



# Using generative AI to transform peptide hits into small molecule leads

Joshua Mills<sup>\*1,2</sup> and Yu Heng Lau<sup>\*1,2</sup>

## Perspective

Open Access

### Address:

<sup>1</sup>School of Chemistry, The University of Sydney, Camperdown NSW 2006, Australia and <sup>2</sup>ARC Centre for Innovations in Peptide and Protein Science, The University of Sydney, Australia

### Email:

Joshua Mills<sup>\*</sup> - joshua.mills@sydney.edu.au; Yu Heng Lau<sup>\*</sup> - yuheng.lau@sydney.edu.au

\* Corresponding author

### Keywords:

diffusion models; drug discovery; generative AI; peptides; small molecules

*Beilstein J. Org. Chem.* **2026**, *22*, 672–679.

<https://doi.org/10.3762/bjoc.22.51>

Received: 13 February 2026

Accepted: 17 April 2026

Published: 30 April 2026

This article is part of the thematic issue "Design and synthesis of bioactive molecules".

Associate Editor: D. Spring



© 2026 Mills and Lau; licensee Beilstein-Institut.  
License and terms: see end of document.

## Abstract

There is a wealth of structural data on peptides that bind potential drug targets, serving as rich inspiration for designing small molecule inhibitors which can recapitulate key binding interactions while improving pharmacokinetic properties. With the rapid advancement of artificial intelligence capabilities including powerful generative models, there is significant potential for new AI-based tools to expedite the structure-based transformation of peptide hits into small molecule leads. In this Perspective, we highlight how AI-enabled prediction and design tools can potentially span the entire workflow from peptide to small molecule: target protein structure prediction, de novo peptide binder generation, diffusion models for generating novel small molecule scaffolds, and deep-learning predictors of binding affinity to rapidly triage candidates.

## Introduction

In drug discovery, peptides serve as accessible starting points for designing molecules that bind and inhibit a chosen protein target. Many enzyme targets (e.g., proteases) natively process peptides as their substrates. Peptide domains are also common motifs in protein–protein interactions, mediating the binding between protein partners that form functional complexes [1–3]. Furthermore, powerful high-throughput library screening technologies such as mRNA and phage display have revolutionised our ability to identify novel peptide binders, frequently gener-

ating binders with nanomolar affinity directly from screening campaigns [4–6].

While there has been increasing recognition of the translational potential of peptides in drug discovery, significant optimisation is typically required to convert a naïve peptide into a bona fide drug candidate [7]. Many promising peptide hits fail to translate beyond academic research, hampered by pharmacokinetic liabilities such as poor metabolic stability, limited membrane

permeability, and low oral bioavailability. In some cases, peptides may be suitable for therapeutic use after installing a small number of chemical modifications (e.g., lipidation of peptide hormones such as GLP-1 receptor agonists). In other cases, however, challenging medicinal chemistry is required to achieve clinical efficacy, as exemplified by the complex development of the heavily modified tricyclic peptide MK-0616, a PCSK9 inhibitor in phase III clinical trials that was originally derived from an mRNA display screen of monocyclic peptides [8].

As an alternative to optimising peptide-based molecules, knowledge of key binding interactions exhibited by potent peptide binders can be leveraged to inform small-molecule design. The field of peptidomimetics seeks to design small molecules that mimic the binding mode of a peptide, while mitigating the pharmacokinetic liabilities associated with unmodified peptides. This peptide-first approach to drug discovery is particularly valuable for challenging targets such as protein–protein interactions, where traditional small molecule high-throughput screening often fails to yield hits [9]. While the term peptidomimetics has been used to describe a variety of molecular classes, including modified peptides and molecular scaffolds for display of amino acid side-chains, this Perspective focuses on non-peptidic small molecules with the appropriate shape and functionality to recapitulate the pharmacophore of the original peptide (Class D according to the classification proposed by Grossmann and co-workers [10]).

The challenge of transforming a peptide binder into a small molecule mimic is not new [11]. Indeed, there are well-known historical examples of blockbuster drugs derived from native

peptide substrates. A classic example is the ACE inhibitor captopril, an analogue of a snake venom peptide, the development of which has been cited as an early success story for structure-based rational drug design [12,13]. Despite the long history, there is still no straightforward and generalisable workflow with a reasonable success rate for achieving the transformation from peptide to small molecule inhibitor.

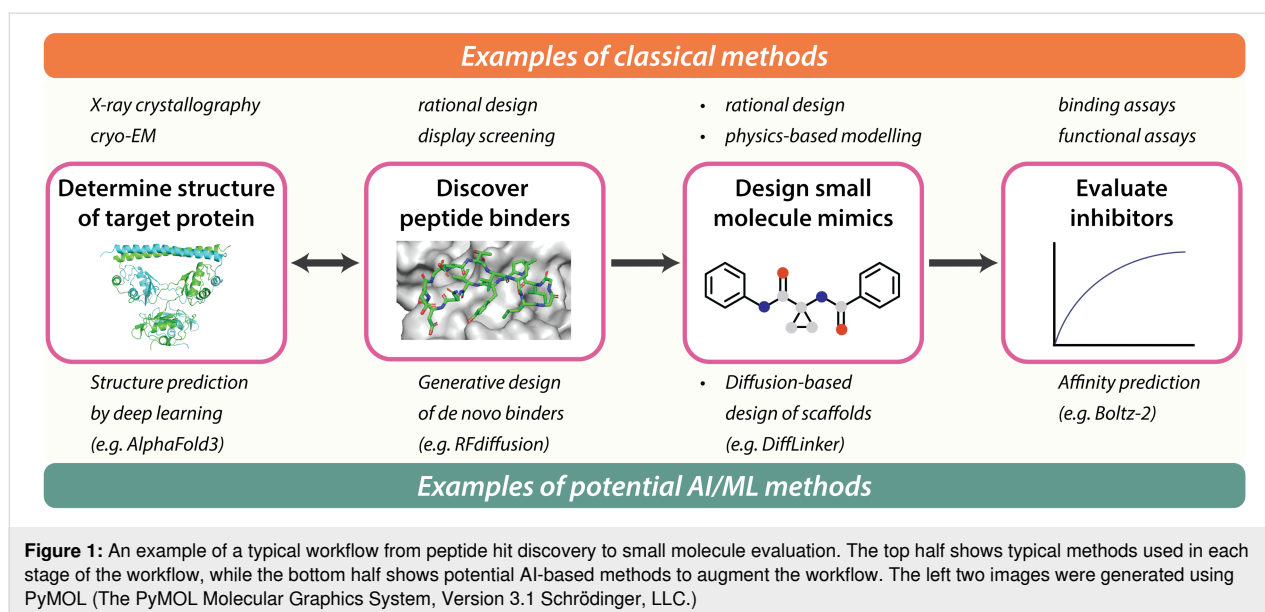
With the rapid maturation of computational tools based on artificial intelligence (AI), this Perspective highlights selected nascent examples where AI has been used for small molecule design by leveraging data on peptide binders, and proposes potential opportunities where generative AI and machine learning (ML) tools may augment various stages throughout the pipeline from peptide hit discovery to small molecule lead.

## Perspective

### Current workflows for designing small molecules that mimic peptide pharmacophores

#### Illustrative example of a non-AI workflow

To understand where generative AI may play a role in transforming peptides into small molecules, we first briefly outline how traditional non-AI tools are typically used in the field. Starting from an experimental co-crystal structure of the bound complex, standard medicinal chemistry principles are used to probe structure–activity relationships (SAR) and determine the key interactions that form a minimal pharmacophore, supported by classical physics-based molecular modelling methods such as molecular dynamics (MD) and docking (Figure 1).



As an illustrative example, Yoshida and co-workers at Shionogi reported the design of small molecule inhibitors of nicotinamide *N*-methyltransferase [14] and  $\beta$ -herpesvirus proteases [15], starting from cyclic peptide hits obtained by mRNA display with flexible in vitro translation (also known as RaPID, random non-standard peptides integrated discovery [4]). Key interactions that comprise the pharmacophores were experimentally determined by alanine scanning and other amino acid substitutions to explore SAR, coupled with analysis of co-crystal structures obtained for the bound complexes, including computational analysis of the amino acid interactions (SiteMap [16]) and hydration in the binding pocket (grid inhomogeneous solvation theory [17] in AmberTools [18]). The pharmacophores were then used as the basis for virtual screening with docking (Glide [19,20]) to obtain candidate small molecules for experimental validation in functional assays.

### Nascent examples of ML approaches

In contrast to physics-based models for extracting pharmacophores and predicting small molecule binders, ML methods are now emerging as alternatives for accomplishing these tasks.

An early application of ML for small molecule design based on peptide datasets was reported in 2023 by Hou and co-workers, where models were trained to classify ligands as ghrelin receptor binders [21]. The complete dataset for training and evaluation consisted of 548 peptides/peptidomimetics and 2193 small molecules spanning known binders, non-binders, and random compounds, each represented using one-hot encoding. A variety of well-established machine-learning classifiers were tested, including random forests, support vector machines, and gradient boosted decision trees. Strategies that used a larger proportion of the dataset for training showed reasonable predictive ability when evaluated on the remaining compounds as the validation set, as well as on an external set of 30 binders that was not in the original dataset. This result hints at the potential of ML models in predicting small molecule binders based on key peptide attributes, although the classifier was trained on more small molecules than peptides/peptidomimetics, making it unclear if peptide and peptidomimetic data alone would provide sufficient training to predict small molecule binders. The generality of this approach to other targets with no known small molecule binders has also yet to be evaluated.

Also in 2023, a deep learning-driven molecular generator for designing peptidomimetics was reported by Nakamura, Bajorath and co-workers [22]. Named DeepCubist, the method superimposes scaffolds from an  $sp^3$ -rich skeleton library onto the target peptide to identify spatially aligned structures. Once a scaffold is selected, a transformer model converts the three-dimensional scaffolds into complete chemical structures by

decorating the backbone with heteroatoms and unsaturated bonds. The resulting small molecules are designed to be synthetically tractable, although synthetic accessibility was noted to be somewhat restricted by the highly  $sp^3$ -rich nature of the generated molecules. New iterations of these methods have since been reported by the authors [23], though experimental validation of predictions remains to be reported at the time of writing this Perspective.

### Opportunities for using generative AI and ML tools

Here, we highlight how multiple stages of the typical peptide to small molecule drug discovery workflow may be amenable to AI-enabled structure-based tools (Table 1). Starting from a target protein structure (which may be modelled by AI), generative models can be used to design de novo peptide binders. Upon identifying the minimal peptide pharmacophore, fragment-linking diffusion-based models can be used to recapitulate key binding features on a novel molecular scaffold. Subsequently, machine learning models can assist in triaging generated candidates prior to more resource-intensive modelling, laboratory synthesis, and evaluation in functional assays.

### Deep learning for protein structure prediction

The growing wealth of structural data in the Protein Data Bank, along with ongoing improvements in computational processing power, has fuelled the success of protein structure prediction tools based on deep learning algorithms. Most notably, DeepMind's AlphaFold2 heralded a breakthrough in protein structure prediction with its exemplary performance in the 14th Critical Assessment of protein Structure Prediction (CASP14) experiment in 2020 [24]. Since then, numerous models with advanced functionality have been released (Table 1), including AlphaFold3 which incorporates diffusion-based components, providing a range of options for predicting the structure of protein targets (including protein complexes and interactions with other biomolecules) where no experimental structure has been reported. While the accuracy of these models continues to increase with time, protein structures modelled on experimental data (X-ray crystallography, cryo-electron microscopy) still remain the gold standard as starting points for structure-based drug design.

### Generative AI for peptide design

Building off the success of structure prediction tools, a major area of growth has been in new AI-based tools for designing de novo protein binders, many of which can be adapted to peptides. The release of RFDiffusion [42] and ProteinMPNN [35] from David Baker and co-workers at the Institute of Protein Design in 2022–23 popularised the use of diffusion models to build de novo protein backbones followed by sequence

**Table 1:** Selected examples of structure-based AI/ML tools for potential end-to-end coverage of the peptide to small molecule pipeline. Tools are listed in alphabetical order. Year corresponds to the final publication date if peer-reviewed, or preprint date otherwise.

Capability	Tools	Developer
predicting structure of target protein (and complexes)	AlphaFold3 (2024) [25] Boltz-1 (2024) [26] Chai-1 (2024) [27] ESMFold (2023) [28] RoseTTAFold (2021) [29]	Google DeepMind MIT Jameel Clinic Chai Discovery Meta AI Institute for Protein Design, UW
generative design of peptide binders (including cyclic peptides)	AfCycDesign (2025) [30] BindCraft (2025) [31] BoltzGen (2025) [32] EvoBind2 (2025) [33] PepMimic (2025) [34] ProteinMPNN (2022) [35] RFpeptides (2025) [36]	Institute for Protein Design, UW Correia Lab, EPFL MIT Bryant Lab, Stockholm Ma Lab, Tsinghua University Institute for Protein Design, UW Institute for Protein Design, UW
transforming peptides hits into small molecules	DiffLinker (2024) [37] Peptide2Mol (2025) [38] ShEPhERD (2024) [39]	Correia Lab, EPFL Ma Lab, Tsinghua University Coley Research Group, MIT
predicting affinities	Boltz-2 (2025) [40] RosettaVS (2024) [41]	MIT and Recursion Institute for Protein Design, UW

design. As with structure prediction, there has since been an explosion of new methods improving upon these initial methods (Table 1), including RFDiffusion2 [43] and RFDiffusion3 [44], along with integrated pipelines such as BindCraft [31] which are built upon tools such as ProteinMPNN and AlphaFold2. Other design tools that are specific for peptide binders also include the evolution-based EvoBind2 that only requires the protein target sequence as an input [33], and the diffusion-based PepMimic that uses structural information from binding interfaces [34]. Ongoing evaluation by independent groups suggests that reliable one-shot prediction of binders is currently target-dependent [45], and continued improvements may see these methods reach parity with traditional peptide discovery methods in the future.

### Diffusion models to generate small molecule mimics

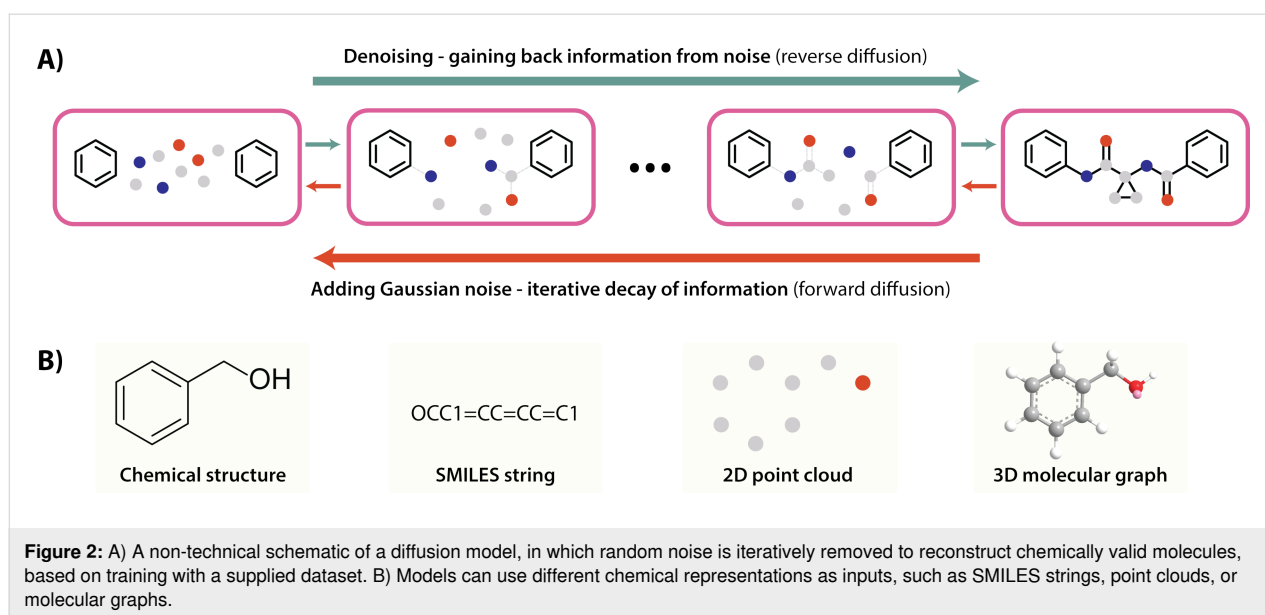
The reliable design of synthesisable small molecules that can accurately mimic a peptide pharmacophore remains the core challenge in the overall peptide to small molecule workflow. There are many AI tools for general ligand-based drug design under active development, spanning a range of generative architectures (e.g., variational autoencoders, generative adversarial networks). Here, we specifically highlight the potential of using diffusion-based models for generating small molecules that recapitulate key components of a peptide pharmacophore.

In general, diffusion models function by progressively adding Gaussian noise to training data over a series of timesteps (forward diffusion). A neural network is then trained to reverse this process, iteratively reconstructing the original data structure through predicting and removing noise (reverse diffusion). Once trained, the model initialises and iteratively denoises

random Gaussian noise to generate novel outputs (Figure 2a). Diffusion models have been particularly adept at generating new visual outputs such as images and videos, hence their suitability for de novo molecular design is not unexpected. In the case of small molecule design, the chemical inputs can be represented in different formats depending on the level of information and complexity required [46]. One-dimensional SMILES strings are the simplest form of molecular representation that lacks spatial information, while two- and three-dimensional molecular graph and point cloud representations are less compact but carry additional spatial information that is likely to be critical for small molecule binder design (Figure 2b).

In the context of transforming peptides into small molecules, diffusion models can be applied using a workflow reminiscent of fragment-based drug design. Starting from the key elements of a peptide pharmacophore (e.g., critical side-chains or backbone functional groups that interact with the target), diffusion models can generate candidate small molecules that spatially link or merge all these elements together. To preserve three-dimensional spatial information, the following examples of diffusion models all use three-dimensional atomic point clouds and an E(3)-equivariant (or SE(3)-equivariant) graph neural network for symmetry-consistent modelling of molecular structures [47].

DiffLinker is an example of a conditional diffusion model developed by Correia and co-workers that enables linking of multiple fragments in a predefined orientation [37]. After pretraining on linkers extracted from molecules originating from several compound databases (ZINC [48], CASF [49], GEOM [50]), input fragments were represented as three-dimen-



sional point clouds, while the appropriately-sized linker was iteratively generated and refined by sampling atom types and positions using Gaussian noise as the starting point. The model could also be trained on a protein–ligand dataset (from Binding MOAD [51]) to factor in the constraints of typical binding pockets.

ShEPHERD is an interaction-aware diffusion model developed by Coley and co-workers that is capable of bioisosteric fragment merging [39]. For fragment merging, the model was trained on a subset from the MOSES database [52]. Then starting with a set of published fragment data from an experimental screen, ShEPHERD generated small molecule candidates with high similarity in electrostatic and pharmacophore scoring metrics.

Meanwhile, Peptide2Mol is a diffusion model developed by Ma and co-workers that leverages peptide-binder structural data to generate small molecules within the protein pocket. It is trained on a range of target-bound complexes, including small molecule ligands as well as peptides and protein binders [38]. Similar to DiffLinker, small molecule generation is initiated from Gaussian noise, then after iterative denoising within the target pocket, an additional pocket-aware refinement step is used to remove steric clashes and shape complementary issues.

Development of this class of generative molecular design tools is still in its infancy. Outputs typically require manual curation or integration with traditional physics-based molecular modelling to filter out unstable functional groups, strained rings and other unfavourable conformations. Experimental validation of predicted small molecules also remains to be conducted. Never-

theless, continued improvements in model design and quality input data may see these tools mature in future, potentially mirroring the trajectory of the aforementioned structure prediction and binder design tools.

### Deep learning for predicting binding affinities

As a brief note, AI tools are also available for preliminary evaluation of small molecule candidates, providing potential alternatives to traditional computationally expensive physics-based workflows such as docking, MD, and free-energy perturbation (FEP) calculations for affinity estimation. For example, Boltz-2 features a strong emphasis on affinity prediction in addition to its core capability of structure prediction [40]. Other tools integrate AI methods into virtual screening pipelines, such as RosettaVS which trains a target-specific neural network during physics-based docking to prioritise candidates for more expensive computational modelling [41].

### Outlook

There is still significant progress that must be made before a complete AI-assisted computational workflow for transforming peptide hits into small molecule leads can be reliably achieved. While the early stages of structure prediction and peptide binder generation are rapidly maturing, small molecule design and evaluation methods are far less advanced, likely reflecting the more diverse and complex nature of small molecule chemical space, and the relatively sparse coverage of this space in training datasets. Nevertheless, early indications suggest that there could be a significant future role for generative AI in peptide to small molecule drug discovery workflows, even though ultimately, all predictions require validation in an experimental laboratory setting.

An open question is whether a peptide-first approach to drug discovery will continue to hold advantages in the future. Indeed, many target-based computational drug discovery tools that are ligand-agnostic are under development, along with small molecule prediction methods that do not require experimental data on target-specific peptide binders. Nevertheless, we envisage that given the ease of binder design and screening, synthetic accessibility, and in vitro assay validation, peptides will continue to play a significant role in supporting the small molecule drug discovery pipeline.

Just as the Critical Assessment of Structure Prediction (CASP) competition supported the development of protein structure prediction methods, the emergence of small molecule hit prediction competitions, such as the CACHE (Critical Assessment of Computational Hit-finding Experiments) challenge [53], are driving a pivotal shift toward rigorous, blind, and prospectively validated benchmarking. As methods improve, we expect such competitions to raise the bar and focus on increasingly challenging targets, such as protein–protein interactions without any known small molecule binders. Additionally, a focus on 'make-on-demand' constraints would allow generative models to move beyond the commercially available chemical space and into the realm of truly novel, de novo chemical space.

One potential barrier to rapid development and uptake of AI in drug discovery is the restricted accessibility of some models and datasets. While the development of proprietary AI-based platforms in private organisations such as large pharmaceutical companies and small start-up enterprises is driving local innovation within their respective organisations, such platforms are rarely shared for public use and may only be accessible through licensing at substantial cost. At the same time, one could argue that these platforms may have never been developed without the corporate structure that private enterprise offers. Although such debates are not new to drug discovery, we anticipate that the balance between proprietary and open-source development will have a disproportionate influence on the future of AI-based drug discovery, due to the data-intensive nature of the discipline.

Finally, if the full breadth of potential applications of AI in drug discovery come to fruition, one could imagine a fully-automated pipeline from target identification and virtual screening through to retrosynthetic planning, and robotics for hit-to-lead optimisation and functional evaluation in cellular models. The examples highlighted in this Perspective represent only a few of the many ongoing developments in the broader field. With the field moving at such a rapid pace, it is difficult to forecast which AI tools will become

the next field standard akin to AlphaFold2 for structure prediction, and which others will fall short of more traditional computational or experimental methods. Only time will tell whether AI optimism in drug discovery is well-founded or not.

## Funding

JM acknowledges funding from an Australian Government Research Training Program scholarship, and a Centre for Drug Discovery Innovation and Sydney Precision Data Science Centre supplementary scholarship. YHL acknowledges funding from the National Health and Medical Research Foundation Investigator grant (2041692), Synergy grant (2045600), and funding from the Cancer Institute of New South Wales Career Development Fellowship (2024/CDF1222).

## ORCID® iDs

Joshua Mills - <https://orcid.org/0009-0009-3790-9772>

Yu Heng Lau - <https://orcid.org/0000-0001-6560-8410>

## Data Availability Statement

Data sharing is not applicable as no new data was generated or analyzed in this study.

## References

- Bullock, B. N.; Jochim, A. L.; Arora, P. S. *J. Am. Chem. Soc.* **2011**, *133*, 14220–14223. doi:10.1021/ja206074j
- Watkins, A. M.; Arora, P. S. *ACS Chem. Biol.* **2014**, *9*, 1747–1754. doi:10.1021/cb500241y
- Gavenonis, J.; Sheneman, B. A.; Siegert, T. R.; Eshelman, M. R.; Kritzer, J. A. *Nat. Chem. Biol.* **2014**, *10*, 716–722. doi:10.1038/nchembio.1580
- Passioura, T.; Suga, H. *Chem. Commun.* **2017**, *53*, 1931–1940. doi:10.1039/c6cc06951g
- Huang, Y.; Wiedmann, M. M.; Suga, H. *Chem. Rev.* **2019**, *119*, 10360–10391. doi:10.1021/acs.chemrev.8b00430
- Deyle, K.; Kong, X.-D.; Heinis, C. *Acc. Chem. Res.* **2017**, *50*, 1866–1874. doi:10.1021/acs.accounts.7b00184
- Gare, C. L.; White, A. M.; Malins, L. R. *Trends Biochem. Sci.* **2025**, *50*, 467–480. doi:10.1016/j.tibs.2025.01.009
- Li, H.; Thaisrivongs, D. A.; Shang, G.; Chen, Y.; Chen, Q.; Tan, L.; Xiao, K.-J.; Larson, R. T.; Kuethe, J. T.; Lee, J.; Deprez, N. R.; Nolting, A. F.; Poirier, M.; Bulger, P. G.; Regalado, E. L.; Biba, M.; Tsay, F.-R.; DaSilva, J.; Prier, C. K.; Strulson, C. A.; Zawatzky, K.; Liu, Z.; Newman, J. A.; Sokolowsky, K.; Tang, W.; Hullen, K.; Thakur, N.; Welch, C.; Patel, S.; He, Y.; Xu, J.; Variankaval, N.; Klapars, A.; Kong, J.; Desmond, R.; Varsolona, R.; Maligres, P. E.; Pons Siepermann, C. A.; Robison, L.; Piou, T.; Hartmanshenn, C.; Chandra, A.; Patel, A.; Becker, M. R.; Liu, G.; Duan, J.; Wan, B.; Xiao, C.; Yuan, Y.; Cao, X.; Chen, L.; Yi, R.; Wu, Z.; Feng, M.; Li, D.; Song, Z.; Dong, Y.; Sun, J.; Li, B.; Shao, G.; Campeau, L.-C.; Yin, J. *J. Am. Chem. Soc.* **2025**, *147*, 11036–11048. doi:10.1021/jacs.4c15966
- Wells, J. A.; McClendon, C. L. *Nature* **2007**, *450*, 1001–1009. doi:10.1038/nature06526

10. Pelay-Gimeno, M.; Glas, A.; Koch, O.; Grossmann, T. N. *Angew. Chem., Int. Ed.* **2015**, *54*, 8896–8927. doi:10.1002/anie.201412070
11. Gante, J. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 1699–1720. doi:10.1002/anie.199416991
12. Nemeč, K.; Schubert-Zsilavec, M. *Pharm. Unserer Zeit* **2003**, *32*, 11–16. doi:10.1002/pauz.200390001
13. Cushman, D. W.; Ondetti, M. A. *Hypertension* **1991**, *17*, 589–592. doi:10.1161/01.hyp.17.4.589
14. Yoshida, S.; Uehara, S.; Kondo, N.; Takahashi, Y.; Yamamoto, S.; Kameda, A.; Kawagoe, S.; Inoue, N.; Yamada, M.; Yoshimura, N.; Tachibana, Y. *J. Med. Chem.* **2022**, *65*, 10655–10673. doi:10.1021/acs.jmedchem.2c00919
15. Yoshida, S.; Sako, Y.; Nikaide, E.; Ueda, T.; Kozono, I.; Ichihashi, Y.; Nakahashi, A.; Onishi, M.; Yamatsu, Y.; Kato, T.; Nishikawa, J.; Tachibana, Y. *ACS Med. Chem. Lett.* **2023**, *14*, 1558–1566. doi:10.1021/acsmedchemlett.3c00359
16. Halgren, T. A. *J. Chem. Inf. Model.* **2009**, *49*, 377–389. doi:10.1021/ci800324m
17. Nguyen, C. N.; Kurtzman Young, T.; Gilson, M. K. *J. Chem. Phys.* **2012**, *137*, 044101. doi:10.1063/1.4733951
18. Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzes, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O’Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Risheh, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E., III; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M., Jr. *J. Chem. Inf. Model.* **2023**, *63*, 6183–6191. doi:10.1021/acs.jcim.3c01153
19. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750–1759. doi:10.1021/jm030644s
20. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739–1749. doi:10.1021/jm0306430
21. Liu, W.; Hopkins, A. M.; Yan, P.; Du, S.; Luyt, L. G.; Li, Y.; Hou, J. *Mol. Diversity* **2023**, *27*, 2239–2255. doi:10.1007/s11030-022-10555-w
22. Umedera, K.; Yoshimori, A.; Chen, H.; Kouji, H.; Nakamura, H.; Bajorath, J. *J. Comput.-Aided Mol. Des.* **2023**, *37*, 107–115. doi:10.1007/s10822-022-00493-y
23. Yoshimori, A.; Bajorath, J. *Eur. J. Med. Chem. Rep.* **2025**, *13*, 100249. doi:10.1016/j.ejmcr.2025.100249
24. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature* **2021**, *596*, 583–589. doi:10.1038/s41586-021-03819-2
25. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O’Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Židek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. *Nature* **2024**, *630*, 493–500. doi:10.1038/s41586-024-07487-w
26. Wohlwend, J.; Corso, G.; Passaro, S.; Getz, N.; Reveiz, M.; Leidal, K.; Swiderski, W.; Atkinson, L.; Portnoi, T.; Chinn, I.; Silterra, J.; Jaakkola, T.; Barzilay, R. *bioRxiv* **2025**. doi:10.1101/2024.11.19.624167
27. Discovery, C.; Boitreaud, J.; Dent, J.; McPartlon, M.; Meier, J.; Reis, V.; Rogozhnikov, A.; Wu, K. *bioRxiv* **2024**. doi:10.1101/2024.10.10.615955
28. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. *Science* **2023**, *379*, 1123–1130. doi:10.1126/science.ade2574
29. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. *Science* **2021**, *373*, 871–876. doi:10.1126/science.abj8754
30. Rettie, S. A.; Campbell, K. V.; Bera, A. K.; Kang, A.; Kozlov, S.; Bueso, Y. F.; De La Cruz, J.; Ahlrichs, M.; Cheng, S.; Gerben, S. R.; Lamb, M.; Murray, A.; Adebomi, V.; Zhou, G.; DiMaio, F.; Ovchinnikov, S.; Bhardwaj, G. *Nat. Commun.* **2025**, *16*, 4730. doi:10.1038/s41467-025-59940-7
31. Pacesa, M.; Nickel, L.; Schellhaas, C.; Schmidt, J.; Pyatova, E.; Kissling, L.; Barendse, P.; Choudhury, J.; Kapoor, S.; Alcaraz-Serna, A.; Cho, Y.; Ghamary, K. H.; Vinué, L.; Yachnin, B. J.; Wollacott, A. M.; Buckley, S.; Westphal, A. H.; Lindhoud, S.; Georgeon, S.; Goverde, C. A.; Hatzopoulos, G. N.; Gönczy, P.; Muller, Y. D.; Schwank, G.; Swarts, D. C.; Vecchio, A. J.; Schneider, B. L.; Ovchinnikov, S.; Correia, B. E. *Nature* **2025**, *646*, 483–492. doi:10.1038/s41586-025-09429-6
32. Stark, H.; Faltings, F.; Choi, M.; Xie, Y.; Hur, E.; O’Donnell, T.; Bushuiev, A.; Uçar, T.; Passaro, S.; Mao, W.; Reveiz, M.; Bushuiev, R.; Pluskal, T.; Sivic, J.; Kreis, K.; Vahdat, A.; Ray, S.; Goldstein, J. T.; Savinov, A.; Hambalek, J. A.; Gupta, A.; Taquiri-Diaz, D. A.; Zhang, Y.; Hatstat, A. K.; Arada, A.; Kim, N. H.; Tackie-Yarboi, E.; Boselli, D.; Schneider, L.; Liu, C. C.; Li, G.-W.; Hnisz, D.; Sabatini, D. M.; DeGrado, W. F.; Wohlwend, J.; Corso, G.; Barzilay, R.; Jaakkola, T. *bioRxiv* **2025**. doi:10.1101/2025.11.20.689494
33. Li, Q.; Vlachos, E. N.; Bryant, P. *Commun. Chem.* **2025**, *8*, 211. doi:10.1038/s42004-025-01601-3
34. Kong, X.; Jiao, R.; Lin, H.; Guo, R.; Huang, W.; Ma, W.-Y.; Wang, Z.; Liu, Y.; Ma, J. *Nat. Biomed. Eng.* **2025**, 1507. doi:10.1038/s41551-025-01507-4

35. Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; de Haas, R. J.; Bethel, N.; Leung, P. J. Y.; Huddy, T. F.; Pellock, S.; Tischer, D.; Chan, F.; Koepnick, B.; Nguyen, H.; Kang, A.; Sankaran, B.; Bera, A. K.; King, N. P.; Baker, D. *Science* **2022**, *378*, 49–56. doi:10.1126/science.add2187
36. Rettie, S. A.; Juergens, D.; Adebomi, V.; Bueso, Y. F.; Zhao, Q.; Leveille, A. N.; Liu, A.; Bera, A. K.; Wilms, J. A.; Üffing, A.; Kang, A.; Brackenbrough, E.; Lamb, M.; Gerben, S. R.; Murray, A.; Levine, P. M.; Schneider, M.; Vasireddy, V.; Ovchinnikov, S.; Weiergräber, O. H.; Willbold, D.; Kritzer, J. A.; Mougous, J. D.; Baker, D.; DiMaio, F.; Bhardwaj, G. *Nat. Chem. Biol.* **2025**, *21*, 1948–1956. doi:10.1038/s41589-025-01929-w
37. Igashov, I.; Stärk, H.; Vignac, C.; Schneuing, A.; Satorras, V. G.; Frossard, P.; Welling, M.; Bronstein, M.; Correia, B. *Nat. Mach. Intell.* **2024**, *6*, 417–427. doi:10.1038/s42256-024-00815-9
38. He, X.; Zhang, Y.; Lin, H.; Peng, X.; Kong, X.; Li, M.; Ma, J. *arXiv* **2025**, 2511.04984. doi:10.48550/arxiv.2511.04984
39. Adams, K.; Abeywardane, K.; Fromer, J.; Coley, C. W. *arXiv* **2025**, 2411.04130. doi:10.48550/arxiv.2411.04130
40. Passaro, S.; Corso, G.; Wohwend, J.; Reveiz, M.; Thaler, S.; Somnath, V. R.; Getz, N.; Portnoi, T.; Roy, J.; Stark, H.; Kwabi-Addo, D.; Beaini, D.; Jaakkola, T.; Barzilay, R. *bioRxiv* **2025**. doi:10.1101/2025.06.14.659707
41. Zhou, G.; Rusnac, D.-V.; Park, H.; Canzani, D.; Nguyen, H. M.; Stewart, L.; Bush, M. F.; Nguyen, P. T.; Wulff, H.; Yarov-Yarovoy, V.; Zheng, N.; DiMaio, F. *Nat. Commun.* **2024**, *15*, 7761. doi:10.1038/s41467-024-52061-7
42. Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; De Bortoli, V.; Mathieu, E.; Ovchinnikov, S.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek, M.; Baker, D. *Nature* **2023**, *620*, 1089–1100. doi:10.1038/s41586-023-06415-8
43. Ahern, W.; Yim, J.; Tischer, D.; Salike, S.; Woodbury, S. M.; Kim, D.; Kalvet, I.; Kipnis, Y.; Coventry, B.; Altae-Tran, H. R.; Bauer, M. S.; Barzilay, R.; Jaakkola, T. S.; Krishna, R.; Baker, D. *Nat. Methods* **2026**, *23*, 96–105. doi:10.1038/s41592-025-02975-x
44. Butcher, J.; Krishna, R.; Mitra, R.; Brent, R. I.; Li, Y.; Corley, N.; Kim, P. T.; Funk, J.; Mathis, S.; Salike, S.; Muraishi, A.; Eisenach, H.; Thompson, T. R.; Chen, J.; Politanska, Y.; Sehgal, E.; Coventry, B.; Zhang, O.; Qiang, B.; Didi, K.; Kazman, M.; DiMaio, F.; Baker, D. *bioRxiv* **2025**. doi:10.1101/2025.09.18.676967
45. Filius, M.; Patsos, T.; Minnee, H.; Turco, G.; Chong, H. E.; Liu, J.; Gnatzy, M.; Rooth, R. S. M.; Liu, A. C. H.; Ta, R. D. T.; Rijk, I. H. A.; Ziani, S.; Boxman, F. J.; Pomplun, S. J. *ACS Chem. Biol.* **2025**, *20*, 2991–2998. doi:10.1021/acscchembio.5c00774
46. Alakhdar, A.; Poczos, B.; Washburn, N. *J. Chem. Inf. Model.* **2024**, *64*, 7238–7256. doi:10.1021/acs.jcim.4c01107
47. Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. *Nat. Commun.* **2022**, *13*, 2453. doi:10.1038/s41467-022-29939-5
48. Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, *45*, 177–182. doi:10.1021/ci049714+
49. Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. *J. Chem. Inf. Model.* **2019**, *59*, 895–913. doi:10.1021/acs.jcim.8b00545
50. Axelrod, S.; Gómez-Bombarelli, R. *Sci. Data* **2022**, *9*, 185. doi:10.1038/s41597-022-01288-4
51. Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 333–340. doi:10.1002/prot.20512
52. Polykovskiy, D.; Zhebrak, A.; Sanchez-Lengeling, B.; Golovanov, S.; Tatanov, O.; Belyaev, S.; Kurbanov, R.; Artamonov, A.; Aladinskiy, V.; Veselov, M.; Kadurin, A.; Johansson, S.; Chen, H.; Nikolenko, S.; Aspuru-Guzik, A.; Zhavoronkov, A. *Front. Pharmacol.* **2020**, *11*, 565644. doi:10.3389/fphar.2020.565644
53. Ackloo, S.; Al-awar, R.; Amaro, R. E.; Arrowsmith, C. H.; Azevedo, H.; Batey, R. A.; Bengio, Y.; Betz, U. A. K.; Bologna, C. G.; Chodera, J. D.; Cornell, W. D.; Dunham, I.; Ecker, G. F.; Edfeldt, K.; Edwards, A. M.; Gilson, M. K.; Gordijo, C. R.; Hessler, G.; Hillisch, A.; Hogner, A.; Irwin, J. J.; Jansen, J. M.; Kuhn, D.; Leach, A. R.; Lee, A. A.; Lessel, U.; Morgan, M. R.; Moul, J.; Muegge, I.; Oprea, T. I.; Perry, B. G.; Riley, P.; Rousseaux, S. A. L.; Saikatendu, K. S.; Santhakumar, V.; Schapira, M.; Scholten, C.; Todd, M. H.; Vedadi, M.; Volkamer, A.; Willson, T. M. *Nat. Rev. Chem.* **2022**, *6*, 287–295. doi:10.1038/s41570-022-00363-z

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (<https://www.beilstein-journals.org/bjoc/terms>), which is identical to the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0>). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:  
<https://doi.org/10.3762/bjoc.22.51>