiproliferative activity, even better than the anticancer drug 5-fluorouracil (5-FU), against HCT-116 (colon, GI_{50} = 3.16 μM) and H226 (lung, GI_{50} = 4.33 μM) cell lines. Rozas and Wilson reported syntheses, mode of DNA binding mechanism and sequence specificity of a set of conformationally restricted symmetric and asymmetric dicationic DNA binders comprising of 9,10-dihydroanthracene (DHA) **102** and 9*H*-fluorene **103** cores; two conjugates representing each class are shown in the Figure 17 [185]. SPR studies clearly indicated the affinity of these conjugates not only for A·T oligonucleotides, but also for G·C-rich oligonucleotides. Again, they exhibited much stronger binding to DNA in comparison to the flexible core conjugates. Conjugate **103** containing a fluorene core was found to bind A·T oligonucleotides much stronger compared to DHA conjugate **102**. CD and UV experiments revealed DHA analogs bind to DNA via intercalation and fluorine analogs act as intercalator as well as minor groove binder. Nakabayashi et al. reported three cyclometalated ruthenium(II) complexes $[Ru(bpy)_2(C\text{^}N)]Cl$ **104–106** in order to study their ct-DNA binding affinity and cytotoxicity against two tumor (L1210 and HeLa) and a non-tumor (BALB/3T3 clone A31) cell lines [186]. Conjugates **104–106** primarily act as intercalators and/or minor groove binders. Moreover, these conjugates exhibit favorable cytotoxicity against L1210 and HeLa cell lines, much improved in comparison to cis-platin and lower cytotoxicity toward BALB. This research paves a new direction towards the discovery of antitumor drugs. Recently, Rotaru et al. has developed a new fluorescent anthracene-based pyridyl-indolizine derivative (conjugate **V**) via "click" chemistry at the first position of the indolizine core to test their DNA binding efficacy and potential application towards anticancer treatment [187]. Agarose gel electrophoresis, UV–vis and fluorescence experiments along with molecular docking simulations has revealed that conjugate **108** (Figure 17) exhibits higher affinity for the DNA than its precursor containing only a pyridyl-indolizinic skeleton (conjugate **107**) owing to much lower values of binding energy and dissociation constant of the corresponding U-DNA complex.

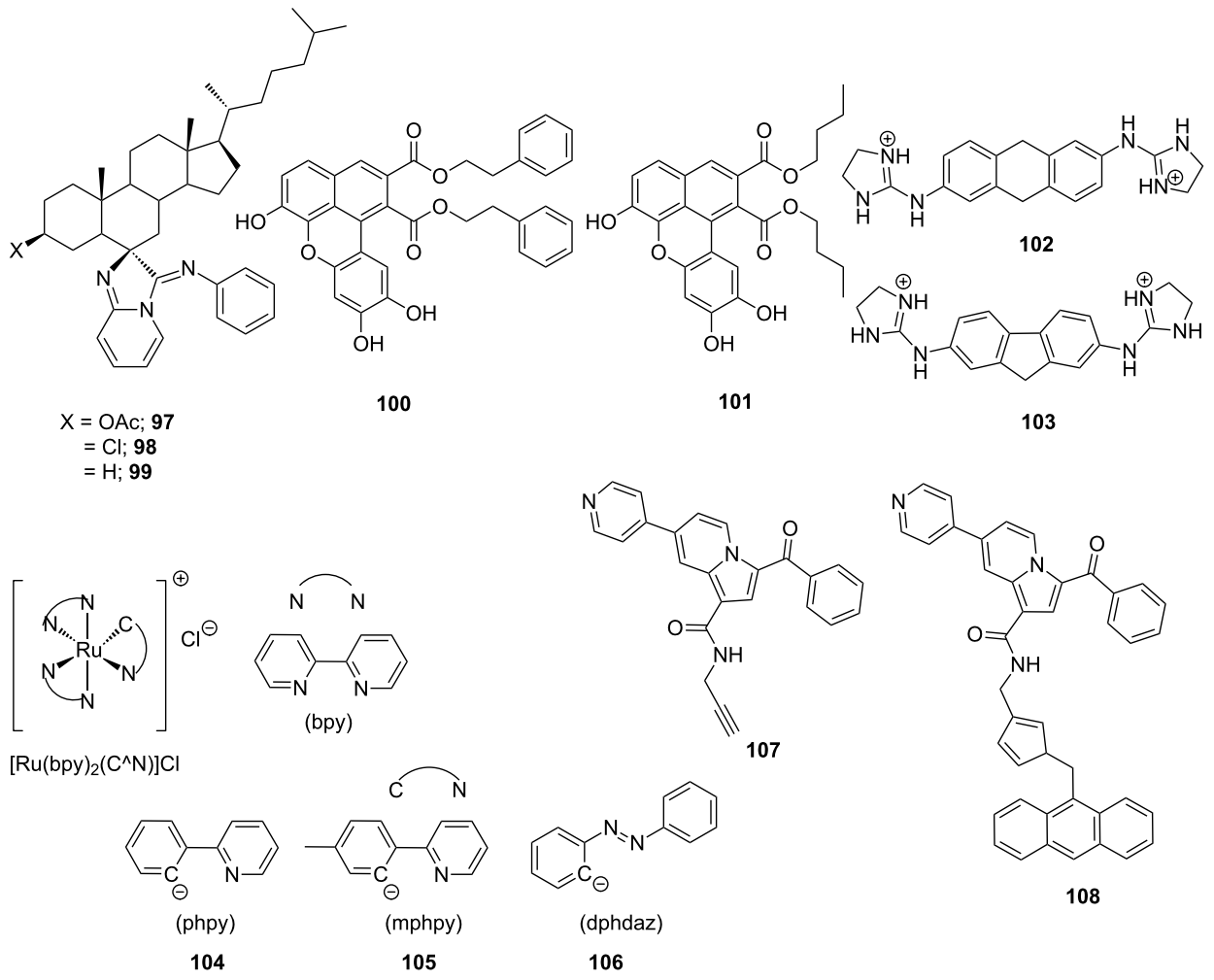


Figure 17: Several recent synthetic varieties of DNA intercalators.

4. Major groove binders

In general, biological macromolecules such as proteins interact with the major groove of ds-DNA via hydrogen bond interactions. In 2012, a detailed review on natural products DNA major groove binders such as pluramycins, aflatoxins, azinomycins, leinamycins, aminosugars, neocarzinostatins was reported, including their binding mechanisms and sequence specificity [188]. The authors clearly demonstrated how major groove binding molecules could block access to various transcription factors by binding to a specific DNA sequence. These natural products primarily act as intercalators; however, some of them interact covalently via alkylation of the nucleophilic sites on DNA. In this section, we will focus on more recent advances in the emergence of modified aminoglycosides (AGs) as reversible major groove binders. AGs are electrostatically attracted to the phosphodiester backbone of nucleic acids due to their polycationic nature. Moreover, they can adapt various conformations due to their flexible ring composition in order to bind within different DNA groove widths. However, B-form duplex DNA has a much larger major groove and the non-aromatic nature of aminoglycosides limits their binding to the DNA major groove due to the lack of shape-complementarity. In this regard, chemical modifications on AGs will lead to the design of novel DNA binding ligands with improved sequence specificity.

It has been observed that neomycin exhibits a much better shape complementarity with A-form DNA due to its narrower groove in comparison to B-DNA. Arya et al. investigated if neomycin, an effective A-form groove binder, could be inserted into the major groove of B-DNA by tethering neomycin with the well-known duplex selective groove binder Hoechst 33258. A neomycin–Hoechst 33258 conjugate **109** showed significant stabilization of DNA duplexes and destabilization of the DNA

triplex which in turn, suggested that neomycin could be forced into the major groove of a B-form DNA duplex (Figure 18) [189].

Inspired by the earlier research, the triple recognition of B-DNA by a novel neomycin–Hoechst 33258–pyrene conjugate **114** was investigated in order to probe the molecular forces that dictate binding within the DNA grooves and base pairs by using spectroscopic, calorimetric, and viscometric techniques [12]. Conjugate **114** was found to bind stronger to B-DNA in comparison to all three constituents such as neomycin, pyrene and Hoechst 33258, thereby stabilizing DNA more efficiently. In addition, fluorescence titrations confirmed that the conjugate **114** could specifically recognize a contiguous stretch of nine A·T base pairs. The conjugate **114** was hypothesized to simultaneously recognize DNA via all three recognition motifs: major groove, minor groove, and intercalation and this research indeed paves the way for the development of multivalent DNA binding molecules. Kumar et al. reported a dimeric neomycin–neomycin conjugate **115** (Figure 18) with a flexible linker 2,20-(ethylenedioxy)bis(ethylamine) which could selectively bind to A·T-rich DNA duplexes preferentially over G·C-rich sequences con-

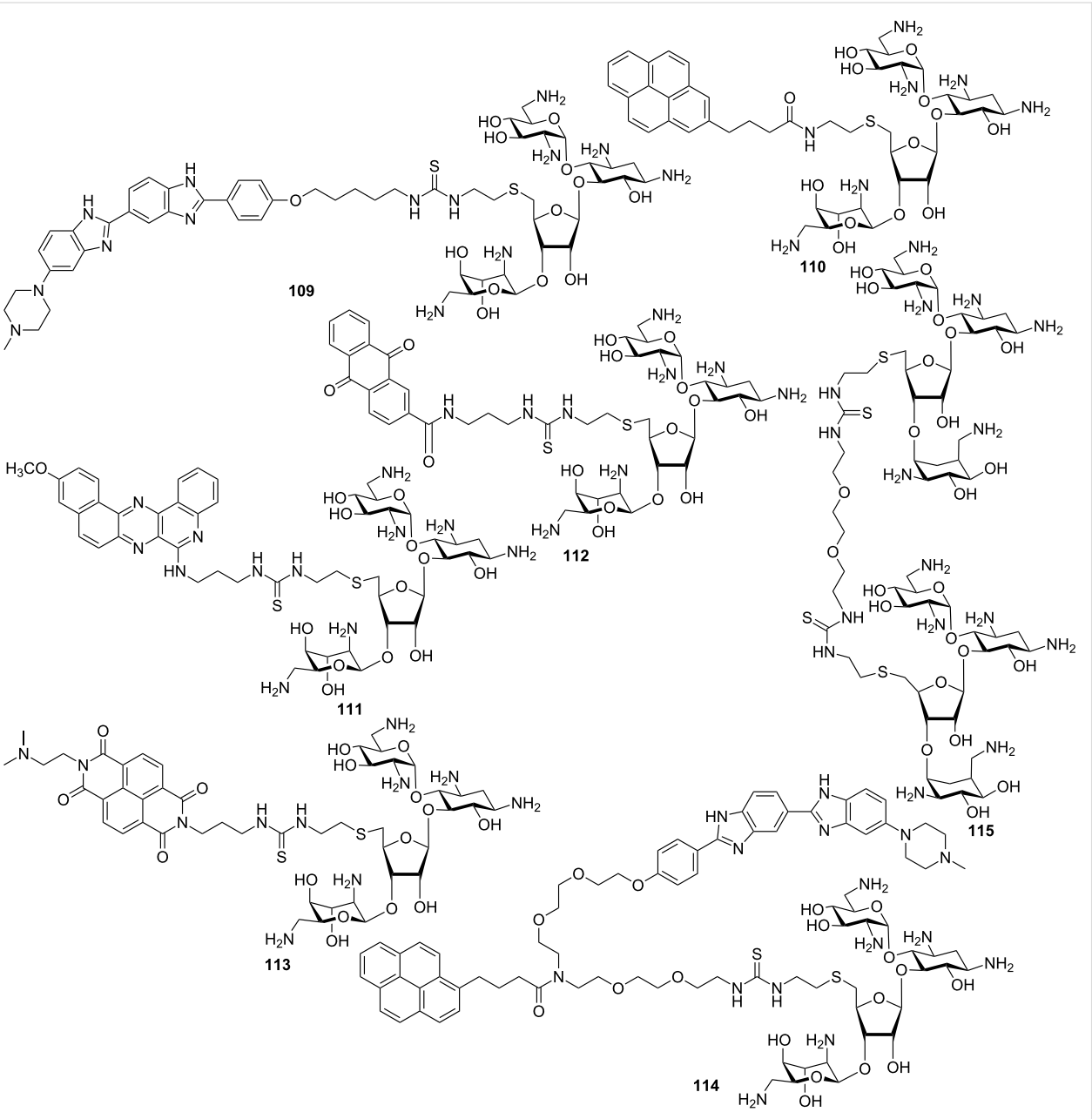


Figure 18: Aminoglycoside (neomycin)–Hoechst 33258/intercalator conjugates.

firmed via ITC, CD, FID, and UV thermal denaturation experiments [190]. Moreover, dimeric conjugate **115** exhibits higher affinity for the major groove of A-tract sequences over those containing alternating A·T bases. In addition, conjugate **115** destabilizes poly(dA)·2poly(dT) triplex but stabilizes poly(dA)·poly(dT) duplex, as opposed to neomycin monomer, suggesting the major groove as the binding site.

4.1. Shape and nucleic acid selectivity (DNA vs. RNA)

One of the major concerns in nucleic acid recognition will be achieving selectivity: selectivity of one form of DNA over the other forms/sequences/shapes and DNA versus RNA selectivity. Previous work has shown that using designed molecules, both types of selectivity can be achieved. Using a competition dialysis assay, it was shown that neomycin is an A-form selective ligand over B-form structures irrespective of its constituent type (DNA or RNA) [191,192]. In a striking contrast, thiourea linked dimeric neomycin conjugates exhibited complete reversal of target selectivity from A-form triplex DNA to B-form duplex DNA structures [193]. Further investigations using a series of thiourea linked neomycin dimers spaced by different linker sizes revealed high affinity ($K_a = 2.26 \times 10^8$ M^{-1}) binding for B-DNA over other forms of DNA. A FID based assay involving 512 DNA duplexes of different sequence compositions revealed that neomycin dimers prefer to bind DNA duplex with the AT-tract [190]. The neomycin dimer **115** (Figure 18) binds to short oligonucleotides (12 mer) with 1:1 ligand to DNA duplex stoichiometry and show a binding site size of 11–12 base pairs with the polymeric DNA. A complete thermodynamic study of neomycin dimer **115** binding to a B-DNA sequence revealed that the first binding event (the high affinity site) is entropically driven and that the ionic strength dependence of the binding is highly dependent on the electrolytic contribution [194]. The neomycin dimers also displayed length dependent shape recognition of the B-DNA [195].

Dimerization of neomycin units using more rigid linkers (triazole linkers) than the thiourea linkers resulted in enhanced binding and more selective recognition of a TAR-RNA conformation over the DNA duplex structure [196,197]. The triazole linked neomycin dimer **116** (Figure 19) displayed close to nanomolar affinity ($K_a = 1.39 \times 10^8$ M^{-1}) and 1:1 binding stoichiometry with a biologically relevant truncated model RNA sequence of TAR. In this case also, the binding was found to be dependent on the linker length joining the two neomycin units and the neomycin dimer conjugates thermally stabilized the TAR RNA structure by up to 10 °C. The neomycin dimers exhibited much improved cytopathic effects in MT-2 cells than neomycin alone [196]. These results showed that subtle changes in the linker composition bring profound differences in the DNA versus RNA nucleic acid selectivity. The linker length was found to have a significant and profound effect in the DNA versus RNA selectivity of a series of neomycin–bisbenzimidazole conjugates. It was found that neomycin–bisbenzimidazole conjugates **117**–**125** with short linkages (up to 11 atoms) stabilized a 12mer duplex DNA $d(CGCAAATTTGCG)_2$ better than its RNA equivalent $r(CGCAAAUUUGCG)_2$. However, neomycin–bisbenzimidazole conjugates with long linkers (15 atoms or higher) stabilized the RNA duplex sequence $r(CGCAAAUUUGCG)_2$ better than the DNA sequence $d(CGCAAATTTGCG)_2$ [198]. The unique selectivity of neomycin–bisbenzimidazole conjugates with long linkers towards RNA duplex was attributed to a linker dependent intercalation of the bisbenzimidazole unit into the RNA duplex, which was maximum (74 °C) with the longest linker (23 atoms). The dual binding of the conjugates allows both neomycin and bisbenzimidazole units binding in a complementary way to impart thermal stabilization of the RNA duplex [198]. The bisbenzimidazole units of the neomycin–bisbenzimidazole conjugates were earlier reported to bind in the minor groove of the DNA [199].

5. Alkylators

Covalent interaction between small molecules and DNA is usually irreversible, which leads to inhibition of DNA functions such as transcription or replication resulting in subsequent cell death. The small molecules can change the overall conformation by cross-linking to the DNA duplexes. However, their low selectivity reflects in their high toxicity in normal cells. Thus, in order to solve this issue, several researchers have designed and developed various synthetic analogs of existing drugs having much improved sequence specific DNA selectivity with reduced side effects, which are discussed in the following sections.

Alkylating agents are strong electrophilic compounds that react chemically with nucleophilic moieties of DNA or proteins to form covalent bonds by transferring an alkyl group to DNA. Their cytotoxicity results from the alkylation of DNA bases that can irreversibly inhibit essential DNA processes such as DNA replication and/or transcription. Nitrogen mustards, derived from sulfur mustards, including bendamustine, estramustine, melphalan, chlormethin, chlorambucil, were the first alkylating agents used for the treatment of leukemias and lymphomas. Alkylation occurs via the formation of an aziridinium ion followed by nucleophilic attack by the N7 of guanine [200]. The other well-known alkylators include platinum derivatives (cisplatin, carboplatin, oxaliplatine), oxazaphosphorines (cyclophosphamide, ifosfamide, trofosfamide), ethylene imines (mitomycin C, thiotepa, altretamine), nitrosoureas (MNU,

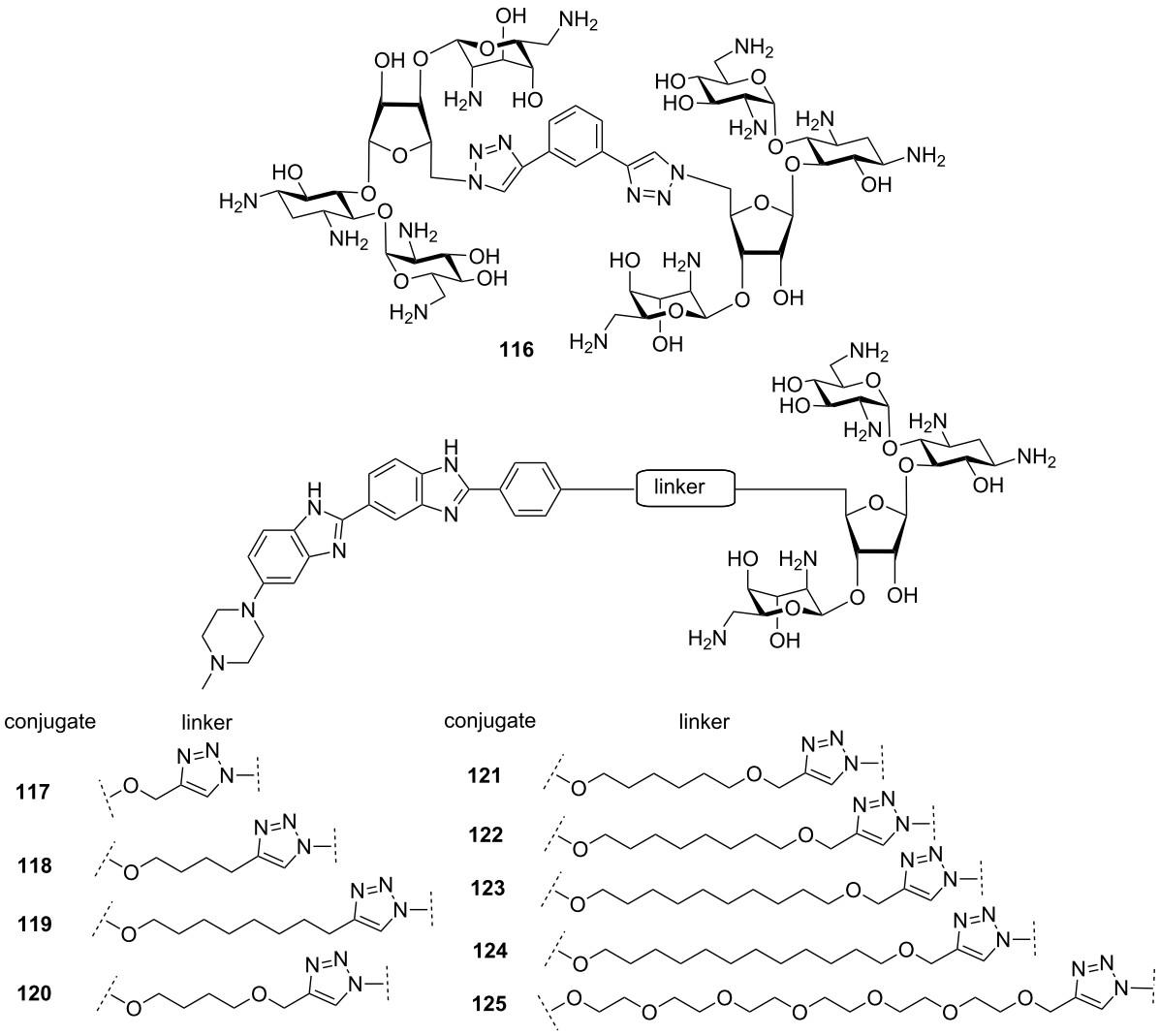


Figure 19: Chemical structures of triazole linked neomycin dimers and neomycin–bisbenzimidazole conjugates.

MNNG, BCNU, CCNU, nimustine), triazenes and hydrazines (dacarbazine temozolomide Procarbazine) [201], trabectidine and so on [21].

In the last few decades, a plethora of natural products and their synthetic analogs were tested for their antineoplastic effect which includes (+)-CC-1065, duocarmycin SA, irofulven, ML-970, *seco*-CBI-indole$_2$ and so on (Figure 20). (+)-CC-1065 and duocarmycin SA are known antitumor drugs, isolated from *Streptomyces species*, which primarily act as minor groove alkylators by forming adenine N3 adducts in A·T-rich regions via the electrophilic cyclopropylindol (CPI) subunit [202,203]. However, these natural products showed significantly reduced antitumor activity mostly due to their low water solubility. Baraldi et al. reported a series of hybrid conjugates by tethering polypyrrole minor groove binders, derived from distamycin A and two pyrazole analogues of the CPI unit of the potent anti-tumor antibiotic (+)-CC-1065 in order to enhance potency, specificity and water solubility of pyrazole CPI analogs [204]. Conjugate **126** (Figure 20) was found to be extremely cytotoxic with IC_{50} values for the different tumor cell lines ranging from 7 to 71 nM.

Additionally, it exhibited the strongest DNA alkylation activity via sequence-specific alkylation of the third adenine located in the sequence 5′-ACAAAAATCG-3′ [204]. The high activity of tripyrrole conjugate **126** than mono- and dipyrrole analogs might result from its stronger binding with in the minor grooves due to multiple hydrogen bonding and van der Waals forces. However, higher toxicity of these natural products and their synthetic analogs forced the researchers to develop less toxic analogs. The newly identified indole-carboxamide ML-970

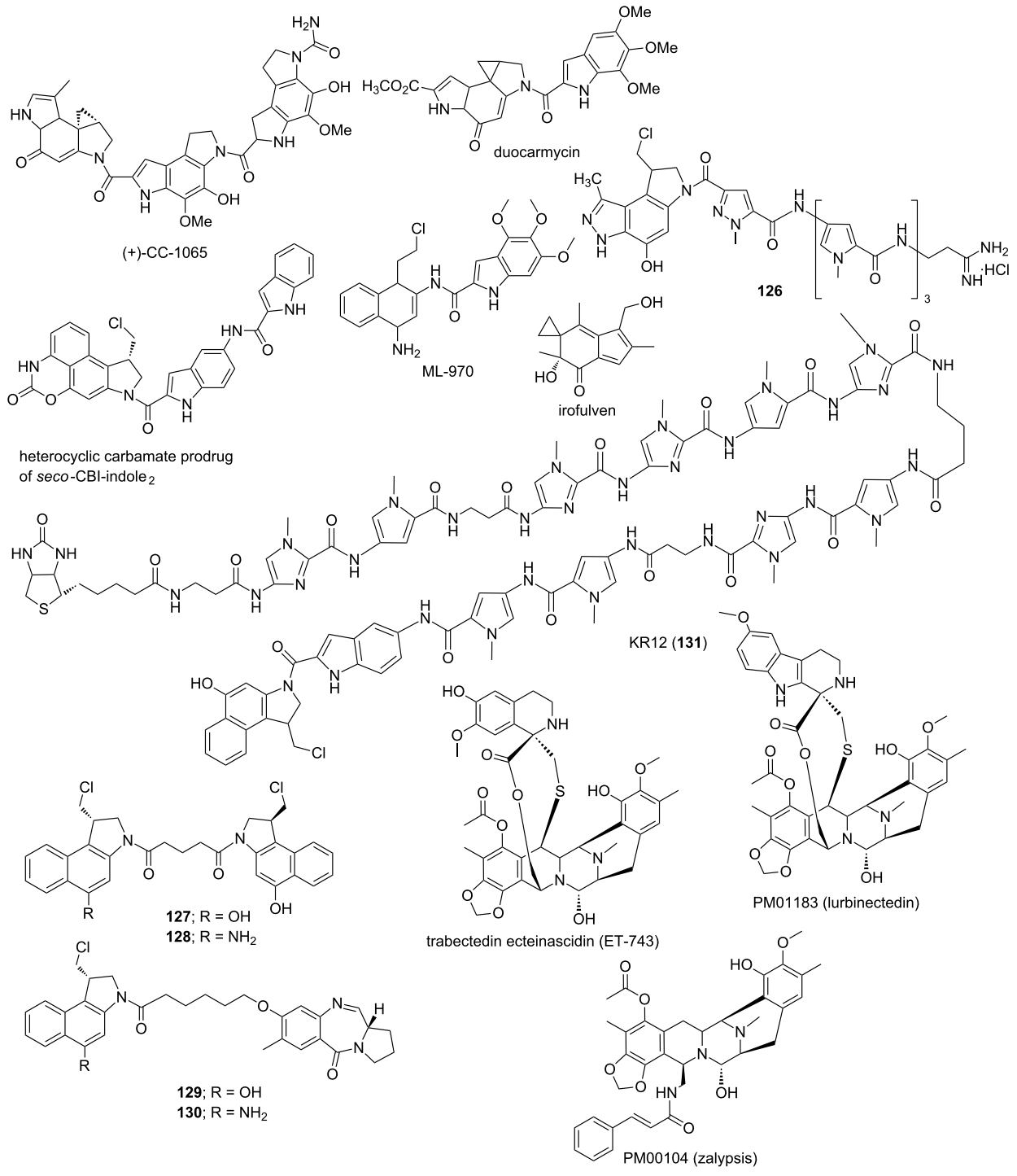


Figure 20: Representative examples of naturally occurring and synthetic analogs of DNA binding alkylating agents.

represents another synthetic derivative which binds the A·T-rich DNA minor groove and alkylates DNA. In addition, it shows potent cytotoxic activity, with an average GI_{50} of 34 nM with much lower myelotoxicity in comparison to (+)-CC-1065 and duocarmycins [205]. Another heterocyclic carbamate prodrug of *seco*-CBI-indole$_2$ was reported which represents a new member of a class of hydrolyzable prodrugs of the duocarmycin and CC-1065 family of natural products [206]. This prodrug is activated by the hydrolysis of the carbamate residue, thereby slowly releasing the active free form of the drug with no

residual byproduct (CO_2). Thus, its slow free drug release allows the safe and efficacious administration of much higher doses than the parent-alkylating agent. Tercel et al. recently developed two new sets of DNA monoalkylating agents **127**, **128** (CBI–CBI dimer) and **129**, **130** [CBI–pyrrolobenzodiazepine (PBD) dimer], with phenol-CBI and amino-CBI residues and their cytotoxicity against nine human tumor cell lines were tested [207]. Interestingly, **128** and **130**, amino-CBI analogs found to be less cytotoxic (2- to 190-fold reduction in potency depending on the particular compound and cell line) in comparison to their phenol analogs **127** and **129** (Figure 20). Irofulven, a semisynthetic derivative of the mushroom-derived compound illudin S, is another extremely promising antitumor agent for solid tumor cells. Its mechanism of action involves an activation step in which nucleophilic attack on the α,β-unsaturated ketone by thiol or NADPH leads to opening of the cyclopropane ring, which results in alkylation of protein and DNA [208].

Recently, Lin et al. reported another attractive versatile sequence-specific DNA alkylating agent (KR12, **131**) by tethering well-known minor groove binder Py–Im polyamides with an alkylating agent such as *seco*-CBI (Figure 20) [209]. The authors have identified KR12 binding sites in the human LS180 colorectal cancer genome and the reduction of KR12-bound gene expressions was also observed. Another marine alkaloid trabectedin (ET-743) comprising of three fused tetrahydroisoquinoline rings has been introduced into clinical trial for the treatment of soft tissue sarcoma. Two of these sulfide-linked substituted isoquinoline rings take part in minor groove binding through covalent interaction with the third ring protruding from DNA duplex allowing interactions with adjacent nuclear proteins [210]. ET-743 interferes with several transcription factors and DNA binding proteins via preventing protein binding by distorting DNA structure. Two other synthetic tetrahydroisoquinolone alkaloid derivatives have been developed. PM01183 (lurbinectedin) [211] and PM00104 (Zalypsis®) [212], which showed broad range of chemotherapeutic activity against solid human tumor cell lines are currently in phase II trials. They both act as DNA binding agents, thereby causing inhibition of the cell cycle and transcription. Varadarajan et al. has developed a strategy for overcoming the deficiencies in current DNA-alkylating chemotherapy drugs by designing a site-specific DNA-methylating agent that can target cancer cells because of its selective uptake via glucose transporters, which are overexpressed in most cancers. A glucosamine unit, which can facilitate uptake via glucose transporters, was conjugated to one end of a bispyrrole triamide unit, which is known to bind to the minor groove of DNA at A/T-rich regions and led to increased activity against resistant glioblastoma cells [213].

6. Pyrrolobenzodiazepines (PBDs)

Pyrrolobenzodiazepines (PBDs) are a class of naturally occurring sequence-selective DNA alkylating agents with antitumor properties, which include DC-81, tomaymycin, and anthramycin, isolated from various actinomycetes (Figure 21). The antitumor activity of these classes of molecules results from the sequence selective covalent binding with the 2-amino group of guanine bases in the minor groove of duplex DNA to the electrophilic imine of the diazepine ring. Anthramycin, isolated in the 1950s, is an active antitumor agent and exhibits antineoplastic activity against various types of tumors including Ehrlich solid carcinoma, sarcoma, epidermal carcinoma and leukemia L1210 cells [214].

However, its high cardio toxicity limits clinical application. In order to enhance their DNA binding affinity, several C8-diether-linked DC-81 dimers such as DSB-120 (dimer of DC-81) were synthesized [215,216]. Unfortunately, these dimers did not exhibit expected in vivo antitumor activity probably due to the low bioavailability and excessive electrophilicity at the N10–C11 imine moiety [217]. This led to develop another PBD dimer (SJG-136) linked by a propane-1,3-diether, which exhibited significant in vivo potential for leukemia treatment. SJG-136 has recently passed phase II clinical trials in patients with leukemia and ovarian cancer. Kamal et al. designed a series of novel PBD dimers **132**–**135** comprising of two DC-81 subunits tethered via piperazine side-armed-alkane spacer [217]. These conjugates, especially conjugate **134**, exhibit much improved cytotoxicity than DSB-120 in nine different human cancer cell lines. The author's demonstrated installation of a piperazine ring in the middle of such an alkanedioxy linker results in several hydrophobic interactions, which in turn, enhances DNA binding ability, confirmed via DNA thermal denaturation studies. A set of novel hybrid conjugates by tethering PBD with polyamides, well-known DNA minor groove binders, was designed by Thurston et al. in order to explore structure/sequence selectivity relationships and target gene promoter regions [218]. Conjugate **136** comprising of *N*-methylpyrrole and a thiazole residue exhibits greater DNA binding affinity as well as selectivity for inverted CCAAT sequences within the topoisomerase IIα promoter region. Recently, Kamal et al. reported a set of C8-linked dithiocarbamate/piperazine bridged PBD conjugates and their cytotoxic potential and DNA binding ability were evaluated [219]. Conjugate **137** has shown promising cytotoxicity against 33 cell lines in nine cancer phenotypes with GI_{50} values of <0.99 μM. Thermal denaturation (ΔT_m) studies revealed that by the introduction of *N*-methylpiperazine dithiocarbamate with five-membered alkane spacer to the PBD core increased the DNA-binding activity considerably in conjugate **138** (ΔT_m = 10.9 °C, Figure 21). Thurston et al. recently reported a thorough review

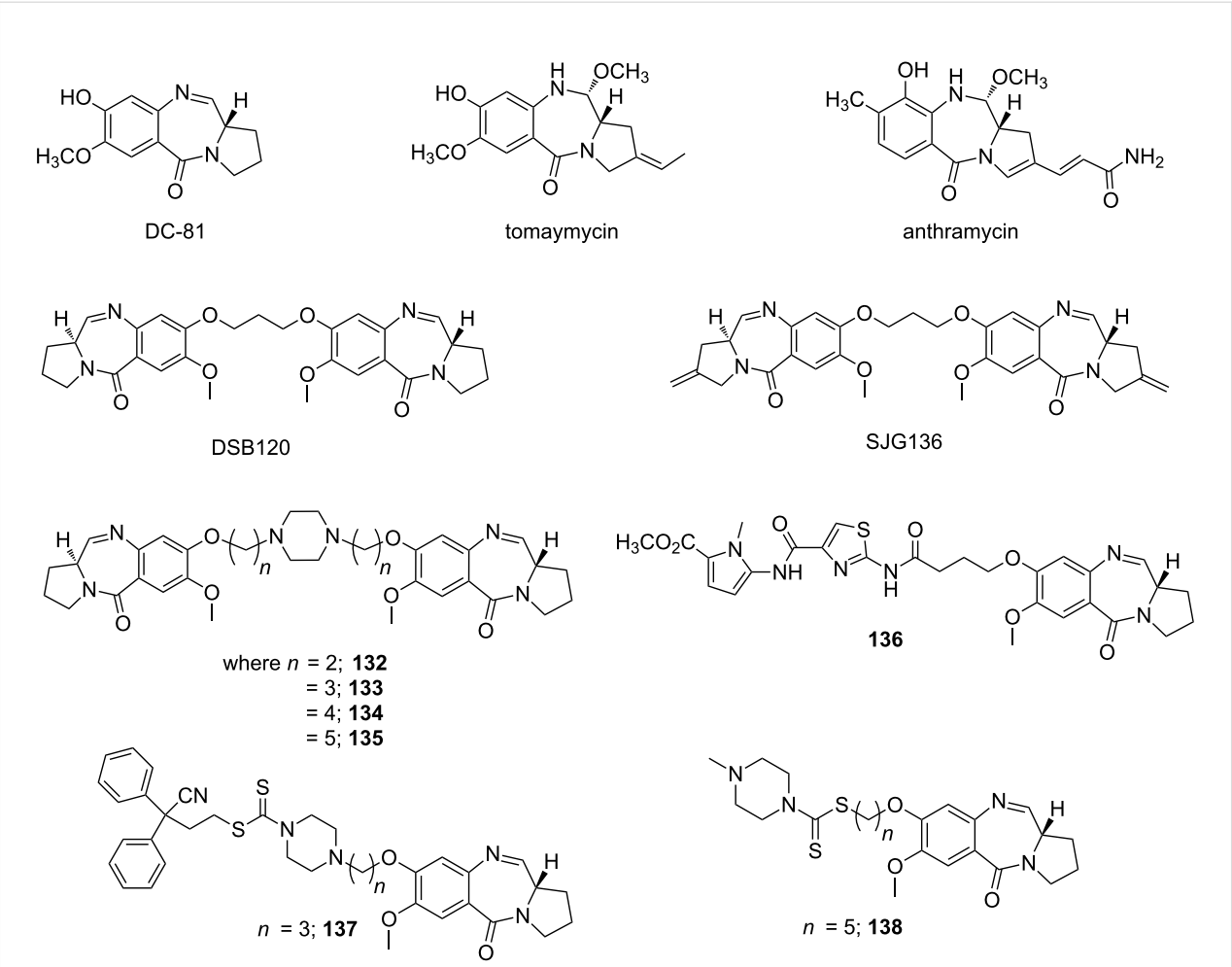


Figure 21: Chemical structures of naturally occurring and synthetic analogs of pyrrolobenzodiazepines.

on the topic, covering the recent developments, SARs and biological applications of PBDs [16].

Conclusion

Regulation of DNA functions with the interference of small molecule DNA binding agents is an established and ongoing area of nucleic acid targeted drug discovery. The clinical success, coupled with high cytotoxicity of DNA binding anticancer agents such as doxorubicin and cis-platin over the past four decades challenges us to design novel agents with reduced toxicity and alternative mechanisms. As covered in this review, new DNA binders are rapidly gaining a foothold in somewhat less explored domains of their application as antibacterial, antifungal and antiparasitic agents beyond their repertoire as anticancer agents. Many of the known sequence specific polyamides have been successfully developed as hairpins, H-pins and hybrid conjugates for enhanced recognition of contiguous DNA bases. The molecules covered in this review show that they indeed are capable of disrupting DNA-transcription factor interactions with high affinity highlighting their emerging importance in chemical biology and potential therapeutics. Recent reports have also shown that end modification of classical bisbenzimidazole (such as Hoechst 33258) based minor groove binding agents leads to dramatic changes in DNA binding, selectivity in bacterial versus human topoisomerase, cellular internalization and cytotoxicity [123]. These findings highlight the sensitivity of DNA sequence selective binders to even modest changes in the chemical structure of the target ligand.

An important aspect of hybrid drug design is the role of linker length and composition on target selectivity and affinity. Optimization of the linker length is an important aspect of fragment-based drug design and appropriate linkage assessment is crucial in optimizing the target binding and cellular uptake of nucleic acid binding ligands. The discoveries summarized in this report

reflect the enormous potential, challenges and expanding diversity of DNA targeted drugs in addressing current therapeutic challenges.

Acknowledgements

DPA thanks the National Institute of Health for financial support (Grants GM097017 and AI114114).

ORCID® iDs

Nihar Ranjan - https://orcid.org/0000-0003-3581-4605
Dev P. Arya - https://orcid.org/0000-0001-5873-1066

References

1. Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U. S. A.* **1976,** *3,* 804–808. doi:10.1073/pnas.73.3.804
2. Garvie, C. W.; Wolberger, C. *Mol. Cell* **2001,** *8,* 937–946. doi:10.1016/S1097-2765(01)00392-6
3. Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. *Nature* **2009,** *461,* 1248–1253. doi:10.1038/nature08473
4. Koutsodontis, G.; Kardassis, D. *Oncogene* **2004,** *23,* 9190–9200. doi:10.1038/sj.onc.1208141
5. Kurmis, A. A.; Yang, F.; Welch, T. R.; Nickols, N. G.; Dervan, P. B. *Cancer Res.* **2017,** *9,* 2207–2212. doi:10.1158/0008-5472.CAN-16-2503
6. Hud, N. V.; Plavec, J. *Biopolymers* **2003,** *69,* 144–158. doi:10.1002/bip.10364
7. Saenger, W. In *Principles of Nucleic Acid Structure;* Cantor, C. R., Ed.; Springer-Verlag: New York, 1984; pp 311 ff. doi:10.1007/978-1-4612-5190-3
8. Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry,* 5th ed.; WH Freeman: New York, 2002.
9. Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **1962,** *48,* 2013–2018. doi:10.1073/pnas.48.12.2013
10. Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2002,** *320,* 189–200. doi:10.1016/S0022-2836(02)00428-X
11. Willis, B.; Arya, D. P. *Bioorg. Med. Chem. Lett.* **2009,** *19,* 4974–4979. doi:10.1016/j.bmcl.2009.07.079
12. Willis, B.; Arya, D. P. *Biochemistry* **2010,** *49,* 452–469. doi:10.1021/bi9016796
13. Willis, B.; Arya, D. P. *Curr. Org. Chem.* **2006,** *10,* 663–673. doi:10.2174/138527206776359739
14. Scott, F. J.; Puig-Sellart, M.; Khalaf, A. I.; Henderson, C. J.; Westrop, G.; Watson, D. G.; Carter, K.; Grant, M. H.; Suckling, C. J. *Bioorg. Med. Chem. Lett.* **2016,** *26,* 3478–3486. doi:10.1016/j.bmcl.2016.06.040
15. Bolhuis, A.; Aldrich-Wright, J. R. *Bioorg. Chem.* **2014,** *55,* 51–59. doi:10.1016/j.bioorg.2014.03.009
16. Mantaj, J.; Jackson, P. J. M.; Rahman, K. M.; Thurston, D. E. *Angew. Chem., Int. Ed.* **2017,** *56,* 462–488. doi:10.1002/anie.201510610
17. Alniss, H. Y.; Salvia, M.-V.; Sadikov, M.; Golovchenko, I.; Anthony, N. G.; Khalaf, A. I.; MacKay, S. P.; Suckling, C. J.; Parkinson, J. A. *ChemBioChem* **2014,** *15,* 1978–1990. doi:10.1002/cbic.201402202
18. Marchal, E.; Smithen, D. A.; Uddin, M. I.; Robertson, A. W.; Jakeman, D. L.; Mollard, V.; Goodman, C. D.; MacDougall, K. S.; McFarland, S. A.; McFadden, G. I.; Thompson, A. *Org. Biomol. Chem.* **2014,** *12,* 4132–4142. doi:10.1039/c3ob42548g
19. Mišković, K.; Bujak, M.; Baus Lončar, M.; Glavaš-Obrovac, L. *Arh. Hig. Rada Toksikol.* **2013,** *4,* 593–602.
20. Wang, M.; Yu, Y.; Liang, C.; Lu, A.; Zhang, G. *Int. J. Mol. Sci.* **2016,** *17,* No. 779. doi:10.3390/ijms17060779
21. Puyo, S.; Montaudon, D.; Pourquier, P. *Crit. Rev. Oncol. Hematol.* **2014,** *89,* 43–61. doi:10.1016/j.critrevonc.2013.07.006
22. Khan, G. S.; Shah, A.; Zia-ur-Rehman; Barker, D. *J. Photochem. Photobiol., B* **2012,** *115,* 105–118. doi:10.1016/j.jphotobiol.2012.07.003
23. Khalaf, A. I.; Al-Kadhimi, A. A. H.; Ali, J. H. *Acta Chim. Slov.* **2016,** *63,* 689–704. doi:10.17344/acsi.2016.2775
24. Neidle, S. *J. Med. Chem.* **2016,** *59,* 5987–6011. doi:10.1021/acs.jmedchem.5b01835
25. Li, T.; Jin, Y.; Vershon, A. K.; Wolberger, C. *Nucleic Acids Res.* **1998,** *26,* 5707–5718. doi:10.1093/nar/26.24.5707
26. Pabo, C. O.; Lewis, M. *Nature* **1982,** *298,* 443–447. doi:10.1038/298443a0
27. Somers, W. S.; Phillips, S. E. *Nature* **1992,** *359,* 387–393. doi:10.1038/359387a0
28. Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. *J. Mol. Biol.* **1997,** *271,* 244–257. doi:10.1006/jmbi.1997.1170
29. Chaires, J. B. *Biopolymers* **1997,** *44,* 201–215. doi:10.1002/(SICI)1097-0282(1997)44:3<201::AID-BIP2>3.0.CO;2-Z
30. Palchaudhuri, R.; Hergenrother, P. J. *Curr. Opin. Biotechnol.* **2007,** *18,* 497–503. doi:10.1016/j.copbio.2007.09.006
31. Tse, W. C.; Boger, D. L. *Chem. Biol.* **2004,** *11,* 1607–1617. doi:10.1016/j.chembiol.2003.08.012
32. Pelton, J. G.; Wemmer, D. E. *J. Biomol. Struct. Dyn.* **1990,** *8,* 81–97. doi:10.1080/07391102.1990.10507791
33. Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochemistry* **1987,** *26,* 5590–5595. doi:10.1021/bi00392a002
34. Randrianarivelo, M.; Zakrzewska, K.; Pullman, B. *J. Biomol. Struct. Dyn.* **1989,** *6,* 769–779. doi:10.1080/07391102.1989.10507736
35. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U. S. A.* **1985,** *82,* 1376–1380. doi:10.1073/pnas.82.5.1376
36. Deng, J.; Pan, B.; Sundaralingam, M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003,** *59,* 2342–2344. doi:10.1107/S0907444903020730
37. McHugh, M. M.; Woynarowski, J. M.; Sigmund, R. D.; Beerman, T. A. *Biochem. Pharmacol.* **1989,** *38,* 2323–2328. doi:10.1016/0006-2952(89)90472-3
38. Woynarowski, J. M.; McHugh, M.; Sigmund, R. D.; Beerman, T. A. *Mol. Pharmacol.* **1989,** *35,* 177–182.
39. Sessa, C.; Pagani, O.; Zurlo, M. G.; de Jong, J.; Hofmann, C.; Lassus, M.; Marrari, P.; Strolin Benedetti, M.; Cavalli, F. *Ann. Oncol.* **1994,** *5,* 901–907. doi:10.1093/oxfordjournals.annonc.a058728
40. Broggini, M.; Coley, H. M.; Mongelli, N.; Pesenti, E.; Wyatt, M. D.; Hartley, J. A.; D'Incalci, M. *Nucleic Acids Res.* **1995,** *23,* 81–87. doi:10.1093/nar/23.1.81
41. Broggini, M.; Ponti, M.; Ottolenghi, S.; D'Incalci, M.; Mongelli, N.; Mantovani, R. *Nucleic Acids Res.* **1989,** *17,* 1051–1059. doi:10.1093/nar/17.3.1051
42. Bellorini, M.; Moncollin, V.; D'Incalci, M.; Mongelli, N.; Mantovani, R. *Nucleic Acids Res.* **1995,** *23,* 1657–1663. doi:10.1093/nar/23.10.1657

43. Cozzi, P. *Farmaco* **2000,** *55,* 168–173. doi:10.1016/S0014-827X(00)00013-6
44. Baraldi, P. G.; del Carman Nunez, M.; Espinosa, A.; Romagnoli, R. *Curr. Top. Med. Chem.* **2004,** *4,* 231–239. doi:10.2174/1568026043451474
45. Cozzi, P.; Beria, I.; Caldarelli, M.; Capolongo, L.; Geroni, C.; Mazzini, S.; Ragg, E. *Bioorg. Med. Chem. Lett.* **2000,** *10,* 1653–1656. doi:10.1016/S0960-894X(00)00295-X
46. Romagnoli, R.; Baraldi, P. G.; Cruz-Lopez, O.; Lopez-Cara, C.; Preti, D. *Mini-Rev. Med. Chem.* **2009,** *9,* 81–94. doi:10.2174/138955709787001640
47. Sabatino, M. A.; Colombo, T.; Geroni, C.; Marchini, S.; Broggini, M. *Clin. Cancer Res.* **2003,** *9,* 5402–5408.
48. Cozzi, P.; Beria, I.; Caldarelli, M.; Geroni, C.; Mongelli, N.; Pennella, G. *Bioorg. Med. Chem. Lett.* **2000,** *10,* 1273–1276. doi:10.1016/S0960-894X(00)00205-5
49. Pezzola, S.; Antonini, G.; Geroni, C.; Beria, I.; Colombo, M.; Broggini, M.; Marchini, S.; Mongelli, N.; Leboffe, L.; MacArthur, R.; Mozzi, A. F.; Federici, G.; Caccuri, A. M. *Biochemistry* **2010,** *49,* 226–235. doi:10.1021/bi901689s
50. Khalaf, A. I.; Bourdin, C.; Breen, D.; Donoghue, G.; Scott, F. J.; Suckling, C. J.; Macmillan, D.; Clements, C.; Fox, K.; Sekibo, D. A. *Eur. J. Med. Chem.* **2012,** *56,* 39–47. doi:10.1016/j.ejmech.2012.08.013
51. Baraldi, P. G.; Cozzi, P.; Geroni, C.; Mongelli, N.; Romagnoli, R.; Spalluto, G. *Bioorg. Med. Chem.* **1999,** *7,* 251–262. doi:10.1016/S0968-0896(98)00205-3
52. Baraldi, P. G.; Romagnoli, R.; Beria, I.; Cozzi, P.; Geroni, C.; Mongelli, N.; Bianchi, N.; Mischiati, C.; Gambari, R. *J. Med. Chem.* **2000,** *43,* 2675–2684. doi:10.1021/jm9911229
53. Khalaf, A. I.; Waigh, R. D.; Drummond, A. J.; Pringle, B.; McGroarty, I.; Skellern, G. G.; Suckling, C. J. *J. Med. Chem.* **2004,** *47,* 2133–2156. doi:10.1021/jm031089x
54. James, P. L.; Merkina, E. E.; Khalaf, A. I.; Suckling, C. J.; Waigh, R. D.; Brown, T.; Fox, K. R. *Nucleic Acids Res.* **2004,** *32,* 3410–3417. doi:10.1093/nar/gkh666
55. Lang, S.; Khalaf, A. I.; Breen, D.; Huggan, J. K.; Clements, C. J.; MacKay, S. P.; Suckling, C. J. *Med. Chem. Res.* **2014,** *23,* 1170–1179. doi:10.1007/s00044-013-0723-0
56. Baraldi, P. G.; Cacciari, B.; Guiotto, A.; Leoni, A.; Romagnoli, R.; Spalluto, G.; Mongelli, N.; Howard, P. W.; Thurston, D. E.; Bianchi, N.; Gambari, R. *Bioorg. Med. Chem. Lett.* **1998,** *8,* 3019–3024. doi:10.1016/S0960-894X(98)00544-7
57. Baraldi, P. G.; Romagnoli, R.; Guadix, A. E.; Pineda de las Infantas, M. J.; Gallo, M. A.; Espinosa, A.; Martinez, A.; Bingham, J. P.; Hartley, J. A. *J. Med. Chem.* **2002,** *45,* 3630–3638. doi:10.1021/jm011113b
58. Anthony, N. G.; Breen, D.; Clarke, J.; Donoghue, G.; Drummond, A. J.; Ellis, E. M.; Gemmell, C. G.; Helesbeux, J.-J.; Hunter, I. S.; Khalaf, A. I.; Mackay, S. P.; Parkinson, J. A.; Suckling, C. J.; Waigh, R. D. *J. Med. Chem.* **2007,** *50,* 6116–6125. doi:10.1021/jm070831g
59. Khalaf, A. I.; Anthony, N.; Breen, D.; Donoghue, G.; Mackay, S. P.; Scott, F. J.; Suckling, C. J. *Eur. J. Med. Chem.* **2011,** *46,* 5343–5355. doi:10.1016/j.ejmech.2011.08.035
60. Szerszenowicz, J.; Drozdowska, D. *Molecules* **2014,** *19,* 11300–11315. doi:10.3390/molecules190811300
61. Parkinson, J. A.; Scott, F. J.; Suckling, C. J.; Wilson, G. *MedChemComm* **2013,** *4,* 1105–1108. doi:10.1039/c3md00071k
62. Scott, F. J.; Nichol, R. J. O.; Khalaf, A. I.; Giordani, F.; Gillingwater, K.; Ramu, S.; Elliott, A.; Zuegg, J.; Duffy, P.; Rosslee, M.-J.; Hlaka, L.; Kumar, S.; Ozturk, M.; Brombacher, F.; Barrett, M.; Guler, R.; Suckling, C. J. *Eur. J. Med. Chem.* **2017,** *136,* 561–572. doi:10.1016/j.ejmech.2017.05.039
63. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998,** *5,* 104–109. doi:10.1038/nsb0298-104
64. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998,** *391,* 468–471. doi:10.1038/35106
65. Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998,** *282,* 111–115. doi:10.1126/science.282.5386.111
66. Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998,** *120,* 6219–6226. doi:10.1021/ja980147e
67. Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003,** *13,* 284–299. doi:10.1016/S0959-440X(03)00081-2
68. Hsu, C. F.; Phillips, J. W.; Trauger, J. W.; Farkas, M. E.; Belitsky, J. M.; Heckel, A.; Olenyuk, B. Z.; Puckett, J. W.; Wang, C. C.; Dervan, P. B. *Tetrahedron* **2007,** *63,* 6146–6151. doi:10.1016/j.tet.2007.03.041
69. Wurtz, N. R.; Pomerantz, J. L.; Baltimore, D.; Dervan, P. B. *Biochemistry* **2002,** *41,* 7604–7609. doi:10.1021/bi020114i
70. Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Graves, B. J.; Gottesfeld, J. M. *J. Biol. Chem.* **1999,** *274,* 12765–12773. doi:10.1074/jbc.274.18.12765
71. Raskatov, J. A.; Meier, J. L.; Puckett, J. W.; Yang, F.; Ramakrishnan, P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109,* 1023–1028. doi:10.1073/pnas.1118506109
72. Yang, F.; Nickols, N. G.; Li, B. C.; Marinov, G. K.; Said, J. W.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110,* 1863–1868. doi:10.1073/pnas.1222035110
73. Lenzmeier, B. A.; Baird, E. E.; Dervan, P. B.; Nyborg, J. K. *J. Mol. Biol.* **1999,** *291,* 731–744. doi:10.1006/jmbi.1999.2969
74. Dickinson, L. A.; Burnett, R.; Melander, C.; Edelson, B. S.; Arora, P. S.; Dervan, P. B.; Gottesfeld, J. M. *Chem. Biol.* **2004,** *11,* 1583–1594. doi:10.1016/j.chembiol.2004.09.004
75. Alvarez, D.; Chou, C. J.; Latella, L.; Zeitlin, S. G.; Ku, S.; Puri, P. L.; Dervan, P. B.; Gottesfeld, J. M. *Cell Cycle* **2006,** *5,* 1537–1548. doi:10.4161/cc.5.14.2913
76. Chenoweth, D. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106,* 13175–13179. doi:10.1073/pnas.0906532106
77. Buchmueller, K. L.; Taherbhai, Z.; Howard, C. M.; Bailey, S. L.; Nguyen, B.; O'Hare, C.; Hochhauser, D.; Hartley, J. A.; Wilson, W. D.; Lee, M. *ChemBioChem* **2005,** *6,* 2305–2311. doi:10.1002/cbic.200500179
78. Lai, Y. M.; Fukuda, N.; Ueno, T.; Matsuda, H.; Saito, S.; Matsumoto, K.; Ayame, H.; Bando, T.; Sugiyama, H.; Mugishima, H.; Serie, K. *J. Pharmacol. Exp. Ther.* **2005,** *365,* 571–575. doi:10.1124/jpet.105.089086
79. Wang, S.; Aston, K.; Koeller, K. J.; Harris, G. D., Jr.; Rath, N. P.; Bashkin, J. K.; Wilson, W. D. *Org. Biomol. Chem.* **2014,** *12,* 7523–7536. doi:10.1039/C4OB01456A
80. Kiakos, K.; Pett, L.; Satam, V.; Patil, P.; Hochhauser, D.; Lee, M.; Hartley, J. A. *Chem. Biol.* **2015,** *22,* 862–875. doi:10.1016/j.chembiol.2015.06.005
81. Satam, V.; Babu, B.; Patil, P.; Brien, K. A.; Olson, K.; Savagian, M.; Lee, M.; Mepham, A.; Jobe, L. B.; Bingham, J. P.; Pett, L.; Wang, S.; Ferrara, M.; Bruce, C. D.; Wilson, W. D.; Lee, M.; Hartley, J. A.; Kiakos, K. *Bioorg. Med. Chem. Lett.* **2015,** *25,* 3681–3685. doi:10.1016/j.bmcl.2015.06.055

82. Pjura, P. E.; Grzeskowiak, K.; Dickerson, R. E. *J. Mol. Biol.* **1987,** *197,* 257–271. doi:10.1016/0022-2836(87)90123-9
83. Moreno, T.; Pous, J.; Subirana, J. A.; Campos, J. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66,* 251–257. doi:10.1107/S0907444909055693
84. Li, B. C.; Montgomery, D. C.; Puckett, J. W.; Dervan, P. B. *J. Org. Chem.* **2013,** *78,* 124–133. doi:10.1021/jo302053v
85. Wurtz, N. R.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Org. Lett.* **2001,** *3,* 1201–1203. doi:10.1021/ol0156796
86. Kawamoto, Y.; Bando, T.; Kamada, F.; Li, Y.; Hashiya, K.; Maeshima, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2013,** *135,* 16468–16477. doi:10.1021/ja406737n
87. Guo, C.; Kawamoto, Y.; Asamitsu, S.; Sawatani, Y.; Hashiya, K.; Bando, T.; Sugiyama, H. *Bioorg. Med. Chem.* **2015,** *23,* 855–860. doi:10.1016/j.bmc.2014.12.025
88. Kawamoto, Y.; Sasaki, A.; Chandran, A.; Hashiya, K.; Ide, S.; Bando, T.; Maeshima, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2016,** *138,* 14100–14107. doi:10.1021/jacs.6b09023
89. Nozeret, K.; Loll, F.; Escudé, C.; Boutorine, A. S. *ChemBioChem* **2015,** *16,* 549–554. doi:10.1002/cbic.201402676
90. Nozeret, K.; Bonan, M.; Yarmoluk, S. M.; Novopashina, D. S.; Boutorine, A. S. *Bioorg. Med. Chem.* **2015,** *23,* 5932–5945. doi:10.1016/j.bmc.2015.06.062
91. Wirth-Hamdoune, D.; Ullrich, S.; Scheffer, U.; Radanović, T.; Dürner, G.; Göbel, M. W. *ChemBioChem* **2016,** *17,* 506–514. doi:10.1002/cbic.201500566
92. Haug, R.; Kramer, M.; Richert, C. *Chemistry* **2013,** *19,* 15822–15826. doi:10.1002/chem.201302972
93. Raman, N.; Sobha, S. *Spectrochim. Acta, Part A* **2012,** *85,* 223–234. doi:10.1016/j.saa.2011.09.065
94. Manikandamathavan, V. M.; Rajapandian, V.; Freddy, A. J.; Weyhermuller, T.; Subramanian, V.; Nair, B. U. *Eur. J. Med. Chem.* **2012,** *57,* 449–458. doi:10.1016/j.ejmech.2012.06.039
95. Badia, C.; Souard, F.; Vicent, C. *J. Org. Chem.* **2012,** *77,* 10870–10881. doi:10.1021/jo302238u
96. Blázquez-Sánchez, M. T.; Marcelo, F.; Fernández-Alonso, M. C.; Poveda, A.; Jiménez-Barbero, J.; Vicent, C. *Chem. – Eur. J.* **2014,** *20,* 17640–17652. doi:10.1002/chem.201403911
97. Marković, V.; Debeljak, N.; Stanojković, T.; Kolundžija, B.; Sladić, D.; Vujčić, M.; Janović, B.; Tanić, N.; Perović, M.; Tešić, V.; Antić, J.; Joksović, M. D. *Eur. J. Med. Chem.* **2015,** *89,* 401–410. doi:10.1016/j.ejmech.2014.10.055
98. Khan, G. S.; Pilkington, L. I.; Barker, D. *Bioorg. Med. Chem. Lett.* **2016,** *26,* 804–808. doi:10.1016/j.bmcl.2015.12.090
99. Drozdowska, D.; Rusak, M.; Miltyk, W.; Markowska, A.; Samczuki, P. *Acta Pol. Pharm.* **2016,** *73,* 47–53.
100. Scott, F. J.; Khalaf, A. I.; Duffy, S.; Avery, V. M.; Suckling, C. J. *Bioorg. Med. Chem. Lett.* **2016,** *26,* 3326–3329. doi:10.1016/j.bmcl.2016.05.039
101. Scott, F. J.; Khalaf, A. I.; Giordani, F.; Wong, P. E.; Duffy, S.; Barrett, M.; Avery, V. M.; Suckling, C. J. *Eur. J. Med. Chem.* **2016,** *116,* 116–125. doi:10.1016/j.ejmech.2016.03.064
102. Ahmadi, F.; Valadbeigi, S.; Sajjadi, S. E.; Shokoohinia, Y.; Azizian, H.; Taheripak, G. *Chem.-Biol. Interact.* **2016,** *258,* 89–101. doi:10.1016/j.cbi.2016.08.020
103. Samanta, S. K.; Dutta, D.; Roy, S.; Bhattacharya, K.; Sarkar, S.; Dasgupta, A. K.; Pal, B. C.; Mandal, C.; Mandal, C. *J. Med. Chem.* **2013,** *56,* 5709–5721. doi:10.1021/jm400290q
104. Mitrasinovic, P. M. *J. Chem. Inf. Model.* **2015,** *55,* 421–433. doi:10.1021/ci5006965
105. Chakraborty, B.; Dutta, D.; Mukherjee, S.; Das, S.; Maiti, N. C.; Das, P.; Chowdhury, C. *Eur. J. Med. Chem.* **2015,** *102,* 93–105. doi:10.1016/j.ejmech.2015.07.035
106. Maity, D.; Jiang, J.; Ehlers, M.; Wu, J.; Schmuck, C. *Chem. Commun.* **2016,** *52,* 6134–6137. doi:10.1039/C6CC02138G
107. Alavijeh, N. S.; Zadmard, R.; Balalaie, S.; Alavijeh, M. S.; Soltani, N. *Org. Lett.* **2016,** *18,* 4766–4769. doi:10.1021/acs.orglett.6b01995
108. Shahabadi, N.; Falsafi, M.; Maghsudi, M. *Nucleosides, Nucleotides Nucleic Acids* **2017,** *36,* 49–65. doi:10.1080/15257770.2016.1218021
109. Shahabadi, N.; Hakimi, M.; Morovati, T.; Fatahi, N. *Nucleosides, Nucleotides Nucleic Acids* **2017,** *36,* 497–510. doi:10.1080/15257770.2017.1332370
110. Kennedy, A. R.; Khalaf, A. I.; Scott, F. J.; Suckling, C. J. *Acta Crystallogr., Sect. E: Crystallogr. Commun.* **2017,** *73,* 254–259. doi:10.1107/S2056989017001177
111. Sissi, C.; Dovigo, L.; Greco, M. L.; Ciancetta, A.; Moro, S.; Trzciński, J. W.; Mancin, F.; Rossi, P.; Spalluto, G.; Tecilla, P. *Tetrahedron* **2017,** *73,* 3014–3024. doi:10.1016/j.tet.2017.04.013
112. Chen, A. Y.; Yu, C.; Bodley, A.; Peng, L. F.; Liu, L. F. *Cancer Res.* **1993,** *53,* 1332–1337.
113. Baraldi, P. G.; Bovero, A.; Fruttarolo, F.; Preti, D.; Tabrizi, M. A.; Pavani, M. G.; Romagnoli, R. *Med. Res. Rev.* **2004,** *24,* 475–528. doi:10.1002/med.20000
114. Nelson, S. M.; Ferguson, L. R.; Denny, W. A. *Mutat. Res.* **2007,** *623,* 24–40. doi:10.1016/j.mrfmmm.2007.03.012
115. Yang, Y.-H.; Cheng, M.-S.; Wang, Q.-H.; Nie, H.; Liao, N.; Wang, J.; Chen, H. *Eur. J. Med. Chem.* **2009,** *44,* 1808–1812. doi:10.1016/j.ejmech.2008.07.021
116. Wang, X.-J.; Chu, N.-Y.; Wang, Q.-H.; Liu, C.; Jiang, C.-g.; Wang, X.-Y.; Ikejima, T.; Cheng, M.-S. *Bioorg. Med. Chem. Lett.* **2012,** *22,* 6297–6300. doi:10.1016/j.bmcl.2012.06.102
117. Ivanov, A. A.; Koval, V. S.; Susova, O. Y.; Salyanov, V. I.; Oleinikov, V. A.; Stomakhin, A. A.; Shalginskikh, N. A.; Kvasha, M. A.; Kirsanova, O. V.; Gromova, E. S.; Zhuze, A. L. *Bioorg. Med. Chem. Lett.* **2015,** *25,* 2634–2638. doi:10.1016/j.bmcl.2015.04.087
118. Roopashree, R.; Mohan, C. D.; Swaroop, T. R.; Jagadish, S.; Raghava, B.; Balaji, K. S.; Jayarama, S.; Basappa; Rangappa, K. S. *Bioorg. Med. Chem. Lett.* **2015,** *25,* 2589–2593. doi:10.1016/j.bmcl.2015.04.010
119. Hegde, M.; Kumar, K. S. S.; Thomas, E.; Ananda, H.; Raghavan, S. C.; Rangappa, K. S. *RSC Adv.* **2015,** *5,* 93194–93208. doi:10.1039/C5RA16605E
120. Amirbekyan, K.; Duchemin, N.; Benedetti, E.; Joseph, R.; Colon, A.; Markarian, S. A.; Bethge, L.; Vonhoff, S.; Klussmann, S.; Cossy, J.; Vasseur, J.-J.; Arseniyadis, S.; Smietana, M. *ACS Catal.* **2016,** *6,* 3096–3105. doi:10.1021/acscatal.6b00495
121. Guo, P.; Paul, A.; Kumar, A.; Farahat, A. A.; Kumar, D.; Wang, S.; Boykin, D. W.; Wilson, W. D. *Chem. – Eur. J.* **2016,** *22,* 15404–15412. doi:10.1002/chem.201603422
122. Ranjan, N.; Story, S.; Fulcrand, G.; Leng, F.; Ahmad, M.; King, A.; Sur, S.; Wang, W.; Tse-Dinh, Y.-C.; Arya, D. P. *J. Med. Chem.* **2017,** *60,* 4904–4922. doi:10.1021/acs.jmedchem.7b00191
123. Ranjan, N.; Kellish, P.; King, A.; Arya, D. P. *Biochemistry* **2017,** *56,* 6434–6447. doi:10.1021/acs.biochem.7b00929
124. Picconi, P.; Hind, C.; Jamshidi, S.; Nahar, K.; Clifford, M.; Wand, M. E.; Sutton, J. M.; Rahman, K. M. *J. Med. Chem.* **2017,** *60,* 6045–6059. doi:10.1021/acs.jmedchem.7b00108

125. Mayence, A.; Pietka, A.; Collins, M. S.; Cushion, M. T.; Tekwani, B. L.; Huang, T. L.; Vanden Eynde, J. J. *Bioorg. Med. Chem. Lett.* **2008,** *18,* 2658–2661. doi:10.1016/j.bmcl.2008.03.020
126. Thuita, J. K.; Karanja, S. M.; Wenzler, T.; Mdachi, R. E.; Ngotho, J. M.; Kagira, J. M.; Tidwell, R.; Brun, R. *Acta Trop.* **2008,** *108,* 6–10. doi:10.1016/j.actatropica.2008.07.006
127. Ismail, M. A.; El Bialy, S. A.; Brun, R.; Wenzler, T.; Nanjunda, R.; Wilson, W. D.; Boykin, D. W. *Bioorg. Med. Chem.* **2011,** *19,* 978–984. doi:10.1016/j.bmc.2010.11.047
128. Hu, L.; Arafa, R. K.; Ismail, M. A.; Patel, A.; Munde, M.; Wilson, W. D.; Wenzler, T.; Brun, R.; Boykin, D. W. *Bioorg. Med. Chem.* **2009,** *17,* 6651–6658. doi:10.1016/j.bmc.2009.07.080
129. Suckling, C. *Future Med. Chem.* **2012,** *4,* 971–989. doi:10.4155/fmc.12.52
130. Chai, Y.; Munde, M.; Kumar, A.; Mickelson, L.; Lin, S.; Campbell, N. H.; Banerjee, M.; Akay, S.; Liu, Z.; Farahat, A. A.; Nhili, R.; Depauw, S.; David-Cordonnier, M.-H.; Neidle, S.; Wilson, W. D.; Boykin, D. W. *ChemBioChem* **2014,** *15,* 68–79. doi:10.1002/cbic.201300622
131. Paul, A.; Nanjunda, R.; Kumar, A.; Laughlin, S.; Nhili, R.; Depauw, S.; Deuser, S. S.; Chai, Y.; Chaudhary, A. S.; David-Cordonnier, M.-H.; Boykin, D. W.; Wilson, W. D. *Bioorg. Med. Chem. Lett.* **2015,** *25,* 4927–4932. doi:10.1016/j.bmcl.2015.05.005
132. Laughlin, S.; Wang, S.; Kumar, A.; Farahat, A. A.; Boykin, D. W.; Wilson, W. D. *Chem. – Eur. J.* **2015,** *21,* 5528–5539. doi:10.1002/chem.201406322
133. Farahat, A. A.; Kumar, A.; Say, M.; Wenzler, T.; Brun, R.; Paul, A.; Wilson, W. D.; Boykin, D. W. *Eur. J. Med. Chem.* **2017,** *128,* 70–78. doi:10.1016/j.ejmech.2017.01.037
134. Nagle, P. S.; Rodriguez, F.; Nguyen, B.; Wilson, W. D.; Rozas, I. *J. Med. Chem.* **2012,** *55,* 4397–4406. doi:10.1021/jm300296f
135. Martínez, C. H. R.; Lagartera, L.; Trujillo, C.; Dardonville, C. *MedChemComm* **2015,** *6,* 2036–2042. doi:10.1039/C5MD00292C
136. Kahvedžić, A.; Nathwani, S.-M.; Zisterer, D. M.; Rozas, I. *J. Med. Chem.* **2013,** *56,* 451–459. doi:10.1021/jm301358s
137. McKeever, C.; Kaiser, M.; Rozas, I. *J. Med. Chem.* **2013,** *56,* 700–711. doi:10.1021/jm301614w
138. O'Sullivan, P.; Rozas, I. *ChemMedChem* **2014,** *9,* 2065–2073. doi:10.1002/cmdc.201402264
139. Nagle, P. S.; Rodriguez, F.; Kahvedžić, A.; Quinn, S. J.; Rozas, I. *J. Med. Chem.* **2009,** *52,* 7113–7121. doi:10.1021/jm901017t
140. Stolić, I.; Čipčić Paljetak, H.; Perić, M.; Matijašić, M.; Stepanić, V.; Verbanac, D.; Bajić, M. *Eur. J. Med. Chem.* **2015,** *90,* 68–81. doi:10.1016/j.ejmech.2014.11.003
141. Bordello, J.; Sánchez, M. I.; Vázquez, M. E.; Mascareñas, J. L.; Al-Soufi, W.; Novo, M. *Chem. – Eur. J.* **2015,** *21,* 1609–1619. doi:10.1002/chem.201404926
142. Antony-Debré, I.; Paul, A.; Leite, J.; Mitchell, K.; Kim, H. M.; Carvajal, L. A.; Todorova, T. I.; Huang, K.; Kumar, A.; Farahat, A. A.; Bartholdy, B.; Narayanagari, S.-R.; Chen, J.; Ambesi-Impiombato, A.; Ferrando, A. A.; Mantzaris, I.; Gavathiotis, E.; Verma, A.; Will, B.; Boykin, D. W.; Wilson, W. D.; Poon, G. M. K.; Steidl, U. *J. Clin. Invest.* **2017,** *127,* 4297–4313. doi:10.1172/JCI92504
143. Aich, P.; Dasgupta, D. *Biochemistry* **1995,** *34,* 1376–1385. doi:10.1021/bi00004a032
144. Hou, C.; Weidenbach, S.; Cano, K. E.; Wang, Z.; Mitra, P.; Ivanov, D. N.; Rohr, J.; Tsodikov, O. V. *Nucleic Acids Res.* **2016,** *44,* 8990–9004. doi:10.1093/nar/gkw761
145. Weidenbach, S.; Hou, C.; Chen, J.-M.; Tsodikov, O. V.; Rohr, J. *J. Inorg. Biochem.* **2016,** *156,* 40–47. doi:10.1016/j.jinorgbio.2015.12.011
146. Martinez, R.; Chacon-Garcia, L. *Curr. Med. Chem.* **2005,** *12,* 127–151. doi:10.2174/0929867053363414
147. Lerman, L. S. *J. Mol. Biol.* **1961,** *3,* 18–30. doi:10.1016/S0022-2836(61)80004-1
148. Terakawa,, T.; Miyake, H.; Muramaki, M.; Takenaka, A.; Fujisawa, M. *Int. J. Urol.* **2010,** *17,* 881–885. doi:10.1111/j.1442-2042.2010.02618.x
149. Milano, A.; Apice, G.; Ferrari, E.; Fazioli, F.; de Rosa, V.; de Luna, A. S.; Iaffaioli, R. V.; Caponigro, F. *Crit. Rev. Oncol. Hematol.* **2006,** *59,* 74–84. doi:10.1016/j.critrevonc.2005.12.002
150. He, Y.; Zhang, L.; Song, C. *Int. J. Nanomed.* **2010,** *5,* 697–705. doi:10.2147/IJN.S12129
151. Salas, J. A.; Méndez, C. *Curr. Opin. Chem. Biol.* **2009,** *13,* 152–160. doi:10.1016/j.cbpa.2009.02.003
152. Bailly, C. *Curr. Med. Chem.* **2000,** *7,* 39–58. doi:10.2174/0929867003375489
153. Davies, D. B.; Eaton, R. J.; Baranovsky, S. F.; Veselkov, A. N. *J. Biomol. Struct. Dyn.* **2000,** *17,* 887–901. doi:10.1080/07391102.2000.10506577
154. Frederick, C. A.; Williams, L. D.; Ughetto, G.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Wang, A. H. J. *Biochemistry* **1990,** *29,* 2538–2549. doi:10.1021/bi00462a016
155. Plosker, G. L.; Faulds, D. *Drugs* **1993,** *45,* 788–856. doi:10.2165/00003495-199345050-00011
156. Crow, S. D. G.; Bailly, C.; Garbay-Jaureguiberry, C.; Roques, B.; Shaw, B. R.; Waring, M. J. *Biochemistry* **2002,** *41,* 8672–8682. doi:10.1021/bi012207q
157. Lambert, B.; Segal-Bendirdjian, E.; Esnault, C.; Le Pecq, J. B.; Roques, B. P.; Jones, B.; Yeung, A. T. *Anti-Cancer Drug Des.* **1990,** *5,* 43–53.
158. Leon, P.; Garbay-Jaureguiberry, C.; Barsi, M. C.; Le Pecq, J. B.; Roques, B. P. *J. Med. Chem.* **1987,** *30,* 2074–2080. doi:10.1021/jm00394a024
159. Pelaprat, D.; Delbarre, A.; Le Guen, I.; Roques, B. P.; Le Pecq, J. B.; Roques, B. P. *J. Med. Chem.* **1980,** *23,* 1336–1343. doi:10.1021/jm00186a010
160. Wu, C.-C.; Li, Y.-C.; Wang, Y.-R.; Li, T.-K.; Chan, N.-L. *Nucleic Acids Res.* **2013,** *41,* 10630–10640. doi:10.1093/nar/gkt828
161. Parker, B. S.; Buley, T.; Evison, B. J.; Cutts, S. M.; Neumann, G. M.; Iskander, M. N.; Phillips, D. R. *J. Biol. Chem.* **2004,** *279,* 18814–18823. doi:10.1074/jbc.M400931200
162. Vollmer, T.; Stewart, T.; Baxter, N. *Neurology* **2010,** *74* (Suppl. 1), S41–S46. doi:10.1212/WNL.0b013e3181c97f5a
163. Advani, A. S.; Shadman, M.; Ali-Osman, F.; Barker, A.; Rybicki, L.; Kalaycio, M.; Sekeres, M. A.; de Castro, C. M.; Diehl, L. F.; Moore, J. O.; Beaven, A.; Copelan, E.; Sobecks, R.; Talea, P.; Rizzieri, D. A. *Clin. Lymphoma, Myeloma Leuk.* **2010,** *10,* 473–476. doi:10.3816/CLML.2010.n.082
164. Konda, S. K.; Wang, H.; Cutts, S. M.; Phillips, D. R.; Collins, J. G. *Org. Biomol. Chem.* **2015,** *13,* 5972–5982. doi:10.1039/C5OB00526D
165. Bailly, C.; Qu, X.; Anizon, F.; Prudhomme, M.; Riou, J.-F.; Chaires, J. B. *Mol. Pharmacol.* **1999,** *55,* 377–385. doi:10.1124/mol.55.2.377
166. Facompré, M.; Baldeyrou, B.; Bailly, C.; Anizon, F.; Marminon, C.; Prudhomme, M.; Colson, P.; Houssier, C. *Eur. J. Med. Chem.* **2002,** *37,* 925–932. doi:10.1016/S0223-5234(02)01423-X

167. Gamage, S. A.; Spicer, J. A.; Finlay, G. J.; Stewart, A. J.; Charlton, P.; Baguley, B. C.; Denny, W. A. *J. Med. Chem.* **2001,** *44,* 1407–1415. doi:10.1021/jm0003283
168. Lin, C.; Yang, D. *Curr. Top. Med. Chem.* **2015,** *15,* 1385–1397. doi:10.2174/1568026615666150413155608
169. Sappal, D. S.; McClendon, A. K.; Fleming, J. A.; Thoroddsen, V.; Connolly, K.; Reimer, C.; Blackman, R. K.; Bulawa, C. E.; Osheroff, N.; Charlton, P.; Rudolph-Owen, L. A. *Mol. Cancer Ther.* **2004,** *3,* 47–58.
170. Harris, S. M.; Scott, J. A.; Brown, J. L.; Charlton, P. A.; Mistry, P. *Anticancer Drugs* **2005,** *16,* 945–951. doi:10.1097/01.cad.0000176499.17939.56
171. Harris, S. M.; Mistry, P.; Freathy, C.; Brown, J. L.; Charlton, P. A. *Br. J. Cancer* **2005,** *92,* 722–728. doi:10.1038/sj.bjc.6602403
172. Bailly, C. *Curr. Med. Chem.* **2000,** *7,* 39–58. doi:10.2174/0929867003375489
173. Fujimoto, S. *Biol. Pharm. Bull.* **2007,** *30,* 1923–1929. doi:10.1248/bpb.30.1923
174. Varvaresou, A.; Iakovou, K. *J. Mol. Model.* **2011,** *17,* 2041–2050. doi:10.1007/s00894-010-0891-5
175. Manderville, R. A. *Curr. Med. Chem.: Anti-Cancer Agents* **2001,** *1,* 195–218. doi:10.2174/1568011013354688
176. Bernardes, N.; Seruca, R.; Chakrabarty, A. M.; Fialho, A. M. *Bioeng. Bugs* **2010,** *1,* 178–190. doi:10.4161/bbug.1.3.10903
177. Nguyen, M.; Marcellus, R. C.; Roulston, A.; Watson, M.; Serfass, L.; Murthy Madiraju, S. R.; Goulet, D.; Viallet, J.; Bélec, L.; Billot, X.; Acoca, S.; Purisima, E.; Wiegmans, A.; Cluse, L.; Johnstone, R. W.; Beauparlant, P.; Shore, G. C. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104,* 19512–19517. doi:10.1073/pnas.0709443104
178. Chawrai, S. R.; Williamson, N. R.; Salmond, G. P. C.; Leeper, F. J. *Chem. Commun.* **2008,** 1862–1864. doi:10.1039/b719353j
179. Hu, D. X.; Withall, D. M.; Challis, G. L.; Thomson, R. J. *Chem. Rev.* **2016,** *116,* 7818–7853. doi:10.1021/acs.chemrev.6b00024
180. Forkuo, A. D.; Ansah, C.; Boadu, K. M.; Boampong, J. N.; Ameyaw, E. O.; Gyan, B. A.; Arku, A. T.; Ofori, M. F. *Malar. J.* **2016,** *15,* No. 89. doi:10.1186/s12936-016-1137-5
181. Bonjean, K.; De Pauw-Gillet, M. C.; Defresne, M. P.; Colson, P.; Houssier, C.; Dassonneville, L.; Bailly, C.; Greimers, R.; Wright, C.; Quetin-Leclercq, J.; Tits, M.; Angenot, L. *Biochemistry* **1998,** *37,* 5136–5146. doi:10.1021/bi972927q
182. Lisgarten, J. N.; Coll, M.; Portugal, J.; Wright, C. W.; Aymami, J. *Nat. Struct. Biol.* **2002,** *9,* 57–60. doi:10.1038/nsb729
183. Dar, A. M.; Shamsuzzaman; Gatoo, M. A. *Steroids* **2015,** *104,* 163–175. doi:10.1016/j.steroids.2015.09.010
184. Spatafora, C.; Barresi, V.; Bhusainahalli, V. M.; Di Micco, S.; Musso, N.; Riccio, R.; Bifulco, G.; Condorelli, D.; Tringali, C. *Org. Biomol. Chem.* **2014,** *12,* 2686–2701. doi:10.1039/c3ob42521e
185. Nagle, P. S.; McKeever, C.; Rodriguez, F.; Nguyen, B.; Wilson, W. D.; Rozas, I. *J. Med. Chem.* **2014,** *57,* 7663–7672. doi:10.1021/jm5008006
186. Matsui, T.; Sugiyama, H.; Nakai, M.; Nakabayashi, Y. *Chem. Pharm. Bull.* **2016,** *64,* 282–286. doi:10.1248/cpb.c15-00903
187. Marangoci, N.-L.; Popovici, L.; Ursu, E.-L.; Danac, R.; Clima, L.; Cojocaru, C.; Coroaba, A.; Neamtu, A.; Mangalagiu, I.; Pinteala, M.; Rotaru, A. *Tetrahedron* **2016,** *72,* 8215–8222. doi:10.1016/j.tet.2016.10.052
188. Hamilton, P. L.; Arya, D. P. *Nat. Prod. Rep.* **2012,** *29,* 134–143. doi:10.1039/C1NP00054C
189. Arya, D. P.; Willis, B. *J. Am. Chem. Soc.* **2003,** *125,* 12398–12399. doi:10.1021/ja036742k
190. Kumar, S.; Xue, L.; Arya, D. P. *J. Am. Chem. Soc.* **2011,** *133,* 7361–7375. doi:10.1021/ja108118v
191. Arya, D. P.; Xue, L.; Willis, B. *J. Am. Chem. Soc.* **2003,** *125,* 10148–10149. doi:10.1021/ja035117c
192. Xi, H.; Davis, E.; Ranjan, N.; Xue, L.; Hyde-Volpe, D.; Arya, D. P. *Biochemistry* **2011,** *50,* 9088–9113. doi:10.1021/bi201077h
193. Arya, D. P.; Coffee, R. L.; Xue, L. *Bioorg. Med. Chem. Lett.* **2004,** *14,* 4643–4646. doi:10.1016/j.bmcl.2004.07.002
194. Kumar, S.; Spano, M. N.; Arya, D. P. *Biopolymers* **2014,** *101,* 720–732. doi:10.1002/bip.22448
195. Kumar, S.; Spano, M. N.; Arya, D. P. *Bioorg. Med. Chem.* **2015,** *23,* 3105–3109. doi:10.1016/j.bmc.2015.04.082
196. Kumar, S.; Kellish, P.; Robinson, W. E.; Wang, D.; Appella, D. H.; Arya, D. P. *Biochemistry* **2012,** *51,* 2331–2347. doi:10.1021/bi201657k
197. Kumar, S.; Arya, D. P. *Bioorg. Med. Chem. Lett.* **2011,** *21,* 4788–4792. doi:10.1016/j.bmcl.2011.06.058
198. Ranjan, N.; Arya, D. P. *Bioorg. Med. Chem. Lett.* **2016,** *26,* 5989–5994. doi:10.1016/j.bmcl.2016.10.076
199. Willis, B.; Arya, D. P. *Biochemistry* **2006,** *45,* 10217–10232. doi:10.1021/bi0609265
200. Povirk, L. F.; Shuker, D. E. *Mutat. Res.* **1994,** *318,* 205–226. doi:10.1016/0165-1110(94)90015-9
201. Fong, L. Y. Y.; Bevill, R. F.; Thurmon, J. C.; Magee, P. N. *Carcinogenesis* **1992,** *13,* 2153–2159. doi:10.1093/carcin/13.11.2153
202. Bhuyan, B. K.; Newell, K. A.; Crampton, S. L.; Von Hoff, D. D. *Cancer Res.* **1982,** *42,* 3532–3537.
203. Boger, D. L.; Johnson, D. S. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92,* 3642–3649. doi:10.1073/pnas.92.9.3642
204. Baraldi, P. G.; Balboni, G.; Pavani, M. G.; Spalluto, G.; Tabrizi, M. A.; De Clercq, E.; Balzarini, J.; Bando, T.; Sugiyama, H.; Romagnoli, R. *J. Med. Chem.* **2001,** *44,* 2536–2543. doi:10.1021/jm0108404
205. Rayburn, E.; Wang, W.; Li, M.; Zhang, X.; Xu, H.; Li, H.; Qin, J.-J.; Jia, L.; Covey, J.; Lee, M.; Zhang, R. *Cancer Chemother. Pharmacol.* **2012,** *69,* 1423–1431. doi:10.1007/s00280-012-1851-9
206. Wolfe, A. L.; Duncan, K. K.; Parelkar, N. K.; Weir, S. J.; Vielhauer, G. A.; Boger, D. L. *J. Med. Chem.* **2012,** *55,* 5878–5886. doi:10.1021/jm300330b
207. Giddens, A. C.; Lee, H. H.; Lu, G. L.; Miller, C. K.; Guo, J.; Lewis Phillips, G. D.; Pillow, T. H.; Tercel, M. *Bioorg. Med. Chem.* **2016,** *24,* 6075–6081. doi:10.1016/j.bmc.2016.09.068
208. McMorris, T. C.; Staake, M. D.; Kelner, M. J. *J. Org. Chem.* **2004,** *69,* 619–623. doi:10.1021/jo035084j
209. Lin, J.; Hiraoka, K.; Watanabe, T.; Kuo, T.; Shinozaki, Y.; Takatori, A.; Koshikawa, N.; Chandran, A.; Otsuki, J.; Sugiyama, H.; Horton, P.; Nagase, H. *PLoS One* **2016,** *11,* e0165581. doi:10.1371/journal.pone.0165581
210. D'Incalci, M.; Galmarini, C. M. *Mol. Cancer Ther.* **2010,** *9,* 2157–2163. doi:10.1158/1535-7163.MCT-10-0263
211. Elez, M. E.; Tabernero, J.; Geary, D.; Macarulla, T.; Kang, S. P.; Kahatt, C.; Pita, A. S.; Teruel, C. F.; Siguero, M.; Cullell-Young, M.; Szyldergemajn, S.; Ratain, M. J. *Clin. Cancer Res.* **2014,** *20,* 2205–2214. doi:10.1158/1078-0432.CCR-13-1880
212. Petek, B. J.; Jones, R. L. *Molecules* **2014,** *19,* 12328–12335. doi:10.3390/molecules190812328
213. Buchanan, M. K.; Needham, C. N.; Neill, N. E.; White, M. C.; Kelly, C. B.; Mastro-Kishton, K.; Chauvigne-Hines, L. M.; Goodwin, T. J.; McIver, A. L.; Bartolotti, L. J.; Frampton, A. R.; Bourdelais, A. J.; Varadarajan, S. *Biochemistry* **2017,** *56,* 421–440. doi:10.1021/acs.biochem.6b01075

214. Kohn, K. W.; Bono, V. H., Jr.; Kann, H. E., Jr. *Biochim. Biophys. Acta* **1968,** *155,* 121–129. doi:10.1016/0005-2787(68)90342-0
215. Smellie, M.; Kelland, L. R.; Thurston, D. E.; Souhami, R. L.; Hartley, J. A. *Br. J. Cancer* **1994,** *70,* 48–53. doi:10.1038/bjc.1994.248
216. Thurston, D. E.; Bose, D. S.; Thompson, A. S.; Howard, P. W.; Leoni, A.; Croker, S. J.; Jenkins, T. C.; Neidle, S.; Hartley, J. A.; Hurley, L. H. *J. Org. Chem.* **1996,** *61,* 8141–8147. doi:10.1021/jo951631s
217. Kamal, A.; Murali Mohan Reddy, P. S.; Reddy, D. R.; Laxman, E. *Bioorg. Med. Chem.* **2006,** *14,* 385–394. doi:10.1016/j.bmc.2005.08.020
218. Brucoli, F.; Hawkins, R. M.; James, C. H.; Wells, G.; Jenkins, T. C.; Ellis, T.; Hartley, J. A.; Howard, P. W.; Thurston, D. E. *Bioorg. Med. Chem. Lett.* **2011,** *21,* 3780–3783. doi:10.1016/j.bmcl.2011.04.054
219. Kamal, A.; Sreekanth, K.; Shankaraiah, N.; Sathish, M.; Nekkanti, S.; Srinivasulu, V. *Bioorg. Chem.* **2015,** *59,* 23–30. doi:10.1016/j.bioorg.2015.01.002

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Oligonucleotide analogues with cationic backbone linkages

Melissa Meng and Christian Ducho*

Review

Open Access

Address:
Department of Pharmacy, Pharmaceutical and Medicinal Chemistry, Saarland University, Campus C2 3, 66123 Saarbrücken, Germany

Email:
Christian Ducho* - christian.ducho@uni-saarland.de

* Corresponding author

Keywords:
backbone modifications; cations; DNA; oligonucleotides; zwitterions

Beilstein J. Org. Chem. **2018,** *14,* 1293–1308.
doi:10.3762/bjoc.14.111

Received: 05 January 2018
Accepted: 26 April 2018
Published: 04 June 2018

This article is part of the Thematic Series "Nucleic acid chemistry II".

Guest Editor: H.-A. Wagenknecht

Abstract

Their unique ability to selectively bind specific nucleic acid sequences makes oligonucleotides promising bioactive agents. However, modifications of the nucleic acid structure are an essential prerequisite for their application in vivo or even in cellulo. The oligoanionic backbone structure of oligonucleotides mainly hampers their ability to penetrate biological barriers such as cellular membranes. Hence, particular attention has been given to structural modifications of oligonucleotides which reduce their overall number of negative charges. One such approach is the site-specific replacement of the negatively charged phosphate diester linkage with alternative structural motifs which are positively charged at physiological pH, thus resulting in zwitterionic or even oligocationic backbone structures. This review provides a general overview of this concept and summarizes research on four according artificial backbone linkages: aminoalkylated phosphoramidates (and related systems), guanidinium groups, *S*-methylthiourea motifs, and nucleosyl amino acid (NAA)-derived modifications. The synthesis and properties of the corresponding oligonucleotide analogues are described.

Introduction

Oligonucleotides have the unique ability to bind endogenous nucleic acids in a selective and sequence-specific manner. They can therefore modulate biological functions via different mechanisms [1]. Single-stranded oligonucleotides (ONs) can act in cellulo mainly via two different pathways (Figure 1). In the antigene pathway [2], the ON enters the nucleus and binds to double-stranded DNA to form a triple helix. The triple helix is not a substrate for the transcription machinery, and hence, RNA biosynthesis (and therefore protein formation) is blocked. In the antisense pathway [3], the ON binds to single-stranded mRNA in the cytoplasm, thus furnishing a duplex structure (usually a DNA–RNA heteroduplex) which cannot undergo ribosomal protein biosynthesis. Alternatively, the DNA–RNA heteroduplex can be a substrate for RNAse H-mediated degradation of the mRNA strand. This way, catalytic amounts of the ON can mediate the efficient cleavage of mRNA encoding a specific protein, which leads to effective (though reversible) and selective downregulation of the protein's activity. A third option for

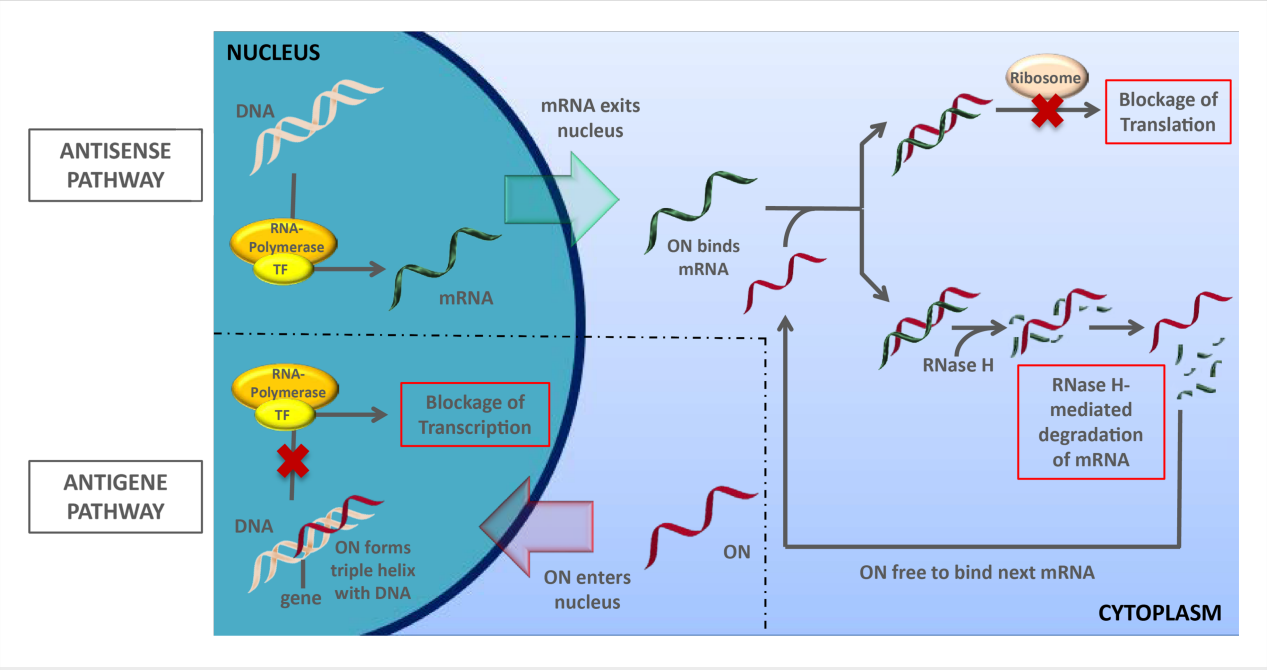


Figure 1: Biological action of single-stranded oligonucleotides (ON): antigene and antisense pathways.

the biological action of oligonucleotide structures is the triggering of the RNA interference mechanism by double-stranded 'small interfering' RNA (siRNA, mechanism not shown) [4]. Alternatively, single-stranded oligonucleotides (anti-miRNA oligonucleotides, 'AMOs', 'antimiRs') can inhibit endogenous microRNA-mediated RNA interference by blocking the RNA strand in the involved protein–RNA complex (RISC) [5].

The capability of ONs to exert the aforementioned biological mechanisms via sequence-specific molecular recognition makes them highly attractive candidates for drug development. However, their pharmacokinetic properties are problematic and represent a significant hurdle for their therapeutic application. First, the high polarity of ONs, mainly caused by their oligoanionic phosphate diester backbone, severely hampers the penetration of biological barriers such as cellular membranes, thus leading to low cellular uptake. Second, unmodified ON structures are good substrates for nuclease-mediated degradation. Consequently, it is of vital importance to chemically modify ON structures in order to make them suitable drug candidates or chemical probes, e.g., for diagnostic purposes [6,7].

The relevance of the polyanionic phosphate diester-linked backbone to the overall function of nucleic acids has been discussed by Westheimer [8], Benner [9,10], and others. In spite of these considerations, many artificial internucleotide linkages were investigated in order to reduce the overall negative charge of the backbone and to enhance nuclease stability. One apparent approach to achieve these goals is the introduction of non-native electroneutral backbone linkages, with the nucleic acid mimic 'peptide nucleic acid' (PNA) [11-13] representing a striking example. Although the achiral PNA backbone is pronouncedly different from native nucleic acid structures, PNAs are capable of sequence-specific hybridization to native nucleic acids. However, their moderate water solubility and peptide-like folding properties [9] are hurdles for their biological application. As an alternative strategy, the (deoxy)ribose part of the backbone has been retained and only some of the internucleotide phosphate diesters have been selectively replaced by electroneutral motifs. Such artificial neutral linkages include, among others, sulfone [14], amide [15-22], triazole [23-27], phosphoramidate [28] and phosphate triester [29] moieties.

Using a different approach, positive charges have been introduced into nucleic acid structures. Positively charged moieties were either employed (i) as additional charged structural motifs compensating for the negative charges in the backbone linkages or (ii) as replacements of the native negatively charged phosphate diester linkages. The first option has found considerable attention, with positively charged moieties attached to nucleobases or the ribose sugar. Some selected examples **1**–**6** of resulting nucleic acid structures are provided in Figure 2 [30-37]. Oligonucleotides of this type are at least partially zwitterionic, but overall densely charged. With respect to the aspired improvement of cellular uptake, fully cationic oligonucleotide analogues might also be attractive candidate structures, as indicated by the advantageous properties of cationic cell-penetrating peptides (CPPs) [38]. However, the design of

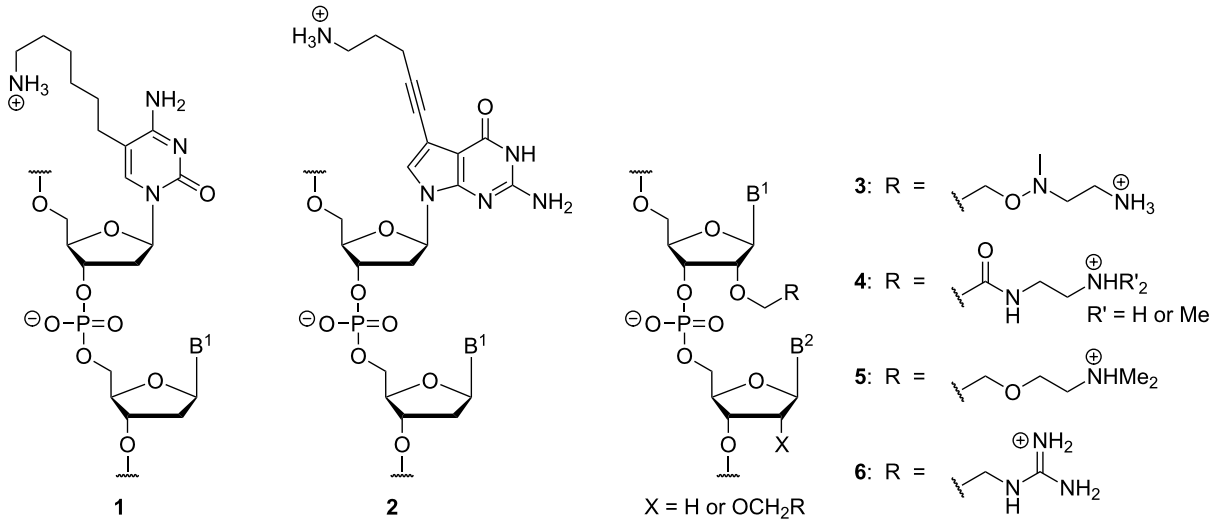


Figure 2: Selected examples **1**–**6** of nucleic acid modifications based on additionally attached positively charged moieties, but retaining an intact phosphate diester backbone (B^1, B^2 = nucleobases) [30-37].

modifications of type **1**–**6** precludes the preparation of fully cationic oligonucleotide analogues.

This review focusses on the second aforementioned option to employ cationic motifs in oligonucleotide structures, i.e., as replacements of the native phosphate diester linkages [39]. In principle, this approach enables the preparation of partially or fully zwitterionic as well as cationic backbones. This strategy has been studied less frequently, with research on four artificial cationic linkages summarized in this review: aminoalkylated phosphoramidates (and related systems), guanidinium groups, *S*-methylthiourea motifs, and nucleosyl amino acid (NAA)-derived modifications. The synthesis and properties of the corresponding oligonucleotide analogues of types **7**–**10** (Figure 3) with cationic backbone linkages are described.

Review

Aminoalkyl phosphoramidate linkages and related systems

Pioneering work in the field has been reported by Letsinger and co-workers. In 1986, they introduced a deoxyadenosyl dinucleotide linked by an aminoethyl phosphoramidate moiety which is positively charged under acidic and neutral conditions [40]. Based on these results, they subsequently reported the synthesis of short, cationic DNA oligonucleotides with phosphoramidate linkages of type **7**, which were N-alkylated with substituents containing basic structural motifs [41].

The synthesis of the modified deoxyadenosyl dinucleotide **11** was achieved using solution-phase chemistry (reactions not shown, for structure of **11** see Scheme 1) [40]. Subsequently,

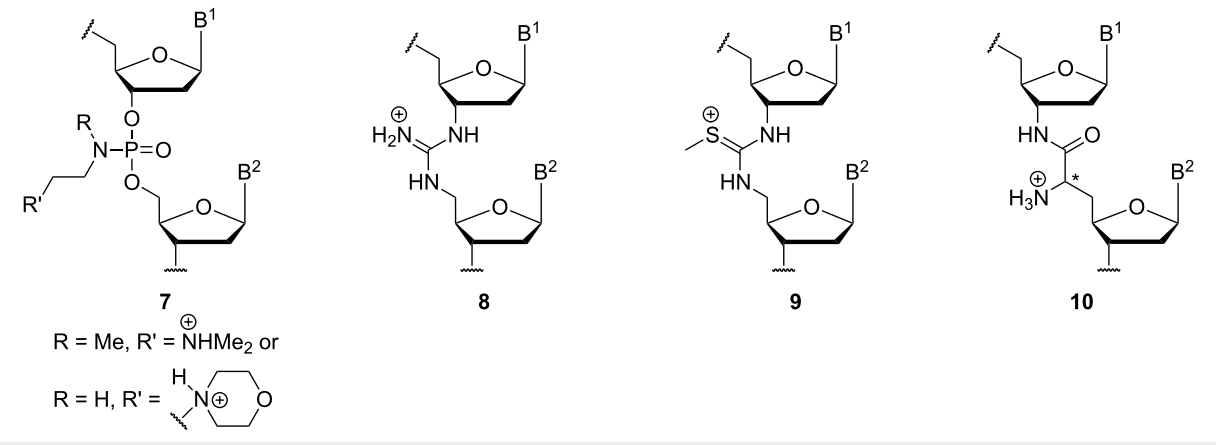
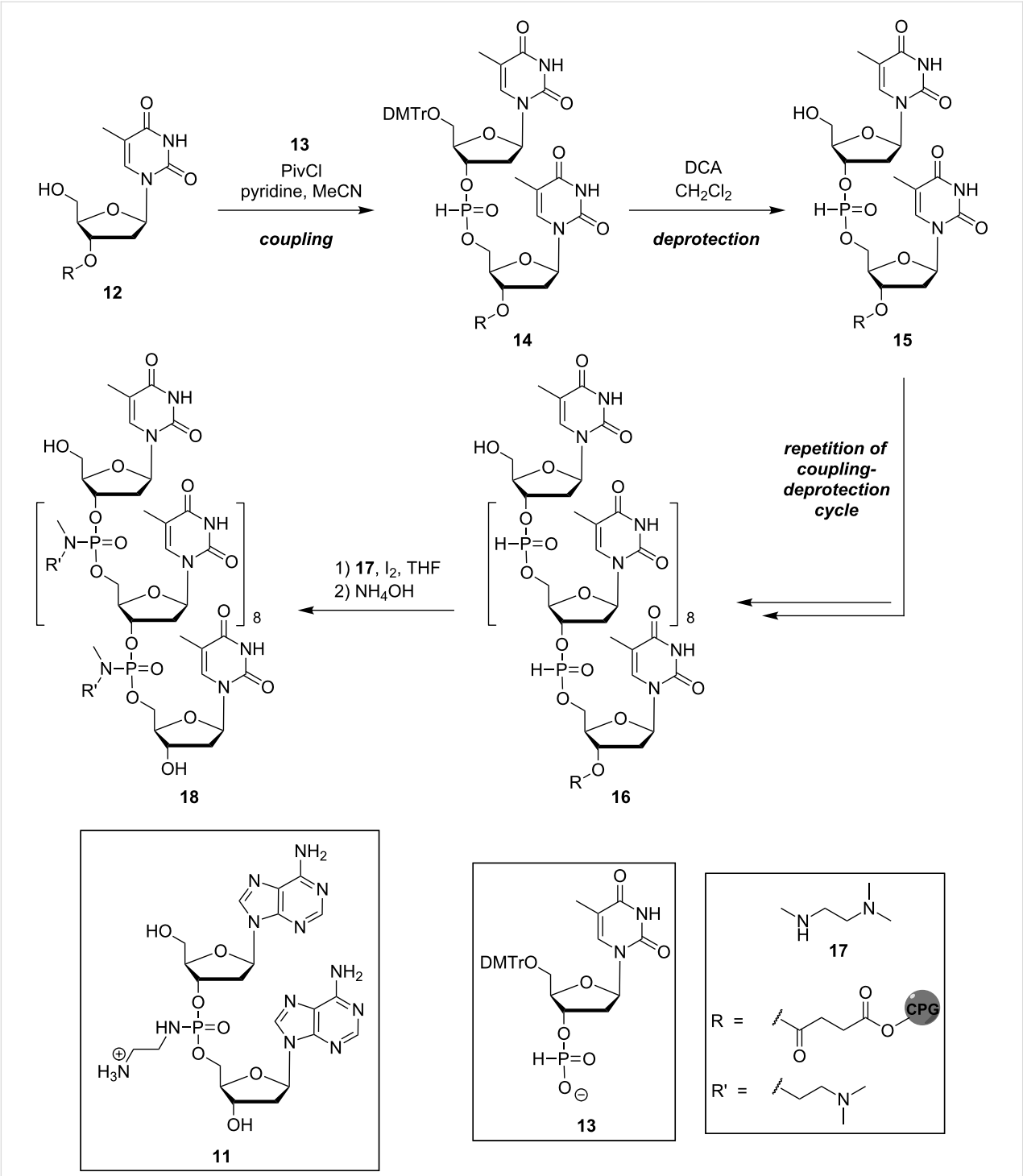


Figure 3: Oligonucleotide analogues with artificial cationic backbone linkages discussed in this review: aminoalkylated phosphoramidates **7** (and related systems, not shown), guanidinium-linked 'DNG' **8**, *S*-methylthiourea-linked oligomers **9**, and nucleosyl amino acid (NAA)-modified oligonucleotides **10** (B^1, B^2 = nucleobases).



Scheme 1: Structure of Letsinger's modified deoxyadenosyl dinucleotide **11** and synthesis of cationic oligonucleotide analogue **18** containing amino-alkyl phosphoramidate linkages. CPG = controlled pore glass (solid support).

the preparation of corresponding oligonucleotide analogues was performed on solid support using H-phosphonate chemistry (Scheme 1). Thus, solid phase-linked thymidine **12** was coupled with 5'-dimethoxytrityl-(DMTr)-protected thymidine 3'-H-phosphonate **13** to give dimeric H-phosphonate **14**, which was then acidically DMTr-deprotected to furnish **15**. After the desired number of such coupling-deprotection cycles, the phosphite-linked oligo-thymidine **16** was transformed in an oxidative amidation reaction [42] in the presence of iodine and *N*,*N*,*N*′-trimethylethylenediamine (**17**) to yield, after basic cleavage

from the solid support, the envisioned aminoalkyl phosphoramidate-linked oligonucleotide **18**.

To study the hybridization properties of such cationic oligonucleotide analogues with native DNA and RNA, Letsinger and co-workers performed UV-monitored thermal denaturation experiments [40,41]. In the case of the modified deoxyadenosyl dimer **11**, hybridization with native RNA-T_{Poly} as well as with DNA-T_{Poly} strands was evident and the complex formed more stable than comparable complexes involving the native d(ApA) DNA reference. An increase of the measured T_m of ≈10 °C for complexes of the aminoethyl phosphoramidate-linked dinucleoside **11** with RNA and ≈25 °C for the according hybridization with DNA was observed [40]. In addition, the cationic dimer **11** was shown to bind more tightly to native RNA and DNA strands in the presence of magnesium chloride [40].

For the cationic T-oligomer **18**, Letsinger and co-workers reported a strongly reduced absorbance of a mixture of **18** with DNA-A_{Poly} in thermal melting studies, as compared to the non-hybridized, single-stranded oligonucleotides [41]. This indicated a successful complex formation with ordered base stacking of the positively charged oligonucleotide analogue and its native DNA counterstrand. When exposed to high ionic strength (1.0 M NaCl), the complex was shown to undergo a significant decrease in stability. This effect of high salt concentrations was inverse to the corresponding effect for native anionic DNA duplexes and obviously resulted from electrostatic shielding mediated by the salt ions, thus weakening the attraction of the oppositely charged backbones [41].

In order to elucidate the stability of aminoethyl phosphoramidate-linked oligonucleotides to nuclease-catalysed degradation, Letsinger and co-workers described the incubation of such oligomers, the deoxyadenosyl dimer **11** and DNA-T_{Poly} (as a reference) with snake venom phosphodiesterase and spleen phosphodiesterase, respectively [40,41]. In these assays, neither the modified dimer **11** nor oligonucleotides of type **7** (such as **18**) showed any degradation by either enzyme, while native DNA reference strands were rapidly cleaved.

Other groups have subsequently employed Letsinger's aminoalkyl phosphoramidate linkage (or variations thereof) in biochemical and biological studies on the properties of corresponding oligonucleotides. Weeks and co-workers have demonstrated that a triplex-forming antigene oligonucleotide modified with a variant of Letsinger's linkages can efficiently inhibit the expression of plasmid DNA injected into *Xenopus* oocytes [43]. The presence of the cationic backbone modification and a sufficiently long mismatch-free target DNA sequence were essential for this gene-silencing effect, thus indicating the relevance of enhanced nuclease stability and sequence-specific DNA binding. However, the gene-silencing effect could only be achieved if the modified oligonucleotide and the plasmid DNA were either mixed prior to cellular injection or if the oligonucleotide was injected first, pointing out a likely competition of the cationic antigene oligonucleotide with cellular histones for DNA binding [43].

Vasseur, Debart and co-workers have combined a variant of Letsinger's linkages with an α-configuration at the anomeric centers of antisense oligonucleotides [44,45]. They have found that such zwitterionic to fully cationic α-oligonucleotides bound to single-stranded DNA and RNA targets with high affinity, with duplex stabilization being proportional to the number of cationic modifications. It was also reported that these oligonucleotides showed retained base pairing fidelity, i.e., the T_m value was significantly reduced in the presence of a base mismatch. This specificity in binding suggested that such oligonucleotides should be promising sterically blocking antisense agents as their RNA targets were not digested by RNAse H. This anticipated bioactivity was confirmed in whole cell assays without the presence of transfection agents, suggesting that the altered charge pattern of the oligonucleotide backbone enabled its cellular self-delivery [44]. The same authors then also studied similar oligonucleotides with guanidinium groups as cationic moieties, which were obtained by postsynthetic guanidinylation of the congeners with amino-functionalized phosphoramidate linkages (reaction not shown) [46]. The presence of the guanidinium units furnished high hybridization affinities, in particular with single-stranded RNA targets, and also in triplex formation with double-stranded DNA, though the amino-functionalized analogues gave similar triplex stabilities. A fully cationic and fluorescently labelled guanidinylated oligonucleotide was subjected to comparative cellular uptake studies. Relative to its fluorescently labelled anionic phosphorothioate congener, it showed vastly enhanced cellular uptake. Fluorescence microscopy revealed a cytoplasmic localization of the oligonucleotide without accumulation in the nuclei. This indicated an endocytotic uptake mechanism with (at least partial) retention of the material in the endocytotic vesicles. No unspecific cytotoxic effect of the guanidinylated oligonucleotide was observed.

Other types of oligonucleotides with aminoalkyl moieties as part of artificial internucleotide linkages have also been reported. With respect to their structural similarity to Letsinger's aminoalkyl phosphoramidate linkages, these variants are categorized as 'related systems' in this review. Fathi et al. have established the aminoethylphosphonate linkage **19** (i.e., a phosphonate analogue of amidate **7**) [47], and Rahman, Obika and

co-workers have described cationic phosphorothioates of type **20** [48] (Figure 4).

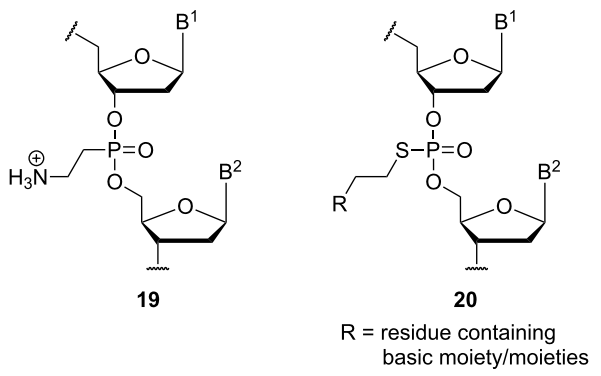


Figure 4: Artificial cationic backbone linkages **19** and **20** which are structurally related to aminoalkylated phosphoramidates of type **7** (B^1, B^2 = nucleobases).

The preparation of phosphonate linkage **19** was achieved in diastereomerically pure form, i.e., with defined configuration at the stereogenic phosphorus atom [47]. Corresponding R_P-configured zwitterionic oligonucleotides formed duplexes with complementary DNA or RNA that were more stable than their respective native counterparts. The modified oligonucleotides showed pronounced nuclease and serum stability as well as significantly enhanced cellular uptake relative to their native congeners. As for the aforementioned phosphoramidates, fluorescence microscopy indicated a cytoplasmic localization of the tested zwitterionic oligonucleotide without significant accumulation in the nuclei, thus pointing to endocytotic uptake with retention of the compound in endocytotic vesicles (vide supra).

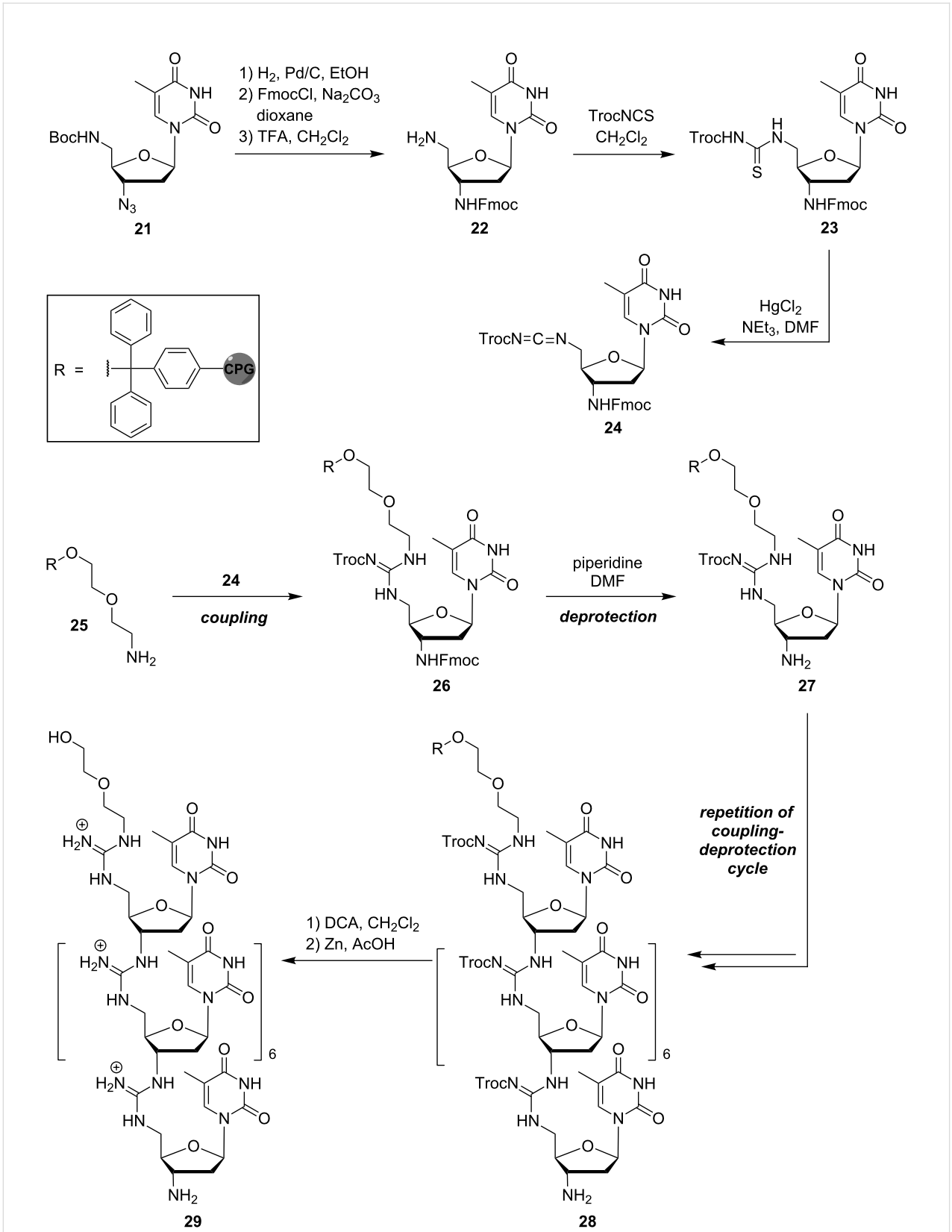
Cationically functionalized phosphorothioates of type **20** were also prepared as diastereomerically pure compounds with defined configuration at the stereogenic phosphorothioate unit [48]. A series of different residues (R in Figure 4) bearing one or two basic amino functionalities was introduced. The resulting 12-mer oligonucleotides with one cationic internucleotide linkage (all other linkages were phosphates) were tested for their ability to form duplexes with single-stranded DNA or RNA as well as triplexes with double-stranded DNA. The aminoalkylated R_P-phosphorothioates showed an increased stability of DNA duplexes while the S_P-isomers gave destabilized duplexes. Both the cationically functionalized R_P- and S_P-oligonucleotides displayed decreased affinity towards RNA, while triplex formation was enhanced for all tested R_P congeners. The aminoalkylation generally provided an increased nuclease stability, which was more pronounced for the R_P isomers.

Deoxyribonucleic guanidines (DNG) with guanidinium linkages

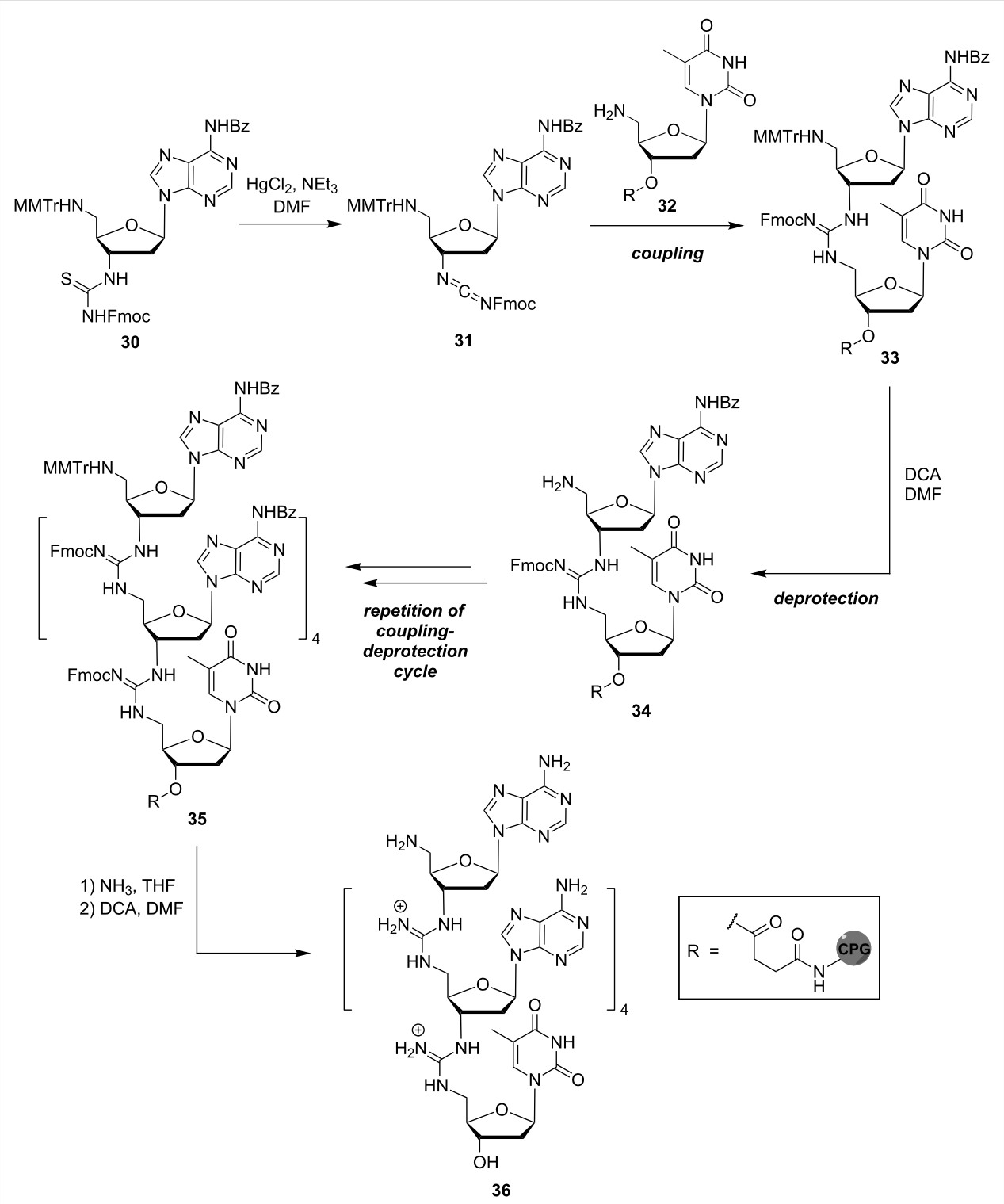
In their design of cationic oligonucleotide analogues, Bruice et al. did not just attach a cationic moiety to the modified phosphate diester backbone, but they completely replaced it with a guanidinium linkage to give 'deoxyribonucleic guanidines (DNG)' of type **8** [49]. The guanidinium group was selected owing to its maintenance of a positive charge over a broad pH range and its ability to form both intermolecular electrostatic interactions and hydrogen bonds [50]. Letsinger's aminoalkyl phosphoramidate modification was stereogenic at the phosphorus atom, thus leading to complex mixtures of diastereomeric oligomers (with the exception of the aforementioned related systems, vide supra) as the stereoselective synthesis of stereogenic phosphate derivatives is challenging. Therefore, achiral artificial linkages such as guanidinium groups may be considered advantageous from a stereochemical perspective.

For the first synthesis of a pentameric thymidinyl DNG in 1996, Bruice and co-workers used an iterative solution-phase protocol (reactions not shown) [51]. This method was associated with some limitations, such as its moderate yields and the need for purification after each synthetic step. Subsequently, two different approaches for the solid phase-supported synthesis of DNG oligomers were introduced. They enabled chain elongation either in the 5'→3' [52] or 3'→5' [53] direction, respectively. Starting from protected 3',5'-dideoxy-5'-amino-3'-azidothymidine **21**, the 5'→3' route was based on the synthesis of the diamino intermediate **22** and thiourea monomer **23**, which was then converted into a reactive carbodiimide **24** and coupled to a terminal amino group of the solid phase **25** (Scheme 2). This coupling furnished solid phase-attached intermediate **26**, which was Fmoc-deprotected to the amine **27**. Iterative repetition of this coupling-deprotection cycle gave oligomer **28**, which was then acidically cleaved from the solid support and reductively Troc-deprotected to afford octameric thymidinyl DNG **29**.

Based on this method, the solid phase-supported synthesis operating in the 3'→5' direction was later developed. As described by Bruice and co-workers, it was compatible with the cleavage conditions used in the solid phase-supported synthesis of native DNA and also allowed the introduction not only of pyrimidine, but also of purine bases into the oligonucleotide analogue [53]. The method was based on the activation of the 5′-monomethoxytrityl (MMTr)-protected 3'-thiourea monomer **30** to the corresponding carbodiimide **31** (Scheme 3). Using long-chain alkylamine controlled pore glass (CPG) loaded with 5′-amino-5′-deoxythymidine (**32**) as solid phase, the reaction cycle started with the guanidine-forming coupling of **31** and **32** to give **33**, followed by acidic cleavage of the MMTr protecting group to yield the free 5'-amine **34**. Subsequent iterative coupling–depro-



Scheme 2: Bruice's synthesis of guanidinium-linked DNG oligomer **29** in the 5'→3' direction (Troc = 2,2,2-trichloroethyloxycarbonyl).



Scheme 3: Bruice's synthesis of purine-containing guanidinium-linked DNG oligomer **36** in the 3'→5' direction (MMTr = monomethoxytrityl).

tection cycles resulted in the formation of the guanidinium-linked oligomer **35**. After basic guanidine and purine deprotection and concomitant cleavage from the solid support, final acidic deprotection furnished A_5T oligonucleotide analogue **36**.

In addition to these protocols, the solid phase-supported syntheses of DNG-DNA chimeras with partially zwitterionic backbone structures [54,55] as well as of further mixed DNG sequences [56,57] have been described (reactions not shown). It is

also noteworthy that Bruice and co-workers succeeded in the preparation of corresponding guanidine-linked RNA analogues [58,59], though this is not within the main scope of this review.

Bruice et al. reported that oligonucleotide analogues containing the cationic DNG-modification bind to DNA with retention of base-pairing fidelity, furnishing thermally highly stable complexes with native complementary DNA and RNA counterstrands [51,60-64]. The increase in melting temperature for the DNG-DNA complex was reported to be around 15–25 °C per bp under nearly physiological conditions, dependent on the surrounding ionic strength. As shown by Job plot analysis, an oligo-thymidinyl DNG forms triple-stranded complexes in a 2:1 mixture with its native DNA counterstrand, i.e., the resulting triplex contains two DNG oligo-thymidylate analogues and one oligo-adenylate DNA strand [64]. The same binding stoichiometry was observed for an oligo-deoxyadenosyl DNG in complex with a native oligo-thymidylate DNA [53]. Overall, the obtained results suggest that adenosine- and thymidine-derived DNG oligomers support the formation of triplex structures, but that the DNG-DNA ratio within the complex is determined by the respective nucleobases. Remarkably, neither cytidinyl nor 7-deazaguanyl DNG oligomers furnish triplexes, but bind their complementary DNA counterstrand in a 1:1 ratio [65,66]. Furthermore, it was shown that an increase in ionic strength shields the oppositely charged backbones, thus destabilizing both DNG-DNA duplexes and triple-stranded DNG-DNA complexes, respectively. The triple-stranded DNG-DNA complex was less affected than its duplex congener though [51,60,61].

Regarding base-pairing fidelity, Bruice and co-workers have reported significantly reduced stabilities of DNG-DNA duplexes and triplexes, respectively, upon the insertion of base mismatches in the DNA counterstrand. Analyzing a 2:1 complex formed from two octameric thymidinyl DNG strands and one native DNA A_8-mer, they concluded that base mismatches at either end of the DNA counterstrand sequence do not hamper hybridization as strongly as a single base mismatch in the center of the DNA strand. Two base mismatches in the center of the DNA counterstrand led to a complete loss of hybridization [64].

In addition to these thermal denaturation experiments, Bruice et al. also reported circular dichroism (CD) spectroscopic studies to obtain further information on the solution structures of DNG strands and their complexes with DNA. The corresponding analysis of the aforementioned triplex (DNG-T_8)$_2$/DNA-A_8 indicated a usual B-DNA-derived triple helix structure, while the comparison of single-stranded DNG-T_8 with native DNA-T_8 furnished two very different CD spectra [64].

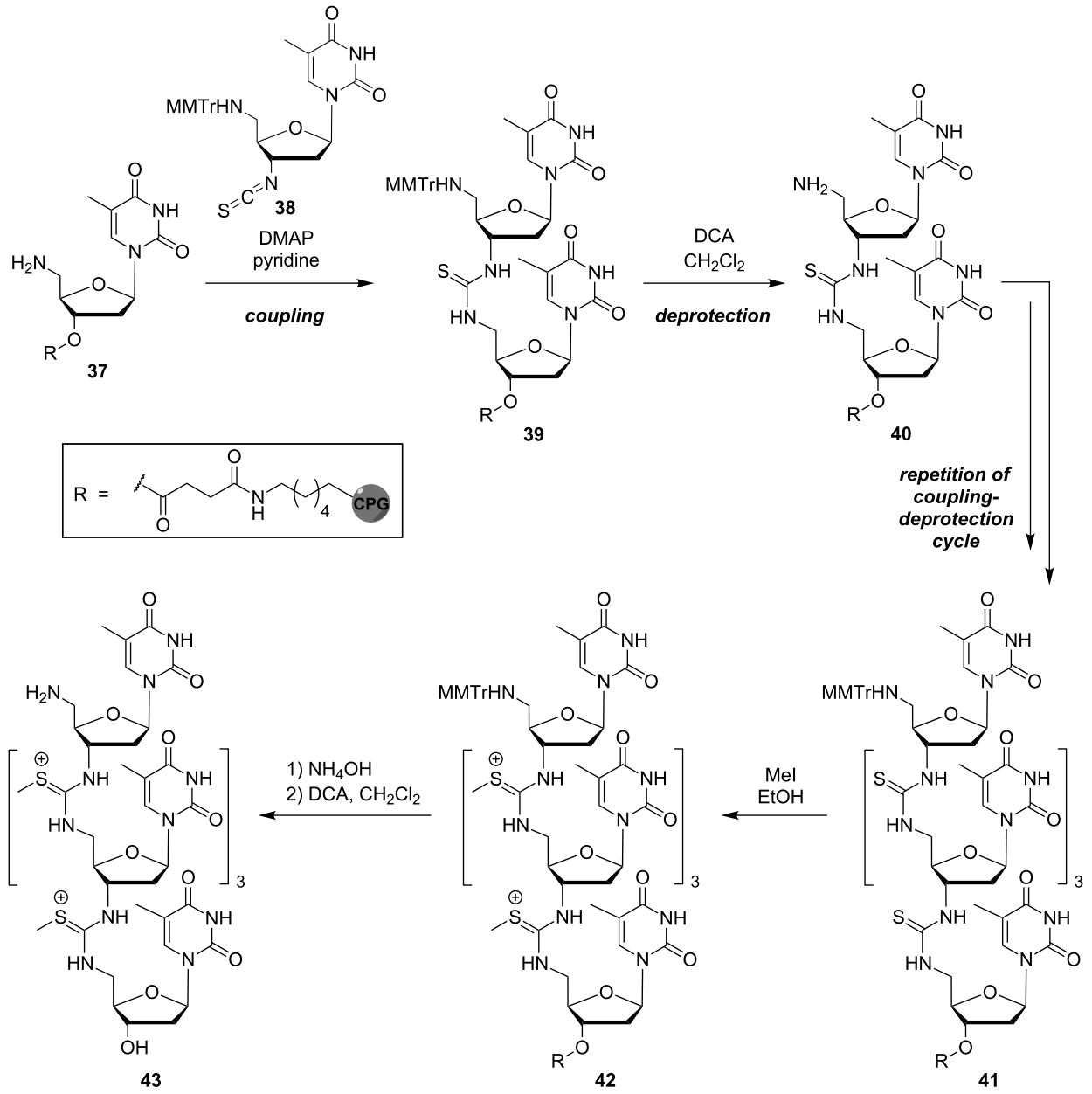
S-Methylthiourea linkages

In addition to their work on DNG oligonucleotide analogues, Bruice et al. also reported the positively charged *S*-methylthiourea backbone modification as an artificial internucleotide linkage [67-69]. For oligomers containing this replacement of the backbone phosphate diesters, the term 'DNmts' was coined. Just like the guanidinium linkage in DNGs, the *S*-methylthiourea modification is not stereogenic and stable towards nuclease-mediated cleavage. Furthermore, it retains its positive charge independent of pH conditions.

Bruice and co-workers initially reported a solution-phase synthesis that enabled the formation of pentameric thymidinyl DNmt in the 3'→5' direction (reactions not shown) [68]. They then introduced an automated solid phase-supported synthesis which was compatible with standard techniques of DNA synthesis (Scheme 4) [69]. A derivative of 5'-amino-5'-deoxythymidine attached to CPG (**37**) served as the solid phase. The construction of the oligomer, achieved in 3'→5' direction, was based on the coupling of 3'-isothiocyanate **38** with the 5'-amino group of **37** to give **39** and, after acidic MMTr cleavage, **40**. Iterative repetition of this coupling-deprotection cycle afforded thiourea-linked oligonucleotide analogue **41**. Subsequent reaction of the thiourea internucleotide linkages with methyl iodide furnished the protected *S*-methylthiourea-linked oligomer **42** and finally, after cleavage from the solid support and acidic deprotection, the envisioned DNmt oligomer **43**.

As for the pentameric DNG congener (vide supra), the DNmt-T_5 oligonucleotide analogue was shown to bind more tightly to complementary DNA than DNA itself [68]. Under nearly physiological conditions with respect to pH and ionic strength, the T_m value for the DNmt-T_5/DNA-A_{Poly} complex was reported to be above 80 °C whereas a comparable DNA–DNA duplex was only stable up to 13 °C. DNmt-T_5 complexes with native RNA-A_{Poly} showed an even higher thermal stability. Job plot analysis revealed the formation of triple-stranded complexes between the DNmt pentamer and DNA-A_{Poly} or RNA-A_{Poly}, respectively [68,70]. Similar to the results obtained for DNG-T_5 (vide supra), a triplex with 2:1 stoichiometry (DNmt:DNA and DNmt:RNA, respectively) was confirmed.

Remarkably, Bruice et al. identified two different hyperchromic shifts for the DNmt-T_5/DNA-A_{Poly} complex, but not for comparable DNmt-RNA aggregates when these mixtures were exposed to higher ionic strength, denoting the thermal denaturation of the (DNmt-T_5)$_2$/DNA-A_{Poly} triplex and, subsequently, the DNmt-DNA duplex. However, the corresponding melting temperatures were significantly lower than T_m values measured in aqueous solutions with physiological ionic



Scheme 4: Bruice's synthesis of *S*-methylthiourea-linked DNmt oligomer **43**.

strength. This indicates a pronounced destabilization of the DNmt-DNA complex with increasing ionic strength [70]. Comparable DNmt-RNA complexes were less destabilized under identical conditions.

Bruice and co-workers also performed further thermal denaturation studies to elucidate base-pairing fidelity of the pentameric thymidinyl DNmt. No increase in hyperchromicity was observed for combinations of DNmt-T_5 with either DNA-G_{Poly}, DNA-C_{Poly} or DNA-T_{Poly}, over a temperature range from 5 to 93 °C, thus ruling out complex formation with these fully mismatched native DNA counterstrands. Furthermore, a pronounced drop in thermal stability of DNmt–DNA complexes containing 50% T–C mismatches and also for congeners containing 20% T–C mismatches was described [71].

In CD spectroscopic studies performed on the thymidinyl DNmt pentamer, Bruice et al. further confirmed the base-pairing specificity of oligonucleotides containing the artificial *S*-methylthiourea backbone linkage [70,71]. CD spectra of DNmt-T_5 in complex with five different DNA oligonucleotides containing an increasing number of C mismatches showed significant

changes dependent on the mismatch content. While the combination of DNmt-T_5 with DNA-A_{20} resulted in a CD difference spectrum with distinct amplitude, the addition of DNA oligonucleotides with an increasing number of C mismatches led to continuous slackening of signals in the difference spectra, until those were almost flat for DNA oligonucleotides containing 50% C mismatches. Hence, this indicates that the ability of the DNmt pentamer to associate with a native DNA oligomer is dependent on Watson–Crick base pairing and is severely hampered by an increasing amount of base-pairing mismatches.

Nucleosyl amino acid (NAA)-derived linkages

Both Letsinger's and Bruice's approaches for the introduction of positive charges into artificial backbone linkages have characteristic conformational features. Letsinger's aminoalkyl phosphoramidate modification and related systems involve a pronounced conformational flexibility of the moieties carrying the positively charged groups. Hence, it cannot be ruled out that interactions with the phosphate groups occur which would be less likely if the positively charged units were more rigidly fixed to the backbone. In contrast, both Bruice's DNG and DNmt oligonucleotide analogues are characterized by conformationally rigid internucleotide linkages. Apparently, an alternative strategy providing a positively charged backbone linkage with 'intermediate' conformational flexibility is missing.

These considerations have stimulated our design of a new artificial internucleotide linkage named 'nucleosyl amino acid (NAA)-modification' (Figure 5) [72-74]. In principle, the NAA-modification is inspired by 'high-carbon' nucleoside structures (i.e., nucleosides having more than five carbon atoms in the sugar unit) found in naturally occurring nucleoside antibiotics [75-77]. In muraymycin- and caprazamycin-type nucleoside antibiotics, among others, such 'high-carbon' nucleosides are uridine-derived amino acid structures ('glycyluridine', GlyU) [78-80], which are aminoribosylated at the 5'-hydroxy group. As part of our ongoing research program on muraymycin nucleoside antibiotics (e.g., muraymycin A1 (**44**)) and their analogues [81-88], we have reported the synthesis of simplified (i.e., 5'-defunctionalized) GlyU derivatives of type **45** (Figure 5) [86-88]. The formal amalgamation of this 'nucleosyl amino acid (NAA)' structure **45** with previously reported amide internucleotide linkages of types **46** and **47** [15-22] furnished the structure of an 'NAA-modified oligonucleotide' **48** (Figure 5). The 6'-amino group of the NAA-modification is positively charged at physiological pH values, thus providing a (partially) zwitterionic backbone structure if some phosphate diester units are replaced with the NAA-modification. In the NAA-modification, several rotatable bonds are combined with the rigid amide group, and it is therefore expected to represent an example of the aforementioned positively charged backbone linkage with 'intermediate' conformational flexibility (vide supra).

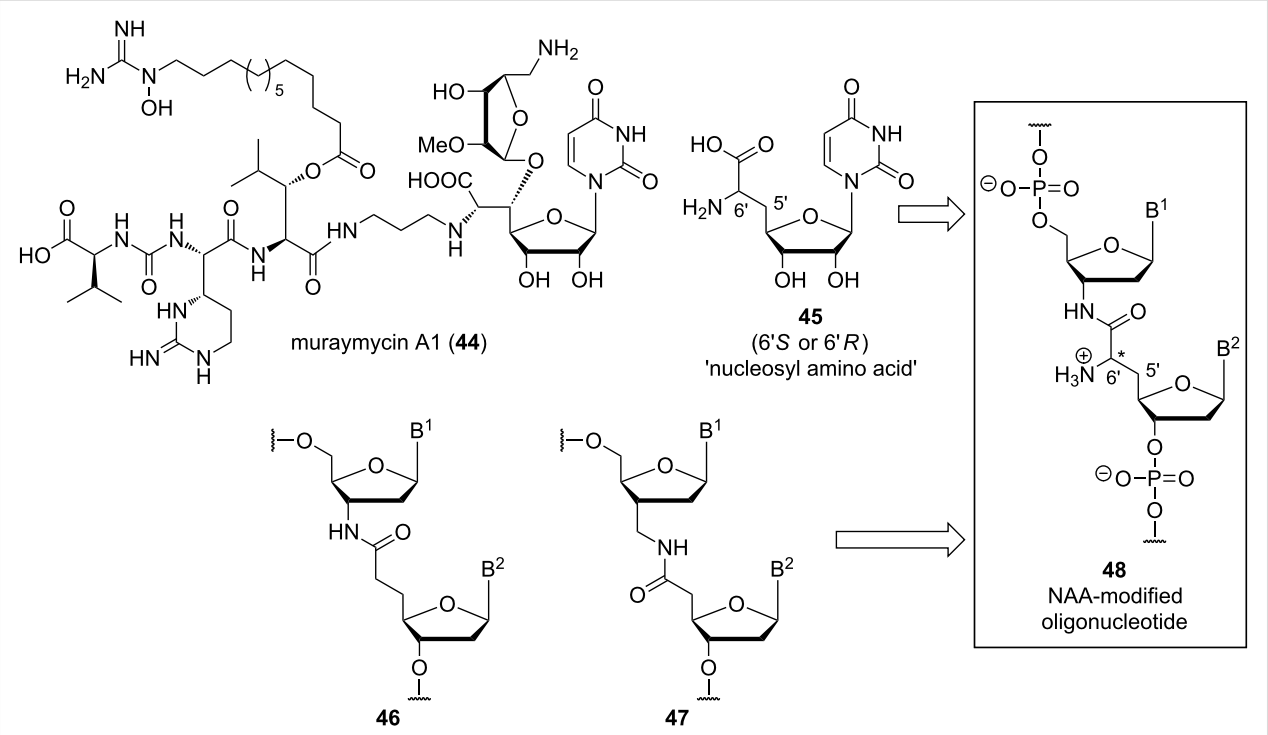
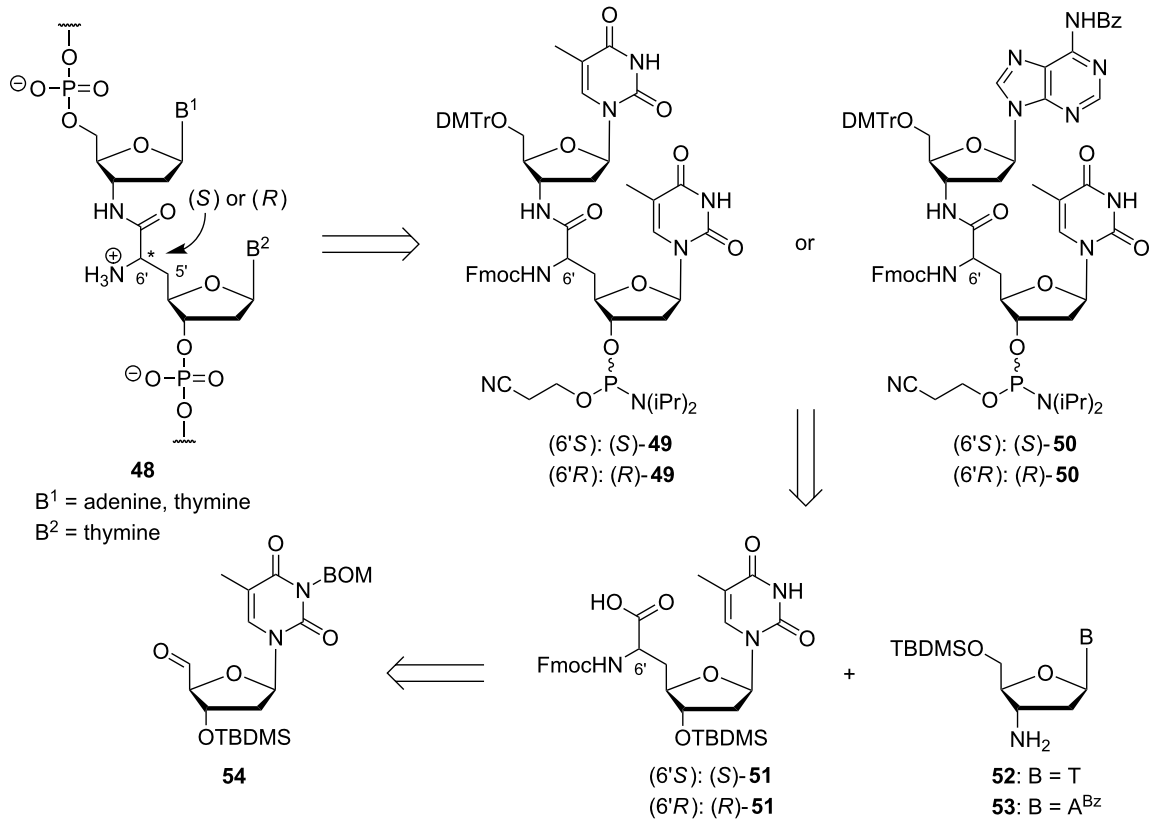


Figure 5: Structure of the natural product muraymycin A1 (**44**) and design concept of nucleosyl amino acid (NAA)-modified (partially) zwitterionic oligonucleotides of type **48** formally derived from structures **45**–**47** (B^1, B^2 = nucleobases).

We have reported that partially zwitterionic NAA-modified DNA oligonucleotides can be obtained by standard solid phase-supported automated DNA synthesis if 'dimeric' phosphoramidite building blocks **49** and **50** are employed (Scheme 5) [72,73]. For the synthesis of 'dimeric' phosphoramidites **49** and **50**, protected thymidinyl amino acids (*S*)-**51** or (*R*)-**51** were coupled with protected 3'-amino-3'-deoxythymidine **52** or protected 3'-amino-2',3'-dideoxyadenosine **53** [73,89], respectively. Thymidinyl amino acids **51** were obtained from 3'-*O*-silylated thymidine-5'-aldehyde **54** via a previously established route using Wittig–Horner olefination and catalytic asymmetric hydrogenation as key steps (reactions not shown) [86,87,90-92].

Using 'dimeric' building blocks (*S*)-**49**, (*R*)-**49**, (*S*)-**50**, and (*R*)-**50** (Scheme 5), automated DNA synthesis under standard conditions enabled the preparation of partially zwitterionic NAA-modified oligonucleotides with defined configuration at the 6'-position, i.e., with control over the spatial orientation of the positive charge [72,73]. Thus, the NAA-modification was placed in T–T ('TxT', with x representing the NAA-linkage) and A–T segments ('AxT') of the oligonucleotide sequence, respectively. Further variation of the 3'-aminonucleoside component (**52** and **53** in Scheme 5) should potentially also allow the introduction of the NAA-modification at C–T and G–T sites within a given sequence.

So far, 24 different oligonucleotides with one to four TxT NAA-modifications at various positions [72] as well as two oligonucleotides with two AxT NAA-modifications [73] have been reported. The properties of the TxT-containing congeners have been studied in detail [72]. Thermal denaturation experiments showed that the TxT NAA-modified DNA oligonucleotides formed duplexes with complementary native DNA or RNA counterstrands, but with moderate destabilization relative to unmodified native duplexes, in particular for DNA–RNA hybrids. The fidelity of base pairing was studied using native DNA counterstrands containing a single base mismatch. Furthermore, structures of the duplexes were investigated by CD spectroscopy. The following properties of TxT NAA-modified DNA oligonucleotides were reported [72]: (i) they formed reasonably stable duplexes with complementary counterstrands, in particular with native DNA; (ii) the influence of the spatial



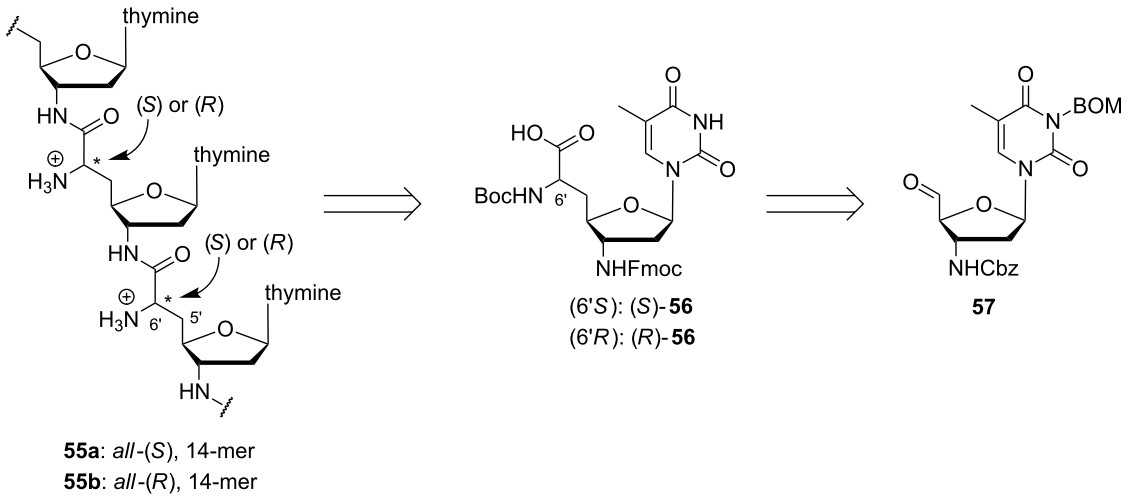
Scheme 5: Retrosynthetic summary of Ducho's synthesis of partially zwitterionic NAA-modified oligonucleotides **48** (BOM = benzyloxymethyl).

orientation of the positive charge, i.e., of the configuration at the 6'-position, was moderate, with a tendency that (6'*R*)-configured linkages furnished slightly more stable duplexes; (iii) the modified oligonucleotides showed no impairment of mismatch discrimination, i.e., single base mismatches led to a significant drop in duplex stability; (iv) the formed duplexes were devoid of significant structural distortion, i.e., their CD spectra indicated B-type helices for DNA–DNA duplexes and A-type helices for DNA–RNA duplexes. Overall, these results demonstrated that typical chemical properties of nucleic acids are retained in partially zwitterionic NAA-modified DNA oligonucleotides. However, corresponding studies on NAA-modified DNA oligonucleotides with a fully zwitterionic backbone have not been conducted yet.

With respect to the aforementioned favourable properties of zwitterionic NAA-modified oligonucleotides, the obvious aim was to synthesize fully cationic oligomers, i.e., oligonucleotide analogues with the cationic NAA-modification as their sole internucleotide linkage. The phosphoramidite-based synthetic strategy depicted in Scheme 5 was not suitable to reach this goal as it furnishes phosphate diester linkages at least at every second position within a given sequence. Therefore, a different synthetic route was developed (Scheme 6) [74]. The envisioned fully cationic thymidine-derived oligomers **55a** (*all*-(*S*)-configured at the 6'-positions) and **55b** (*all*-(*R*)-configured at the 6'-positions) were assembled by manual Fmoc-based solid phase-supported peptide synthesis using the monomeric 3'-amino-nucleosyl amino acids (*S*)-**56** and (*R*)-**56**, respectively, as building blocks. The synthesis of thymidinyl amino acids **56** was again started from a corresponding 5'-aldehyde **57** using Wittig–Horner olefination and catalytic asymmetric hydrogenation as key steps (reactions not shown) [74].

The properties of fully cationic oligonucleotide analogues **55a** and **55b** were studied in detail [74]. Thermal denaturation experiments demonstrated a strong hybridization of both thymidinyl oligomers with native complementary A_{14} DNA, with T_m values being 9 and 17 °C higher, respectively, than the T_m value of an unmodified T_{14}–A_{14} DNA reference duplex. As anticipated based on Letsinger's and Bruice' work (vide supra), the T_m value of the **55**–DNA complex decreased with increasing ionic strength. Studies on base-pairing fidelity gave the remarkable result that both **55a** and **55b** were largely insensitive to the presence of a single base mismatch in the counterstrand, thus indicating that electrostatic attraction overruled Watson–Crick base-pairing specificity in these cases. CD spectroscopy indicated that both **55a** and **55b** formed double-helical duplex structures with complementary DNA, apparently with slight distortions in case of the **55b**–DNA duplex.

The hampered base-pairing fidelity of **55a** and **55b** raised the question if the hybridization of these oligocations with oligoanionic DNA was dependent on Watson–Crick base-pairing at all or if it was mainly mediated by electrostatic attraction. Thermal denaturation studies of mixtures of **55a** or **55b**, respectively, with a fully mismatched DNA counterstrand (G_6TTG_6) showed a pronounced hyperchromicity upon heating in both cases, but also indicated that no transition between two defined states occurred [74]. It was derived from these results that **55a** and **55b** probably formed less defined, unspecific aggregates with the fully mismatched counterstrand, which then disassembled at elevated temperatures. This hypothesis was further supported by CD-spectroscopic studies. The overall conclusion was that the formation of defined double-helical duplex structures of **55a** and **55b** with DNA was mainly steered by Watson–Crick base-



Scheme 6: Retrosynthetic summary of Ducho's and Grossmann's synthesis of fully cationic NAA-modified oligonucleotides **55a** and **55b**.

pairing, but that unspecific electrostatic attraction also contributed to the hybridization of the strands.

Conclusion

In summary, this review provides an overview of four different approaches to introduce cationic backbone linkages as replacements of the phosphate diester units into oligonucleotide structures: i) aminoalkylated phosphoramidates and related systems; ii) guanidinium groups; iii) *S*-methylthiourea motifs and iv) nucleosyl amino acid (NAA)-derived modifications. All of these artificial internucleotide linkages are accessible by means of chemical synthesis, which is either based on the application of H-phosphonate (for i) or phosphoramidite-based (for iv) DNA synthesis, or on a massively modified version of DNA synthesis (for ii and iii), or on solid phase-supported peptide synthesis (for iv).

Studies on the properties of resulting oligomers are not fully conclusive yet. Some data, for instance on base-pairing fidelity, are missing for Letsinger's originally reported aminoalkylated phosphoramidates, while subsequently reported variants thereof and related systems have been studied in more detail. Thus, both retained base-pairing fidelity and improved cellular uptake have been reported for some oligonucleotides with structural similarity to Letsinger's first-generation aminoalkylated phosphoramidates. Bruice's guanidinium- and *S*-methylthiourea-linked systems have a pronounced tendency to form triple-helical structures with native nucleic acids, which makes a direct comparison with the other approaches difficult. Bruice's data suggest retained base-pairing fidelity for fully cationic oligomers, which is in remarkable contrast to our results obtained for NAA-modified oligonucleotides. The latter showed excellent base-pairing fidelity in the case of partially zwitterionic backbones, but insensitivity to single base mismatches for the hybridization of fully cationic oligomers with native DNA. Recently reported results on such fully cationic NAA oligomers [74] indicate that in addition to Watson–Crick base-pairing, unspecific electrostatic attraction also plays a role in the hybridization process. Overall, one must state that the interplay of the structural and conformational properties of cationic internucleotide linkages and the physicochemical behaviour of corresponding oligomers in their binding to anionic nucleic acids is only scarcely understood and will require further research efforts.

Studies on the biological properties of (partially) zwitterionic and cationic oligonucleotide analogues in cellular systems, in particular with respect to their cellular uptake, are currently only available for some aminoalkylated phosphoramidate-linked oligonucleotides and a related phosphonate analogue. The anticipated vast improvement of cellular uptake due to the presence of the cationic internucleotide linkages was proven for these systems, even though they displayed hampered endosomal release. On the other hand, our results on NAA-derived cationic oligomers suggest that, as a paradigm for the design of cationic oligonucleotide analogues for biological applications, one should potentially be cautious with respect to the number of positive charges in the backbone: base-pairing fidelity might be hampered, dependent on the structure of the artificial internucleotide linkage. It will therefore also be of significant relevance to further investigate the influence of the charge pattern in the backbone on the oligonucleotides' cellular uptake. The stage is set to perform such studies, which will further advance the development of cationically linked oligonucleotide analogues for potential applications as drug candidates, diagnostic agents or chemical tool compounds.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG, grant DU 1095/2-1) and the Fonds der Chemischen Industrie (FCI, Sachkostenzuschuss) for financial support of our research on NAA-modified oligonucleotides.

ORCID® iDs

Christian Ducho - https://orcid.org/0000-0002-0629-9993

References

1. Sharma, V. K.; Rungta, P.; Prasad, A. K. *RSC Adv.* **2014,** *4,* 16618–16631. doi:10.1039/c3ra47841f
2. Xodo, L. E.; Cogoi, S.; Rapozzi, V. *Curr. Pharm. Des.* **2004,** *10,* 805–819. doi:10.2174/1381612043452983
3. Wagner, R. W. *Nature* **1994,** *372,* 333–335. doi:10.1038/372333a0
4. Mello, C. C. *Angew. Chem.* **2007,** *119,* 7114–7124. doi:10.1002/ange.200701713
   Angew. Chem., Int. Ed. **2007,** *46,* 6985–6994. doi:10.1002/anie.200701713
5. Lennox, K. A.; Behlke, M. A. *Gene Ther.* **2011,** *18,* 1111–1120. doi:10.1038/gt.2011.100
6. Cobb, A. J. A. *Org. Biomol. Chem.* **2007,** *5,* 3260–3275. doi:10.1039/b709797m
7. Deleavey, G. F.; Damha, M. *J. Chem. Biol.* **2012,** *19,* 937–954. doi:10.1016/j.chembiol.2012.07.011
8. Westheimer, F. H. *Science* **1987,** *235,* 1173–1178. doi:10.1126/science.2434996
9. Benner, S. A.; Hutter, D. *Bioorg. Chem.* **2002,** *30,* 62–80. doi:10.1006/bioo.2001.1232
10. Benner, S. A. *Acc. Chem. Res.* **2004,** *37,* 784–797. doi:10.1021/ar040004z
11. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991,** *254,* 1497–1500. doi:10.1126/science.1962210
12. Zhilina, Z. V.; Ziemba, A. J.; Ebbinghaus, S. W. *Curr. Top. Med. Chem.* **2005,** *5,* 1119–1131. doi:10.2174/156802605774370892
13. Karkare, S.; Bhatnagar, D. *Appl. Microbiol. Biotechnol.* **2006,** *71,* 575–586. doi:10.1007/s00253-006-0434-2
14. Huang, Z.; Benner, S. A. *J. Org. Chem.* **2002,** *67,* 3996–4013. doi:10.1021/jo0003910

15. Lebreton, J.; De Mesmaeker, A.; Waldner, A.; Fritsch, V.; Wolf, R. M.; Freier, S. M. *Tetrahedron Lett.* **1993,** *34,* 6383–6386. doi:10.1016/0040-4039(93)85051-W
16. De Mesmaeker, A.; Lebreton, J.; Waldner, A.; Fritsch, V.; Wolf, R. M.; Freier, S. M. *Synlett* **1993,** 733–736. doi:10.1055/s-1993-22588
17. De Mesmaeker, A.; Waldner, A.; Lebreton, J.; Hoffmann, P.; Fritsch, V.; Wolf, R. M.; Freier, S. M. *Angew. Chem.* **1994,** *106,* 237–240. doi:10.1002/ange.19941060230
Angew. Chem., Int. Ed. **1994,** *33,* 226–229. doi:10.1002/anie.199402261
18. Lebreton, J.; Waldner, A.; Fritsch, V.; Wolf, R. M.; De Mesmaeker, A. *Tetrahedron Lett.* **1994,** *35,* 5225–5228. doi:10.1016/S0040-4039(00)77069-9
19. Rozners, E.; Katkevica, D.; Bizdena, E.; Strömberg, R. *J. Am. Chem. Soc.* **2003,** *125,* 12125–12136. doi:10.1021/ja0360900
20. Selvam, C.; Thomas, S.; Abbott, J.; Kennedy, S. D.; Rozners, E. *Angew. Chem.* **2011,** *123,* 2116–2118. doi:10.1002/ange.201007012
Angew. Chem., Int. Ed. **2011,** *50,* 2068–2070. doi:10.1002/anie.201007012
21. Tanui, P.; Kennedy, S. D.; Lunstad, B. D.; Haas, A.; Leake, D.; Rozners, E. *Org. Biomol. Chem.* **2014,** *12,* 1207–1210. doi:10.1039/C3OB42532K
22. Mutisya, D.; Selvam, C.; Lunstad, B. D.; Pallan, P. S.; Haas, A.; Leake, D.; Egli, M.; Rozners, E. *Nucleic Acids Res.* **2014,** *42,* 6542–6551. doi:10.1093/nar/gku235
23. El-Sagheer, A. H.; Brown, T. *J. Am. Chem. Soc.* **2009,** *131,* 3958–3964. doi:10.1021/ja8065896
24. El-Sagheer, A. H.; Sanzone, A. P.; Gao, R.; Tavassoli, A.; Brown, T. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108,* 11338–11343. doi:10.1073/pnas.1101519108
25. Sanzone, A. P.; El-Sagheer, A. H.; Brown, T.; Tavassoli, A. *Nucleic Acids Res.* **2012,** *40,* 10567–10575. doi:10.1093/nar/gks756
26. Birts, C. N.; Sanzone, A. P.; El-Sagheer, A. H.; Blaydes, J. P.; Brown, T.; Tavassoli, A. *Angew. Chem.* **2014,** *126,* 2394–2397. doi:10.1002/ange.201308691
Angew. Chem., Int. Ed. **2014,** *53,* 2362–2365. doi:10.1002/anie.201308691
27. Kukwikila, M.; Gale, N.; El-Sagheer, A. H.; Brown, T.; Tavassoli, A. *Nat. Chem.* **2017,** *9,* 1089–1098. doi:10.1038/nchem.2850
28. Ozaki, H.; Kitamura, M.; Yamana, K.; Murakami, A.; Shimidzu, T. *Bull. Chem. Soc. Jpn.* **1990,** *63,* 1929–1936. doi:10.1246/bcsj.63.1929
29. Moody, H. M.; van Genderen, M. H. P.; Koole, L. H.; Kocken, H. J. M.; Meijer, E. M.; Buck, H. M. *Nucleic Acids Res.* **1989,** *17,* 4769–4782. doi:10.1093/nar/17.12.4769
30. Strauss, J. K.; Roberts, C.; Nelson, M. G.; Switzer, C.; Maher, L. J., III. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93,* 9515–9520. doi:10.1073/pnas.93.18.9515
31. Ramzaeva, N.; Mittelbach, C.; Seela, F. *Helv. Chim. Acta* **1997,** *80,* 1809–1822. doi:10.1002/hlca.19970800605
32. Rosemeyer, H.; Ramzaeva, N.; Becker, E.-M.; Feiling, E.; Seela, F. *Bioconjugate Chem.* **2002,** *13,* 1274–1285. doi:10.1021/bc020024q
33. Prakash, T. P.; Kawasaki, A. M.; Lesnik, E. A.; Sioufi, N.; Manoharan, M. *Tetrahedron* **2003,** *59,* 7413–7422. doi:10.1016/S0040-4020(03)01104-9
34. Prakash, T. P.; Kawasaki, A. M.; Lesnik, E. A.; Owens, S. R.; Manoharan, M. *Org. Lett.* **2003,** *5,* 403–406. doi:10.1021/ol027131k
35. Prhavc, M.; Prakash, T. P.; Minasov, G.; Cook, P. D.; Egli, M.; Manoharan, M. *Org. Lett.* **2003,** *5,* 2017–2020. doi:10.1021/ol0340991
36. Prakash, T. P.; Püschl, A.; Lesnik, E.; Mohan, V.; Tereshko, V.; Egli, M.; Manoharan, M. *Org. Lett.* **2004,** *6,* 1971–1974. doi:10.1021/ol049470e
37. Milton, S.; Honcharenko, D.; Rocha, C. S. J.; Moreno, P. M. D.; Smith, C. I. E.; Strömberg, R. *Chem. Commun.* **2015,** *51,* 4044–4047. doi:10.1039/C4CC08837A
38. Brock, R. *Bioconjugate Chem.* **2014,** *25,* 863–868. doi:10.1021/bc500017t
39. Jain, M. L.; Bruice, P. Y.; Szabó, I. E.; Bruice, T. C. *Chem. Rev.* **2012,** *112,* 1284–1309. doi:10.1021/cr1004265
40. Letsinger, R. L.; Bach, S. A.; Eadie, J. S. *Nucleic Acids Res.* **1986,** *14,* 3487–3499. doi:10.1093/nar/14.8.3487
41. Letsinger, R. L.; Singman, C. N.; Histand, G.; Salunkhe, M. *J. Am. Chem. Soc.* **1988,** *110,* 4470–4471. doi:10.1021/ja00221a089
42. Froehler, B. C. *Tetrahedron Lett.* **1986,** *27,* 5575–5578. doi:10.1016/S0040-4039(00)85269-7
43. Bailey, C. P.; Weeks, D. L.; Dagle, J. M. *Nucleic Acids Res.* **1998,** *26,* 4860–4867. doi:10.1093/nar/26.21.4860
44. Michel, T.; Martinand-Mari, C.; Debart, F.; Lebleu, B.; Robbins, I.; Vasseur, J.-J. *Nucleic Acids Res.* **2003,** *31,* 5282–5290. doi:10.1093/nar/gkg733
45. Deglane, G.; Abes, S.; Michel, T.; Laurent, A.; Naval, M.; Martinand-Mari, C.; Prevot, P.; Vives, E.; Robbins, I.; Barvik, I., Jr.; Lebleu, B.; Debart, F.; Vasseur, J.-J. *Collect. Symp. Ser.* **2005,** *7,* 143–147. doi:10.1135/css200507143
46. Deglane, G.; Abes, S.; Michel, T.; Prévot, P.; Vives, E.; Debart, F.; Barvik, I.; Lebleu, B.; Vasseur, J.-J. *ChemBioChem* **2006,** *7,* 684–692. doi:10.1002/cbic.200500433
47. Fathi, R.; Huang, Q.; Coppola, G.; Delaney, W.; Teasdale, R.; Krieg, A. M.; Cook, A. F. *Nucleic Acids Res.* **1994,** *22,* 5416–5424. doi:10.1093/nar/22.24.5416
48. Rahman, S. M. A.; Baba, T.; Kodama, T.; Islam, M. A.; Obika, S. *Bioorg. Med. Chem.* **2012,** *20,* 4098–4102. doi:10.1016/j.bmc.2012.05.009
49. Dempcy, R. O.; Almarsson, O.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91,* 7864–7868. doi:10.1073/pnas.91.17.7864
50. Perreault, D. M.; Cabell, L. A.; Anslyn, E. V. *Bioorg. Med. Chem.* **1997,** *5,* 1209–1220. doi:10.1016/S0968-0896(97)00051-5
51. Blaskó, A.; Dempcy, R. O.; Minyat, E. E.; Bruice, T. C. *J. Am. Chem. Soc.* **1996,** *118,* 7892–7899. doi:10.1021/ja961308m
52. Linkletter, B. A.; Bruice, T. C. *Bioorg. Med. Chem. Lett.* **1998,** *8,* 1285–1290. doi:10.1016/S0960-894X(98)00228-5
53. Linkletter, B. A.; Szabo, I. E.; Bruice, T. C. *Nucleic Acids Res.* **2001,** *29,* 2370–2376. doi:10.1093/nar/29.11.2370
54. Barawkar, D. A.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95,* 11047–11052. doi:10.1073/pnas.95.19.11047
55. Challa, H.; Bruice, T. C. *Bioorg. Med. Chem. Lett.* **2001,** *11,* 2423–2427. doi:10.1016/S0960-894X(01)00455-3
56. Reddy, P. M.; Bruice, T. C. *Bioorg. Med. Chem. Lett.* **2003,** *13,* 1281–1285. doi:10.1016/S0960-894X(03)00119-7
57. Challa, H.; Bruice, T. C. *Bioorg. Med. Chem.* **2004,** *12,* 1475–1481. doi:10.1016/j.bmc.2003.12.043
58. Dempcy, R. O.; Luo, J.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93,* 4326–4330. doi:10.1073/pnas.93.9.4326
59. Kojima, N.; Szabo, I. E.; Bruice, T. C. *Tetrahedron* **2002,** *58,* 867–879. doi:10.1016/S0040-4020(01)01185-1
60. Dempcy, R. O.; Browne, K. A.; Bruice, T. C. *J. Am. Chem. Soc.* **1995,** *117,* 6140–6141. doi:10.1021/ja00127a035

61. Dempcy, R. O.; Browne, K. A.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92,* 6097–6101. doi:10.1073/pnas.92.13.6097
62. Browne, K. A.; Dempcy, R. O.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92,* 7051–7055. doi:10.1073/pnas.92.15.7051
63. Blaskó, A.; Minyat, E. E.; Dempcy, R. O.; Bruice, T. C. *Biochemistry* **1997,** *36,* 7821–7831. doi:10.1021/bi970064v
64. Linkletter, B. A.; Szabo, I. E.; Bruice, T. C. *J. Am. Chem. Soc.* **1999,** *121,* 3888–3896. doi:10.1021/ja984212w
65. Szabo, I. E.; Bruice, T. C. *Bioorg. Med. Chem.* **2004,** *12,* 4233–4244. doi:10.1016/j.bmc.2004.05.010
66. Jain, M. L.; Bruice, T. C. *Bioorg. Med. Chem.* **2006,** *14,* 7333–7346. doi:10.1016/j.bmc.2006.05.074
67. Arya, D. P.; Bruice, T. C. *J. Am. Chem. Soc.* **1998,** *120,* 6619–6620. doi:10.1021/ja980629q
68. Arya, D. P.; Bruice, T. C. *J. Am. Chem. Soc.* **1998,** *120,* 12419–12427. doi:10.1021/ja9829416
69. Arya, D. P.; Bruice, T. C. *Bioorg. Med. Chem. Lett.* **2000,** *10,* 691–693. doi:10.1016/S0960-894X(00)00085-8
70. Arya, D. P.; Bruice, T. C. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96,* 4384–4389. doi:10.1073/pnas.96.8.4384
71. Arya, D. P.; Bruice, T. C. *J. Am. Chem. Soc.* **1999,** *121,* 10680–10684. doi:10.1021/ja992871i
72. Schmidtgall, B.; Spork, A. P.; Wachowius, F.; Höbartner, C.; Ducho, C. *Chem. Commun.* **2014,** *50,* 13742–13745. doi:10.1039/C4CC06371F
73. Schmidtgall, B.; Höbartner, C.; Ducho, C. *Beilstein J. Org. Chem.* **2015,** *11,* 50–60. doi:10.3762/bjoc.11.8
74. Schmidtgall, B.; Kuepper, A.; Meng, M.; Grossmann, T. N.; Ducho, C. *Chem. – Eur. J.* **2018,** *24,* 1544–1553. doi:10.1002/chem.201704338
75. Kimura, K.-i.; Bugg, T. D. H. *Nat. Prod. Rep.* **2003,** *20,* 252–273. doi:10.1039/b202149h
76. Winn, M.; Goss, R. J. M.; Kimura, K.-i.; Bugg, T. D. H. *Nat. Prod. Rep.* **2010,** *27,* 279–304. doi:10.1039/B816215H
77. Ichikawa, S.; Yamaguchi, M.; Matsuda, A. *Curr. Med. Chem.* **2015,** *22,* 3951–3979. doi:10.2174/0929867322666150818103502
78. Chi, X.; Pahari, P.; Nonaka, K.; Van Lanen, S. G. *J. Am. Chem. Soc.* **2011,** *133,* 14452–14459. doi:10.1021/ja206304k
79. Barnard-Britson, S.; Chi, X.; Nonaka, K.; Spork, A. P.; Tibrewal, N.; Goswami, A.; Pahari, P.; Ducho, C.; Rohr, J.; Van Lanen, S. G. *J. Am. Chem. Soc.* **2012,** *134,* 18514–18517. doi:10.1021/ja308185q
80. Funabashi, M.; Baba, S.; Takatsu, T.; Kizuka, M.; Ohata, Y.; Tanaka, M.; Nonaka, K.; Spork, A. P.; Ducho, C.; Chen, W.-C. L.; Van Lanen, S. G. *Angew. Chem.* **2013,** *125,* 11821–11825. doi:10.1002/ange.201305546
Angew. Chem., Int. Ed. **2013,** *52,* 11607–11611. doi:10.1002/anie.201305546
81. McDonald, L. A.; Barbieri, L. R.; Carter, G. T.; Lenoy, E.; Lotvin, J.; Petersen, P. J.; Siegel, M. M.; Singh, G.; Williamson, R. T. *J. Am. Chem. Soc.* **2002,** *124,* 10260–10261. doi:10.1021/ja017748h
82. Wiegmann, D.; Koppermann, S.; Wirth, M.; Niro, G.; Leyerer, K.; Ducho, C. *Beilstein J. Org. Chem.* **2016,** *12,* 769–795. doi:10.3762/bjoc.12.77
83. Spork, A. P.; Koppermann, S.; Ducho, C. *Synlett* **2009,** 2503–2507. doi:10.1055/s-0029-1217742
84. Spork, A. P.; Koppermann, S.; Dittrich, B.; Herbst-Irmer, R.; Ducho, C. *Tetrahedron: Asymmetry* **2010,** *21,* 763–766. doi:10.1016/j.tetasy.2010.03.037
85. Spork, A. P.; Ducho, C. *Synlett* **2013,** *24,* 343–346. doi:10.1055/s-0032-1318117
86. Spork, A. P.; Ducho, C. *Org. Biomol. Chem.* **2010,** *8,* 2323–2326. doi:10.1039/c003092a
87. Spork, A. P.; Wiegmann, D.; Granitzka, M.; Stalke, D.; Ducho, C. *J. Org. Chem.* **2011,** *76,* 10083–10098. doi:10.1021/jo201935w
88. Spork, A. P.; Büschleb, M.; Ries, O.; Wiegmann, D.; Boettcher, S.; Mihalyi, A.; Bugg, T. D. H.; Ducho, C. *Chem. – Eur. J.* **2014,** *20,* 15292–15297. doi:10.1002/chem.201404775
89. Eisenhuth, R.; Richert, C. *J. Org. Chem.* **2008,** *74,* 26–37. doi:10.1021/jo8018889
90. Vineyard, B. D.; Knowles, W. S.; Sabacky, M. J.; Bachman, G. L.; Wienkauff, D. J. *J. Am. Chem. Soc.* **1977,** *99,* 5946–5952. doi:10.1021/ja00460a018
91. Schmidt, U.; Lieberknecht, A.; Schanbacher, U.; Beuttler, T.; Wild, J. *Angew. Chem.* **1982,** *94,* 797–798. doi:10.1002/ange.19820941027
Angew. Chem., Int. Ed. Engl. **1982,** *21, 776–777.* doi:10.1002/anie.198207761
92. Ducho, C.; Hamed, R. B.; Batchelar, E. T.; Sorensen, J. L.; Odell, B.; Schofield, C. J. *Org. Biomol. Chem.* **2009,** *7,* 2770–2779. doi:10.1039/b903312b

License and Terms