



This open access document is posted as a preprint in the Beilstein Archives at <https://doi.org/10.3762/bxiv.2021.11.v1> and is considered to be an early communication for feedback before peer review. Before citing this document, please check if a final, peer-reviewed version has been published.

This document is not formatted, has not undergone copyediting or typesetting, and may contain errors, unsubstantiated scientific claims or preliminary data.

Preprint Title Computational model predicts protein binding sites of a luminescent ligand equipped with guanidiniocarbonyl-pyrrole groups

Authors Neda Rafieiolhosseini, Matthias Killa, Niklas Tötsch, Jean-Noël Grad, Alexander Höing, Christian Ottmann, Shirley K. Knauer, Jens Voskuhl and Daniel Hoffmann

Publication Date 19 Feb. 2021

Article Type Full Research Paper

Supporting Information File 1 SI.docx; 2.3 MB

Supporting Information File 2 Table-S3.xlsx; 9.9 KB

Supporting Information File 3 Table-S4.xlsx; 10.2 KB

ORCID® IDs Neda Rafieiolhosseini - <https://orcid.org/0000-0002-2347-7747>; Niklas Tötsch - <https://orcid.org/0000-0001-7656-1160>; Jean-Noël Grad - <https://orcid.org/0000-0002-5821-4912>; Alexander Höing - <https://orcid.org/0000-0002-3762-4832>; Shirley K. Knauer - <https://orcid.org/0000-0003-4321-0924>; Daniel Hoffmann - <https://orcid.org/0000-0003-2973-7869>

License and Terms: This document is copyright 2021 the Author(s); licensee Beilstein-Institut.

This is an open access work under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>). Please note that the reuse, redistribution and reproduction in particular requires that the author(s) and source are credited and that individual graphics may be subject to special legal provisions.

The license is subject to the Beilstein Archives terms and conditions: <https://www.beilstein-archives.org/xiv/terms>.

The definitive version of this work can be found at <https://doi.org/10.3762/bxiv.2021.11.v1>

Computational model predicts protein binding sites of a luminescent ligand equipped with guanidiniocarbonyl-pyrrole groups

Neda Rafieiolhosseini¹, Matthias Killa², Niklas Tötsch¹, Jean-Noël Grad³,
Alexander Höing⁴, Christian Ottmann⁵, Shirley K. Knauer⁴, Jens Voskuhl², Daniel
Hoffmann^{1*}

¹ Bioinformatics and Computational Biophysics, Center for Medical Biotechnology (ZMB),
University of Duisburg-Essen, Universitätsstraße 5, 45141 Essen, Germany

² Faculty of Chemistry (Organic Chemistry) and CENIDE, University of Duisburg-Essen,
Universitätsstraße 7, 45141 Essen, Germany

³ Institute for Computational Physics, University of Stuttgart, Allmandring 3, 70569 Stuttgart,
Germany

⁴ Department of Molecular Biology II, Center for Medical Biotechnology (ZMB), University of
Duisburg-Essen, Universitätsstraße 5, 45141 Essen, Germany

⁵ Laboratory of Chemical Biology, Department of Biomedical Engineering and Institute for
Complex Molecular Systems, Eindhoven University of Technology, PO Box 513, 5600 MB
Eindhoven, Netherlands

Daniel Hoffmann - daniel.hoffmann@uni-due.de

* = Corresponding author

Abstract

The 14-3-3 protein family, one of the first discovered phosphoserine/phosphothreonine binding proteins, has attracted interest not only because of its important role in the cell regulatory processes but also due to its enormous number of interactions with other proteins. Here, we use a computational approach to find the binding sites of the designed hybrid compound featuring aggregation-induced emission luminophores as a potential supramolecular ligand for 14-3-3 ζ in the presence and absence of C-Raf peptides. Our results suggest that the area above and below the central pore of the dimeric 14-3-3 ζ protein is the most probable binding site for the ligand. Moreover, we predict that the position of the ligand is sensitive to the presence of phosphorylated C-Raf peptides.

Keywords

AIE luminophores; Fluorescence emission; Guanidiniocarbonyl-pyrrole; Ligand binding; 14-3-3 protein

1. Introduction

The 14-3-3 protein family was one of the first discovered phosphoserine/phosphothreonine binding proteins. In total seven isoforms of the 14-3-3 family are known to date (β , ϵ , γ , τ , θ , σ , and ζ) in mammals [1]. They are forming homo and heterodimers with a profile shaped like the Greek letter “ ω ” [2]. 14-3-3 proteins have attracted interest due to their enormous number of interactions with other proteins. Currently, more than 200 interacting proteins are known. 14-3-3 proteins play a significant role in cell signaling [2,3] and they were found to be essential in processes such as differentiation, apoptosis, or migration [4]. 14-3-3 proteins play also a role in several human diseases including cancer and neurodegenerative disorders like Alzheimer’s and Parkinson’s

diseases [5]. Based on these observations and findings it is not surprising, that the 14-3-3 family has attracted interest in pharmacological research as a novel potential target [5,6].

One option to modulate, inhibit, or stabilize protein-protein interactions (PPI) is the use of specific supramolecular ligands [7,8]. One well-known example of efficient protein binders is the so-called guanidiniocarbonyl-pyrrole (GCP) discovered 20 years ago by *Schmuck et al.*. These compounds are known to efficiently bind to oxo-anions such as carboxylates [9]. These compounds were already used to specifically address carboxylates on the surface of proteins. Many artificial receptors based on guanidinium scaffolds use hydrogen bonding, charge pairing, and hydrophobic interactions to complex oxoanions [10]. The guanidiniocarbonyl-pyrrole (GCP) is able to bind oxoanions even in aqueous solvents with competing ions and salts. *Schmuck et al.* also discovered that an additional positive charge increases the binding affinity to oxo-anions [9]. These unique properties make the GCP oxo-anion binder an ideal candidate to be used for protein recognition. In a previous study, it was possible to use GCP containing polycationic ligands for 14-3-3 proteins with a significant effect on PPIs [11,12]. Furthermore, a simple GCP derivative, namely GCP-Lys-OMe, was identified as the first binder for the specific binding area of the 14-3-3 ζ homodimer [13]. Very recently the survivine-histone H3 interaction was disrupted using a GCP dimer, which led to decreased cancer cell proliferation [8].

A major problem in this context is the readout of binding events, which is currently mainly achieved by indirect measurements. One approach to overcome this issue is to use fluorescence emission as a read-out tool, such as an emission “on” or “off” behavior [14]. Selective and sensitive fluorescent ligands have been proven to be essential tools for the study of biological systems by biosensing and imaging [15]. There is an increasing demand for novel luminophores tailor-made for different applications. One unique, promising class of compounds, the so-called aggregation-

induced emitters (AIE) have been used for a wide range of applications, like in OLEDs, liquid crystals, stimuli responses, bioassays, protein and ion detection or imaging [14,16]. In contrast to classical luminophores, these compounds typically show emission “on” behavior upon aggregation or binding which can be explained by a restriction of motion. In this contribution, we designed a hybrid compound featuring AIE luminophores based on aromatic-thioethers [17] as a potential supramolecular ligand for 14-3-3 ζ .

We synthesized a GCP-Lys dimer coupled via Cu(I) catalyzed click reaction to the chosen emitter equipped with two azide functions (see Figure 1 and Supporting Information File 1). This compound (**1**) was tested in initial binding assays using fluorescence emission as well as native gel electrophoresis. We could indeed show that (**1**) binds to 14-3-3 ζ as detected by native gel electrophoresis and fluorescence titration (see Supporting Information File 1). Initial experiments show that probably two molecules of (**1**) can bind simultaneously to one 14-3-3 ζ homodimer (see Supporting Information File 1). However, it is completely unknown where (**1**) binds on 14-3-3 ζ 's surface. In its profile, 14-3-3 ζ has roughly the shape of the Greek letter ω , where the two wells can clamp together pairs of proteins (Figure 2b). Knowledge of the binding site is of critical importance for a potential bioanalytics application: if (**1**) would block one or both wells of the ω , it could prevent binding of ligand proteins, whereas a (**1**) binding site in the outer parts of the ω would not interfere with functional protein-protein binding.

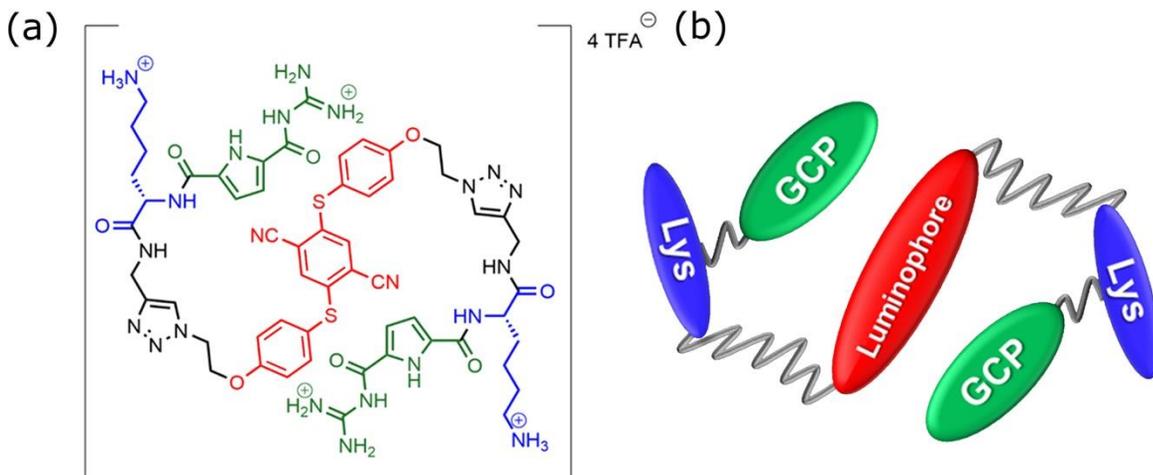


Figure 1: AIE-active molecule (**1**). (a) Structure of (**1**) with color-coded subunits AIE, lysine, and GCP. (b) Coarse-grained bead-spring model of (**1**).

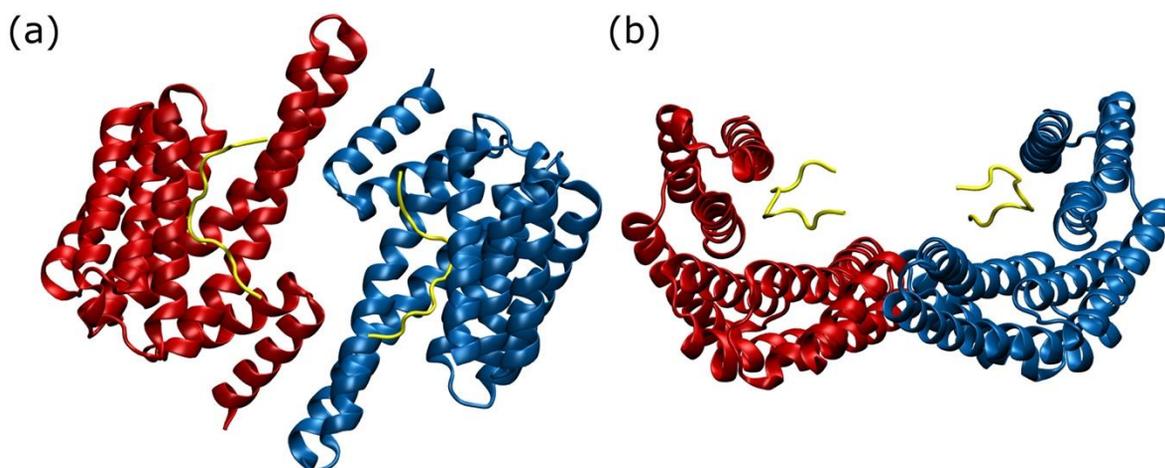


Figure 2: 14-3-3 ζ from (a) top and (b) side with the two monomers in red and blue. In the top view the central pore is clearly visible. The side view shows the ω shape with the two binding grooves for protein ligands, e.g., the C-Raf peptides (yellow). The structure is based on PDB entry 4IHL [18].

To identify the putative binding site of compound (**1**) on 14-3-3 ζ , and possibly to guide further experiments, we have developed and applied a computational model of the molecular system consisting of 14-3-3 ζ , (**1**), and the implicit solvent. Since this system is large and has many degrees of freedom, we did not use the expensive standard methods, such as molecular dynamics simulation with an all-atom force field. Instead, we devised a new method that allowed a complete screening

of the system at a meaningful level of accuracy and that identified putative binding sites of **(1)** on 14-3-3 ζ that can now be tested in a targeted way. We tried to answer basically two questions: (a) where does compound **(1)** bind on 14-3-3 ζ , and (b) how would the presence of phosphorylated C-Raf peptides in the binding grooves of 14-3-3 ζ affect the binding of **(1)**?

2. Results and Discussion

We explored potential binding positions by exhaustive simulations of **(1)** around 14-3-3 ζ with and without C-Raf peptides as described in section Experimental. The distribution of the final total energies are shown in Figure 3. Although the bulk of the energies follows a similar distribution in both simulation series, there is a remarkable difference, namely that in the simulations with C-Raf the log-scaled histogram has a tail to much lower energies. This difference must be due to interactions of **(1)** with C-Raf because otherwise the molecular systems are unchanged.

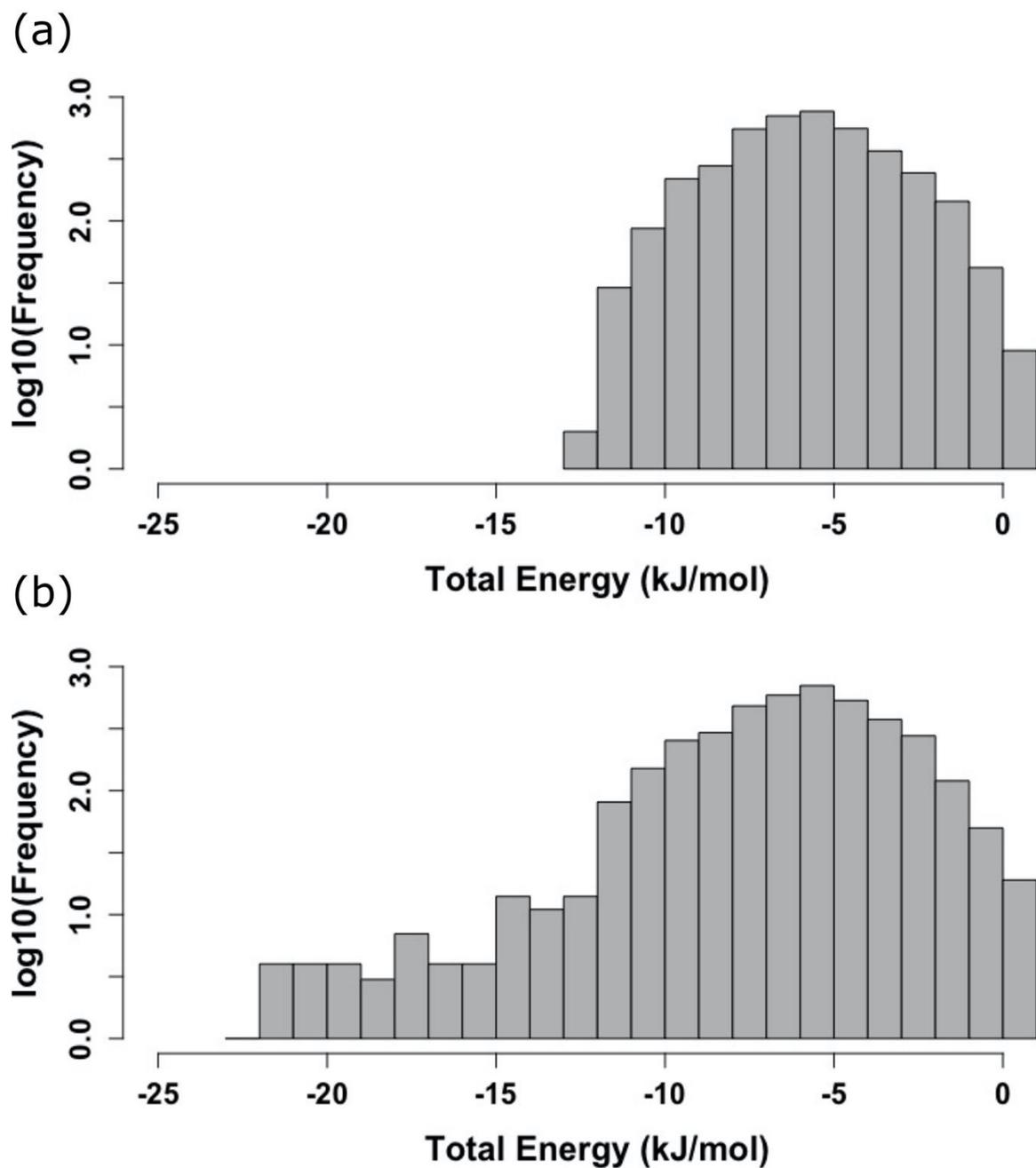


Figure 3: Log scaled histogram of total energies at the final steps of all simulations. (a) Simulated annealing (SA) runs with (1) around 14-3-3 ζ protein, (b) (1) around 14-3-3 ζ /C-Raf complex.

2.1 Predicted binding sites of (1) on 14-3-3 ζ

The total energies at the final positions of simulated annealing (SA) runs (Figure 4) show that (1) could bind in most regions on the surface of 14-3-3 ζ , especially in the absence of C-Raf (top row in Figure 4). Some areas seem to be favored: firstly, the tips of the omega ears, though this should be taken with caution as the model is lacking the very flexible C-termini that are in these regions; secondly, below the pore at the bottom of the omega in the case without C-Raf; thirdly, above the pore between the two binding grooves of the omega in the presence of C-Raf. Overall the clustering and the pattern of energies reflect the C_2 symmetry of the 14-3-3 ζ dimer.

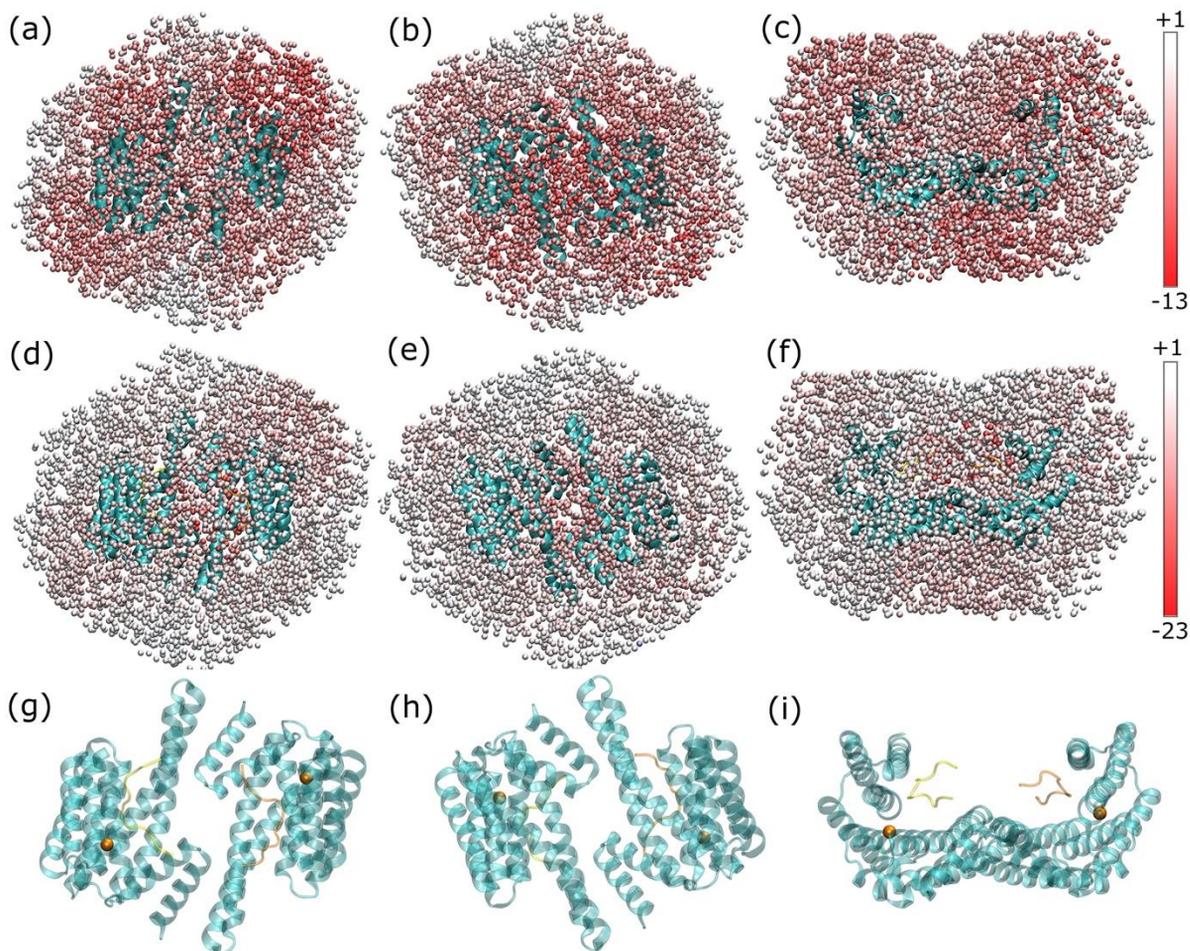


Figure 4: Sampled positions of the AIE moiety (a-f) colored according to the total energy of **(1)** from dark red (lowest energy) to white (highest energy). The values in the color bar are in units of kJ/mol. Top row: in absence of C-Raf; middle row: in presence of C-Raf; bottom row: structure of 14-3-3 ζ in corresponding orientations (same orientation in each column) with the top view in the left column, bottom view in the central column, side view in the right column. Little spheres in the bottom row are cluster medoids.

2.2 Minimum energy conformations

Closer inspection of energies reveals that in the absence of C-Raf peptides, the minimum energy conformation of **(1)** lies below the ω , under the central pore of 14-3-3 ζ (Figure 5a and b). With this conformation, one of the GCP groups falls into the minimum energy position on the affinity grid.

There is another minimum at the C_2 -symmetry related position of the mentioned GCP bead as well. Therefore, the second minimum energy conformation of the ligand (represented with larger beads in Figure 5) has one of its GCP groups in the C_2 -symmetry related position of the GCP bead in the first minimum energy conformation. The same happens to the Lys groups. As Figure 5 shows, the GCP and lysine beads are found near the hydroxyl groups of the aspartic acid residues of the protein.

In contrast, in the presence of C-Raf peptides the minimum energy positions of (**1**) move to the center of the ω where the AIE moieties hover above the central pore (Figure 5c and d). In these two conformations, at least one of the lysine groups falls into the minimum energy position of the lysine affinity map (more information about the affinity map is given in section 4.1). Both GCP groups fall into the two minimum energy positions of the GCP affinity map as well. In both minimum energy conformations of the ligand, at least one GCP or Lys group is found near one of the phosphorylated serine (pSer) residues of the C-Raf peptides (Figure 5d). This is to be expected due to the strong negative charge of pSer.

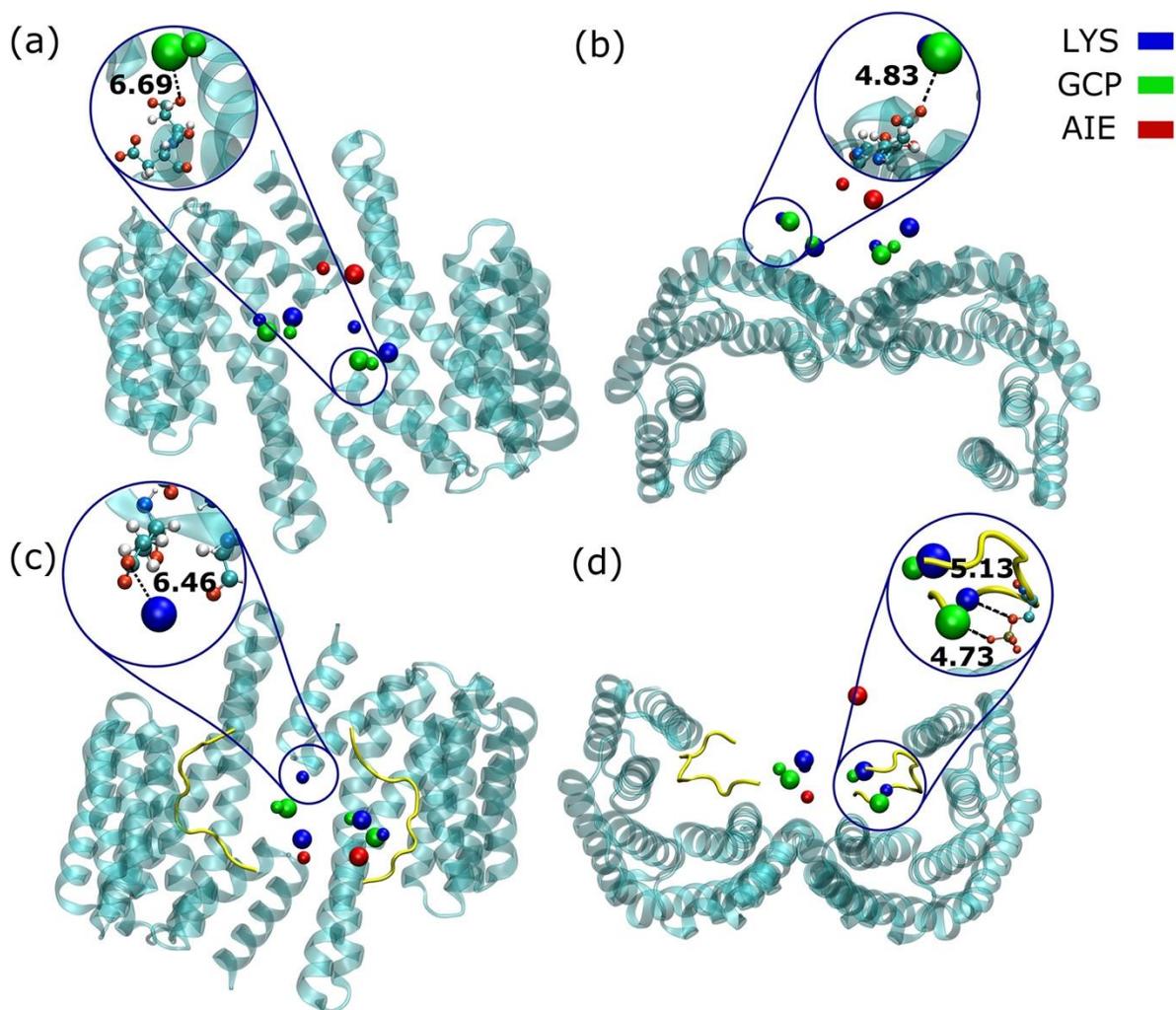


Figure 5: Minimum energy conformations of AIE ligand in the absence (a,b) and presence (c,d) of C-Raf peptides. The first and second minimum energy conformations are represented by small and large beads, respectively. In the absence of the C-Raf peptides, Lys (dark blue beads) and GCP (green beads) groups are found near the aspartic and glutamic residues of 14-3-3 ζ . In the presence of the peptides, at least one Lys or GCP group is found near the phosphorylated serine group (magnified image in (d)) of the C-Raf peptides. In this figure, all the distances are in Ångströms.

2.2.1 Binding experiments

A qualitative proof of binding of **(1)** and 14-3-3 ζ could be achieved with native gel electrophoresis.

It could be shown that **(1)** hinders some of the 14-3-3 ζ from entering the gel and that the mobility of 14-3-3 ζ correlates with the amount of **(1)** added. In fact, using default atomic charges [19, 20],

the 14-3-3 ζ dimer has a net charge of -32 |e| (proton charges), and the binding of two (**1**) leads to a net charge of -24 |e| of the complex.

It is known that aggregation-induced emission is caused by the blocking of non-radiative decay pathways. Moreover, fluorescence is quenched if two fluorophores come into close proximity including π - π -contacts, known as aggregation-caused quenching (ACQ) [21]. Contrary to our expectations we observed a decrease in emission upon titration of (**1**) to 14-3-3 ζ , in initial experiments, which motivated this contribution. Furthermore, we discovered that probably two ligands (**1**) simultaneously bind to one 14-3-3 ζ dimer (see Supporting Information File 1).

Our simulations predict that in the C-Raf-free complex of 14-3-3 ζ with (**1**) two AIE moieties in the minimum energy conformations of the ligand could be very close to each other (at a distance of 5.25 Å) (Figure 5a and b). If two molecules of (**1**) occupy these two positions at the same time, their fluorescence might be quenched. Yet, our model is too coarse to answer this question. Our simulations contain only a single moiety of (**1**) so interactions with another one are omitted. Moreover, our electrostatic model could not account for π - π contacts.

A more detailed experimental validation will be the focus of an experimental study in the near future. Three variants of (**1**) will be synthesized to gain a deeper insight into the binding stoichiometry and affinity, as well as the structure-property relationship on the influence on protein functions.

3. Conclusion

On the basis of the final energy values of 4000 of runs with a simulated annealing approach, we find the area above and below the central pore of 14-3-3 ζ protein to be the most probable binding site for (**1**). The position of the ligand is sensitive to the presence of phosphorylated C-Raf peptides

as interaction with these phosphorylated peptides draws (**1**) with its positive GCP and Lys groups into the region between the two binding grooves of 14-3-3 ζ with a much higher affinity. If presence of another ligand (like C-Raf) leads to rearrangement of binding of AIE-molecules around a target protein (like 14-3-3 ζ), and thus to differences in fluorescence, we have a mechanism that can possibly be exploited in analytical applications.

4. Experimental

4.1 Energy Grids

To compute a map of affinities of 14-3-3 ζ with GCP and lysine ligands, we proceeded as follows. The GCP and lysine ligand geometries were calculated in OpenBabel v2.3.2 [22] starting from a SMILES string of each ligand. Van der Waals radii of ligand atoms were added automatically by OpenBabel. The 14-3-3 ζ structure (PDB 4IHL, [18]) was refined with MODELLER [23] as described in reference [24]. Charges, van der Waals radii and missing hydrogen atoms were added by PDB2PQR v2.0.0 [19, 20] at pH 6.5 with the Amber force field option (values are provided in Table S3 in the Supporting Information File 2 and Table S4 in the Supporting Information File 3). The 14-3-3 ζ electrostatic field was calculated by solving the non-linear Poisson-Boltzmann equation with APBS version 1.5 [25] with ionic concentrations of 0.1 mol/L NaCl and 0.01 mol/L MgCl₂ and relative dielectric permittivities $\epsilon_r^{\text{protein}} = 2$ and $\epsilon_r^{\text{water}} = 79$. The 14-3-3 ζ environment was scanned with the GCP and lysine ligands separately in EpiTopsy [22] using 150 rotations and a grid resolution of 0.4 Å. We have also tried a grid resolution of 0.8 Å which would have allowed more efficient calculations. However, the 0.8 Å resolution was not fine enough to accurately capture essential properties of the system, especially the C₂ symmetry of 14-3-3 ζ . Affinity grids

for lysine and GCP have been obtained once using the structure of 4IHL (complex C-Raf/14-3-3 ζ) and once using only the structure of 14-3-3 ζ (4IHL without C-Raf peptides). These affinity grids are the only difference between the two series of simulations (in the presence and absence of C-Raf peptides). Note that the above approach models only interactions of individual GCP or lysine groups with 14-3-3 ζ . To model the interactions of 14-3-3 ζ with the full multivalent (**1**) ligand, we have combined these and further interactions in a single model as described in the following.

4.2 Coarse-Grained Model

We developed a coarse-grained model of (**1**) known as the Bead-Spring model in the context of polymer studies [26-28]. Our model is a chain of five beads, GCP-Lys-AIE-Lys-GCP, with each pair of neighbors in the chain connected by a harmonic spring potential (Figure 1b). Given the symmetry of (**1**), there are two types of springs, GCP-Lys and Lys-AIE. The spring parameters are chosen such that the five beads are always within a reasonable distance from each other, i.e. close to the distances in an atomistic model of the ligand structure. This atomistic model of the ligand was created in ChemDraw prime 16.0 (PerkinElmer, Waltham, MA, USA). The structure was then imported in Maestro (Schrödinger Maestro Version 11.5.011, MM share Version 4.1.011, Release 2018-1, Platform Windows-x64) and finally, the energy of the structure was minimized.

Overlaps between non-bonded beads are avoided by repelling potentials. Equation 1 represents this in the form of hard sphere potential.

$$E_{repelling} = \begin{cases} \infty & d < a \\ 0 & d \geq a \end{cases} \quad (1)$$

where d is the distance between the centers of two non-bonded beads and a is the average of pairwise addition of the radii of all the non-bonded beads (i.e. Lys-Lys, GCP-GCP, GCP-Lys, and GCP-AIE). According to the length scales in our atomistic model of the ligand, the value of 9.6 Å

is predicted for a . To prevent possible overlaps between the ligand and the protein, atoms of the ligand are not allowed to enter the van der Waals radii of protein atoms. This model interacts with 14-3-3 ζ through the energy grids of lysine and GCP as mentioned above. Equation 2 describes different terms of the total energy.

$$E_{total} = \frac{1}{2} \sum_{i=1}^4 k_i (l_i - l_i^{eq})^2 + E_g^{GCP_1} + E_g^{Lys_1} + E_g^{GCP_2} + E_g^{Lys_2} \quad (2)$$

where k_i is the spring constant, l_i is the distance between bonded beads i and $i+1$, and l_i^{eq} is the corresponding distance at equilibrium, E_g represents the affinity of protein toward GCP or lysine at a specific grid point. In this study, l_1 and l_4 , i.e. the equilibrium lengths for the springs connecting GCP to lysine, have been selected to be 6 Å and l_2 and l_3 , i.e. the equilibrium lengths for the springs connecting lysine to AIE, have been selected to be 13 Å. The value of the spring constants in bead-spring models with Gaussian probability distribution is inversely proportional to the square of the equilibrium length of the spring [29]. Based on that, the values of k_1 and k_4 (spring constants for GCP-Lys bonds) were chosen to be $12 \text{ k}_B\text{T}/\text{Å}^2$ while the values of k_2 and k_3 (spring constants for Lys-AIE bonds) were chosen to be $3 \text{ k}_B\text{T}/\text{Å}^2$. With these parameters, the sum of harmonic potentials in the relaxed structure of the ligand will be comparable to the sum of GCP and lysine affinities (see Equation 2 for total potential). Otherwise, the minimized total energy might be misleading (the harmonic term might be minimized while the term related to the grid affinities might not be minimized). What matters is to find the minimum energy conformation of the ligand with regard to the affinity maps of lysine and GCP.

4.3 Simulated Annealing Schedule

To identify the energy minima of the coarse-grained ligand model around 14-3-3 ζ , we used simulated annealing. This computational method is typically applied to identify the global optimum of an objective function, even if there are many local optima [30-32].

At the heart of our implementation of the SA method we carried out Markov Chain Monte Carlo (MCMC) simulations of (1) around 14-3-3 ζ with the acceptance criterion of *Metropolis et al.* [33], as detailed below. We did many MCMC simulations at temperatures given by the following schedule. We started the simulations at a very high temperature (3000 K) and slowly cooled the system down to low enough temperatures at which an acceptance ratio of less than 1% was achieved. At each temperature level, we ran an MCMC simulation with the Metropolis criterion. The Monte Carlo trajectory ended at each temperature level after 10000 moves, or after 1000 accepted moves, whichever was reached sooner. With this protocol good convergence to minimum energies was achieved (see Figure S2 in the Supplementary Information).

We did 4000 SA runs to identify the globally optimal configuration of (1) around 14-3-3 ζ . At the beginning of each simulation, we put the AIE bead at a random grid position within a volume layer around the protein (we used the same grid geometry as in the energy grids described above). The thickness of the layer corresponded to the distance between AIE and GCP. Its volume is about 500 nm³, so that the 4000 AIE positions sample the layer at a density of about 8 per nm³ (Figure 4a-f). After its random placement, the position of AIE bead was kept fixed throughout the respective simulation. In each MCMC step, just one of the lysine or GCP beads was randomly selected and moved by a maximum of 6 grid lengths along each of the three Cartesian coordinate axes. If in the new position the bead fell within the specified distance of any of the non-bonded beads, the move was rejected due to the repelling potential. Otherwise, the energy with the new

ligand position was evaluated and compared to the old energy in the Metropolis criterion to accept or reject the move. Our code was written in the Julia language and ran with Julia version 1.5.2 [34].

4.4 Clustering

Applying a clustering method to the final positions of (1) provides a sanity check for our simulations. Considering the fact that 14-3-3 ζ is a homodimer with C_2 symmetry, we expect for each binding site (or cluster of binding sites) that does not lie on the symmetry axis a symmetry related binding site (or cluster).

We used a robust silhouette-validated PAM (partitioning around medoids) clustering for this purpose [35]. Optimal number of clusters was extracted and visualized (in Supporting Information File 1) using the *fviz_nbclust* tool from R package *factoextra* version 1.0.7, and clustering was done using the *pam* tool from R package *cluster* version 2.1.0, in R version 3.5.3 [36].

Supporting Information

Supporting Information File 1:

- File name: “SI.docx”
- File format: Microsoft Word
- Title: General information and instrumentation, cluster analysis, electrostatic potential surface of 14-3-3 ζ , total energy of a simulation.

Supporting Information File 2:

- File name: “Table-S3.xlsx”
- File format: Microsoft Excel

- Title: PDB2PQR output for lysine

Supporting Information File 3:

- File name: “Table-S4.xlsx”
- File format: Microsoft Excel
- Title: PDB2PQR output for GCP

Acknowledgements

Steffen Riebe is acknowledged for support during the synthesis of the azide precursor. Ali Khaleghzadegan and Daniel Habermann are warmly appreciated for their help and advice in tuning and optimizing the code performance.

Funding

The authors wish to thank the Deutsche Forschungsgemeinschaft (CRC 1093, projects A1, A7 and A10, B4, and B5) for financial support.

References

1. Aitken, A. *Plant Mol. Biol.* **2002**, *50*, 993-1010.
2. Freeman, A.K.; Morrison, D.K. 14-3-3 Proteins: diverse functions in cell proliferation and cancer progression. In *Seminars in cell & developmental biology*; Elsevier, 2011; pp 681-687.
3. Mhaweche, P. *Cell Research* **2005**, *15*, 228-236.
4. Fu H.; Subramanian R. R.; Masters S. C. *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40*, 617-647.
5. Kaplan, A.; Ottmann, C.; Fournier, A.E. *Pharmacol. Res.* **2017**, *125*, 114-121.
6. Maity, D.; Gigante, A.; Sánchez-Murcia, P.A.; Sijbesma, E.; Li, M.; Bier, D.; Mosel, S.; Knauer, S.; Ottmann, C.; Schmuck, C. *Org. Biomol. Chem.* **2019**, *17*, 4359-4363.
7. Milroy, L.G.; Grossmann, T.N.; Hennig, S.; Brunsveld, L.; Ottmann, C. *Chem. Rev.* **2014**, *114*, 4695-4748.
8. Vallet, C.; Aschmann, D.; Beuck, C.; Killa, M.; Meiners, A.; Mertel, M.; Ehlers, M.; Bayer, P.; Schmuck, C.; Giese, M.; Knauer, S.K. *Angew. Chem. Int. Ed.*, **2020**, *59*, 5567-5571.
9. Schmuck, C.; Bickert, V. *J. Org. Chem.* **2007**, *72*, 6832-6839.
10. Schug, K.A.; Lindner, W. *Chem. Rev.* **2005**, *105*, 67-114.
11. Gigante, A.; Grad, J.N.; Briels, J.; Bartel, M.; Hoffmann, D.; Ottmann, C.; Schmuck, C. *Chem. Commun.* **2019**, *55*, 111-114.
12. Bartsch, L.; Bartel, M.; Gigante, A.; Iglesias-Fernández, J.; Ruiz-Blanco, Y.B.; Beuck, C.; Briels, J.; Toetsch, N.; Bayer, P.; Sanchez-Garcia, E.; Ottmann, C. *ChemBioChem* **2019**, *20*, 2921-2926.

13. Ehlers, M.; Grad, J.N.; Mittal, S.; Bier, D.; Mertel, M.; Ohl, L.; Bartel, M.; Briels, J.; Heimann, M.; Ottmann, C.; Sanchez- Garcia, E. *ChemBioChem* **2018**, *19*, 591-595.
14. Mei, J.; Leung, N.L.; Kwok, R.T.; Lam, J.W.; Tang, B.Z. *Chem. Rev.* **2015**, *115*, 11718-11940.
15. Shi, H.; Liu, J.; Geng, J.; Tang, B.Z.; Liu, B. *J. Am. Chem. Soc.* **2012**, *134*, 9569-9572.
16. Khan, I.M.; Niazi, S.; Khan, M.K.I.; Pasha, I.; Mohsin, A.; Haider, J.; Iqbal, M.W.; Rehman, A.; Yue, L.; Wang, Z. *TrAC, Trends Anal. Chem.* **2019**, *119*, 115637.
17. Riebe, S.; Vallet, C.; van der Vight, F.; Gonzalez- Abradelo, D.; Wölper, C.; Strassert, C.A.; Jansen, G.; Knauer, S.; Voskuhl, J. *Chem. Eur. J.* **2017**, *23*, 13660-13668.
18. Molzan, M.; Kasper, S.; Röglin, L.; Skwarczynska, M.; Sassa, T.; Inoue, T.; Breitenbuecher, F.; Ohkanda, J.; Kato, N.; Schuler, M.; Ottmann, C. *ACS Chem. Biol.* **2013**, *8*, 1869-1875.
19. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A. *Nucleic Acids Res.* **2004**, *32*, W665-W667.
20. Dolinsky, T.J.; Czodrowski, P.; Li, H.; Nielsen, J.E.; Jensen, J.H.; Klebe, G.; Baker, N.A. *Nucleic Acids Res.* **2007**, *35*, W522-W525.
21. Miyawaki, A. *Annu. Rev. Biochem.* **2011**, *80*, 357-373.
22. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. *J. Cheminformatics* **2011**, *3*, 33.
23. Šali, A.; Blundell, T.L. *J. Mol. Biol.* **1993**, *234*, 779-815.
24. Grad, J.N., Gigante, A., Wilms, C., Dybowski, J.N., Ohl, L., Ottmann, C., Schmuck, C. and Hoffmann, D. *J. Chem. Inf. Model.* **2018**, *58*, 315-327.
25. Baker, N.A.; Sept, D.; Joseph, S.; Holst, M.J.; McCammon, J.A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037-10041.

26. Milchev, A.; Binder, K. *Macromolecules* **1996**, *29*, 343-354.
27. Underhill, P.T.; Doyle, P.S. *J. Non-Newtonian Fluid Mech.* **2004**, *122*, 3-31.
28. Liu, B.; Wang, J.; Fan, X.; Kong, Y.; Gao, H. *J. Comput. Phys.* **2008**, *227*, 2794-2807.
29. Baschnagel, J.; Wittmer, J. P.; Meyer, H. Monte Carlo Simulation of Polymers: Coarse-Grained Models. In *Computational soft matter: from synthetic polymers to proteins*; Attig, N.; Binder, K.; Grubmüller, H.; Kremer, K.; John von Neumann Institute for Computing (NIC), Vol. 23; NIC Series: Jülich, 2004; pp 83-140.
30. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. *Science* **1983**, *220*, 671-680.
31. Černý, V. *J. Optim. Theor. Appl.* **1985**, *45*, 41-51.
32. Henderson, D.; Jacobson, S.H.; Johnson, A.W. The theory and practice of simulated annealing. In *Handbook of metaheuristics*; Springer: Boston, MA, 2003; pp 287-319.
33. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087-1092.
34. Bezanson, J.; Edelman, A.; Karpinski, S.; Shah, V.B. *SIAM review* **2017**, *59*, 65-98.
35. Dodge, Y. *Statistical data analysis based on the L1-norm and related methods*; Birkhäuser, 2012.
36. R 3.5.3; R Foundation for Statistical Computing: Vienna, AT, 2019.