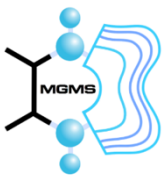
10th Annual CCPBioSim & MGMS 2024 Conference:

Molecular modelling in structure-based drug design

Programme and Abstract Booklet

1st - 3rd July 2024

Newcastle University



Hello everyone, and welcome to the 10th Annual CCPBioSim and Molecular Graphics and Modelling Society 2024 Conference

"Molecular modelling in structure-based drug design"!

We are a highly multidisciplinary community which means we all have something valuable to share, and new things to learn from each other. We hope that you enjoy the posters and talks, see old friends and build new connections.

We hope to see you all in Newcastle either in person or virtually!

The CCPBioSim and MGMS Conference Committee

1. Conference Venue

The conference will take place in the Dame Margaret Barbour Building at Newcastle University:

Dame Margaret Barbour Building, Newcastle upon Tyne NE2 4DR.

Walking to the University Campus from Newcastle Central Station takes around 30 minutes. Newcastle also has a metro system (https://www.nexus.org.uk/metro). The closest station to campus is the Haymarket Station.

The registration, talks, coffee and lunch breaks will take place in the Dame Margaret Barbour Building. On arrival, please register at the registration desk at the entrance, and pick up your badge.

The conference sessions will take place in room DMBB.G.01 on the ground floor of the Dame Margaret Barbour Building.

Dame Margaret Barbour Building, Newcastle University

2. Registration

When you arrive on Monday 1st July, please come and register. You will get a name badge, a printed conference programme, information about the posters, internet connection details etc.

3. Catering

On Monday 1st July, Tuesday 2nd July and Wednesday 3rd July there will be lunch available in the Dame Margaret Barbour Building. The lunch will be buffet style and the food options are clearly labelled.

Our conference dinner, which is included in the registration fee, will take place on Tuesday 2nd July at 7:30 pm at the Biscuit Factory (see section 5 below for details).

In our coffee breaks there will be coffee and tea available, with pastries or biscuits.

During poster sessions on Monday and Tuesday there will be a drinks reception with snacks.

PLEASE NOTE THAT NO DINNER IS PROVIDED FOR MONDAY NIGHT

4. Posters

All posters should be put up during the registration on July 1st and they must be taken down on July 3rd. The poster boards are located in the seminar room in the Baddiley Clark Building NE2 4BN (Please ask for directions at registration).

Please check the poster number you have been assigned (the list of posters follows the programme below). The poster boards will be numbered so please put yours in the right place.

If your poster has an odd number, we ask you to stand in front of your poster in the poster session and engage with the other delegates on Monday; on Tuesday, you can wander around and look at the other posters. If your poster has an even number, we ask you to stand in front of your poster on Tuesday, and you can look at the other posters on Monday.

The poster boards fit A0 posters portrait and only A1 posters fit landscape. **Please note A0 posters will not fit landscape**. Please have your poster with you when you come; it is not possible to print it on the day.

Poster competition:

All posters take part in our poster competition. The posters will be judged on two factors:

1) the visual clarity of the poster
2) the interaction of the poster presenter during poster sessions

We will announce the 4 winners during our conference dinner on Tuesday evening. The prizes are kindly donated by our 4 industry sponsors: Cancer Research Horizons, Kuano, OpenBioSim and RxCelerate, as well as the Royal Society of Chemistry journals – we would like to extend our huge thanks to our sponsors for the prizes.

5. Conference dinner

Time – 7:30 pm Tuesday, July 2nd

The Biscuit Factory, 16 Stoddart St, Shieldfield, Newcastle upon Tyne NE2 1AN

Our conference dinner is included in the registration fee.

Photo of the conference dinner venue at the Biscuit Factory, Newcastle

Everybody is invited to take part in the three-course conference dinner. There will be wine and soft drinks available at the tables and a bar where you can buy more drinks. **IF FOR ANY REASON YOU ARE UNABLE TO ATTEND THE DINNER, PLEASE INFORM A MEMBER OF THE ORGANISING COMMITTEE AS SOON AS POSSIBLE.**

We have informed the venue of any dietary requirements that you requested at registration.

6. Internet access

Eduroam is available or visitors can access free wifi provided by the cloud.

7. Sponsors

We are grateful to our sponsors:

8. Organising Committee

Dr Daniel Cole, Newcastle University
Dr Agnieszka Bronowska, Newcastle University
Dr Natalie Tatum, Newcastle University

We are especially grateful to the following for their contributions to the conference:

Katy Sawyer, Newcastle University
Ben Cree, Newcastle University
Asma Feriel Khoualdi, Newcastle University
Sarah Harris, University of Sheffield
Sarah Fegan, STFC
Georgia Lomas, STFC
Steve Maginn, Chemical Computing Group

Email: ccpbiosim2024@group.newcastle.ac.uk

Programme

Day 1: Monday 1st July

12:00 – 13:00	Registration and Lunch	
13:00 – 13:05	Daniel Cole	Welcome
Session 1: From Desktop to Benchtop and Back Chair: Natalie Tatum		
13:05 – 13:35 T1	**Frank von Delft** Diamond Light Source	Fast-Forward Fragments for rapidly progressing structure-based (XChem) fragment screens
13:35 – 13:55 T2	**Anna Duncan** Aarhus University	Why Losing a Lipid Tail Matters: Cardiolipin, and not Monolysocardiolipin, Preferentially Binds to the Interface of Complexes III and IV
13:55 – 14:15 T3	**Daniella Hares** Institute of Cancer Research	Decoding BCL6 Inhibitors: Computational Insights into the Impact of Water Networks on Potency
14:15 – 14:45 T4	**Giulio Tesei** University of Copenhagen	Data-driven Modelling of the Intrinsically Disordered Proteome
14:45 – 15:20	Coffee & Tea	
15:20 – 15:50 T5	**Matthieu Schapira** (online) University of Toronto	Lessons Learned from the First CACHE Challenge
15:50 – 16:10 T6	**Julian Streit** University College London	The Ribosome Lowers the Entropic Penalty of Protein Folding
16:10 – 16:30 T7	**Isabel Elliott** University of Southampton	Structure-guided Disulfide Engineering Restricts Antibody Conformation and Flexibility to Elicit TNFR Agonism in Anti-cancer Therapeutics
16:30 – 16:50 T8	**James Krieger** National Centre for Biotechnology, Madrid	Improving Continuous Heterogeneity Landscapes and Interpretations using Atomic Structures: A Case Study of SARS-CoV-2 Spike Variants
16:50 – 18:30	Poster session (odd numbers)	

Day 2: Tuesday 2nd July

Session 2: Molecular Modelling Methods Chair: Agnieszka Bronowska		
9:00 – 9:30 T9	**Antonia Mey** University of Edinburgh	From Active Learning to Zinc: Adventures with Alchemical Free Energy Calculations
9:30 – 9:50 T10	**Katarzyna Zator** University of Cambridge	Atom Surface Site Interaction Mapping of Protein-Ligand Complexes
9:50 – 10:10 T11	**Wojciech Kopec** Queen Mary University of London	Computational Electrophysiology of Potassium and Chloride Channels
10:10 – 10:30 T12	**Sofia Oliveira** University of Bristol	Using Dynamical-Nonequilibrium MD Simulations to Understand Drug Resistance and Allostery in Proteins
10:30 – 11:00	Coffee & Tea	
11:00 – 11:30 T13	**Marco de Vivo** Istituto Italiano di Tecnologia	Targeting the Conserved Active Site of Splicing Machines
11:30 – 11:50 T14	**Finlay Clark** University of Edinburgh	Automated Adaptive Absolute Binding Free Energy Calculations
11:50 – 12:20 T15	**Maria Kurnikova** Carnegie Mellon University	Active Learning Driven Hit Mining and Optimization based on Molecular Dynamics Simulated Free Energies
12:20 – 13:30	Lunch	

Day 2: Tuesday 2nd July
Afternoon

Session 3: Machine Learning Chair:		
13:30 – 14:00 T16	**Cecilia Clementi** (online) Freie Universität Berlin	Navigating protein landscapes with a machine-learned transferable coarse-grained model
14:00-14:20 T17	**Elliot Chan** University of Bristol	Electrostatic Embedding of Machine Learned Potentials for Accurate and Efficient Simulation of Enzyme Catalysis
14:20 – 14:40 T18	**Christopher Williams** University of Manchester	Stable and Accurate Simulations of Drug Molecules using Conformationally Generalisable Machine Learned Potentials
14:40 – 15:00 T19	**Francesc Sabanes** Acellera Therapeutics	Enhancing Protein-Ligand Binding Affinity Predictions using Neural Network Potentials
15:00 – 15:30	Coffee & Tea	
15:30 – 16:00 T20	**Mohammed AlQuraishi** Columbia University	The State of Protein Structure Prediction and Friends
16:00 – 16:20 T21	**Jonathan Heal** RxCelerate	Real World Comparison of Dynamic Ultra Large Library Screening with Commercial Library Screening against Different Target Classes
16:20 – 16:40 T22	**Will Gerrard** Kuano	Application of AI Augmented Design in Early Stage Drug Discovery
16:40 – 18:15	Poster Session (even numbers)	
18:15	Move to the Biscuit Factory	
19:00	**Conference Dinner**	

Day 3: Wednesday 3rd July

Session 4: Case Studies in Computer-Aided Design Chair:		
9:00 – 9:30 T23	**Giulia Rossetti** (online) Forschungszentrum Jülich	AI-based Identification of Therapeutic Agents Targeting GPCRs: Introducing Ligand Type Classifiers and Systems Biology
9:30 – 9:50 T24	**Richard Heath** Newcastle University	Using SeeSAR to Identify Novel Therapeutic Starting Points for Friedreich's Ataxia
9:50 – 10:10 T25	**Jordi Juarez-Jimenez** Universitat de Barcelona	Computational Tools for the Rational Design of Molecular Glues
10:10 – 10:30 T26	**Kin Chao** Imperial College London	Modulation of Class B1 GPCRs by the Plasma Membrane Environment
10:30 – 11:00	Coffee & Tea	
11:00 – 11:30 T27	**Joe Bluck** Bayer	Prioritizing Molecules for Synthesis using In-Silico Tools
11:30 – 11:50 T28	**Joshua Horton** Newcastle University	ASAP-Alchemy: A State-of-the-Art Open-Source Alchemical Free Energy Pipeline with Bespoke Force Fields
11:50 – 12:10 T29	**Stefano Serapian** University of Pavia	MD Simulations Map the Havoc Wreaked by One Mutation in $[Hsp60]_{14}$
12:10 – 12:30 T30	**Salome Llabres** Universitat de Barcelona	Unravelling the Energetics of a Small Molecule-Induced Disorder-to-Order Transition
12:30		Closing remarks
12:45	Conference closes & Lunch	

Poster Presentations

1	Adams, Charlie
2	Alhabradi, Thuraya
3	Alyemni, Saleh
4	Begum, Jaida
5	Blanco Gabella, Patricia
6	Burman, Matthew
7	Casteller, Lisa
8	Cherednichenko, Anton
9	Clark-Nicolas, Joan
10	El Khaoudi Enyoury, Hocine
11	Emam, Aya
12	Fegan, Sarah
13	Figueroa Blanco, David Ricardo
14	Guljas, Andrea
15	Holdship, Charlie
16	Hollands, Katy
17	Houppy, William
18	Hsiao, Ya-Wen
19	Ivanova, Varbina
20	Kalayan, Jas
21	Kalita, Papu
22	Kalpokas, Audrius
23	Kazmierczak, Magdalena
24	Khoshravan Azar, Asal
25	Khoualdi, Asma Feriel

26	Mattia, Marco
27	Moore, Harry
28	Morado, Joao
29	Morton, William
30	Notari, Evangelia
31	Pakamwong, Bongkochawan
32	Papa, Ioana
33	Parrag, Matyas
34	Peeters, Jordy
35	Pornprom, Thimpika
36	Puch Giner, Ignasi
37	Pungpo, Pornpan
38	Ramos, Carlos
39	Santos, Sonia
40	Serrano-Aparicio, Natalia
41	Taylor, Matthew
42	Velasco Berrelleza, Victor
43	Veselu, Diana
44	Weston, George
45	White, Elise
46	Windeln, Leonie
47	Yong, Chin
48	Zhang, Chenfeng
49	Zhu, Ryan

Fast-Forward Fragments for rapidly progressing structure-based (XChem) fragment screens

Frank von Delft[1]
[1]*Diamond Light Source*

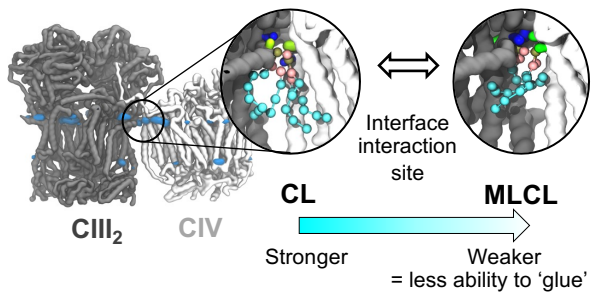
Why losing a lipid tail matters: Cardiolipin, and not monolysocardiolipin, preferentially binds to the interface of complexes III and IV

Anna L Duncan[1]; Noah Harrison[2]; Phillip J Stansfeld[3]; Robin A. Corey[2,4]
[1]*Department of Chemistry, Aarhus University, Denmark*
[2] *Biochemistry Department, University of Oxford, UK*
[3]*Department of Chemistry, University of Warwick, UK*
[4]*Currrent address: School of Physiology, Pharmacology & Neuroscience, University of Bristol*

Corresponding author: anna.duncan@chem.au.dk

Abstract

The mitochondrial electron transport chain comprises a series of protein complexes embedded in the inner mitochondrial membrane that generate a proton motive force via oxidative phosphorylation, ultimately generating ATP. These protein complexes can oligomerize to form larger structures called supercomplexes. Cardiolipin (CL), a conical four-tailed lipid, unique within eukaryotes to the inner mitochondrial membrane, has proven essential in maintaining the stability and function of supercomplexes. Monolysocardiolipin (MLCL) is a CL variant that accumulates in people with Barth syndrome (BTHS), and differs from CL only by the loss of one lipid tail. BTHS is caused by defects in CL biosynthesis and characterised by abnormal mitochondrial bioenergetics and destabilised supercomplexes. However, the mechanisms by which MLCL causes pathogenesis remain unclear. Here, multiscale molecular dynamics characterise the interactions of CL and MLCL with yeast and mammalian mitochondrial supercomplexes containing complex III (CIII) and complex IV (CIV). Coarse-grained simulations reveal that both CL and MLCL bind to sites at the interface between CIII and CIV of the supercomplex. Free energy perturbation calculations show that MLCL interaction is weaker than that of CL and suggest that interaction with CIV drives this difference. Atomistic contact analyses show that, although interaction with CIII is similar for CL and MLCL, CIV makes more contacts with CL than MLCL, demonstrating that CL is a more successful “glue” between the two complexes. The study suggests that MLCL accumulation in people with BTHS disrupts supercomplex stability by formation of relatively weak interactions at the interface lipid binding site.



Cardiolipin (CL) vs monolysocardiolipin (MLCL) interaction at the interface of Complex III_2 and Complex IV

References

A. Corey, R.; Harrison, N.; J. Stansfeld, P.; P. Sansom, M. S.; L. Duncan, A. *Chemical Science* **2022**, *13* (45), 13489–13498. https://doi.org/10.1039/D2SC04072G.

Decoding BCL6 Inhibitors: Computational Insights into the Impact of Water Networks on Potency

Daniella Hares[1]; Andrea Scarpino[1], Michael Bodnarchuk[2*], Swen Hoelder[1*]

[1] *Centre for Cancer Drug Discovery, The Institute of Cancer Research, London SM2 5NG, UK*

[2] *Oncology R&D, AstraZeneca, Cambridge Biomedical Campus, Cambridge CB2 0AA, UK*

Corresponding authors: swen.hoelder@icr.ac.uk, michael.bodnarchuk@astrazeneca.com

Abstract

Water networks can have a critical role in small molecules binding to a target protein and are an important consideration in structure-based drug design. Previous computational studies have predicted protein-bound water molecules to contribute between -3 to -6 kcal mol^{-1} towards binding affinity [1]. When modifying the molecular structure of a lead compound, the rearrangement of surrounding water networks in the binding site can impact potency but this contribution is challenging to measure experimentally. Computational methods are ideally suited to study the interplay between ligand optimisation and water displacement by predicting the effect of structural changes on both the activity of the compound and the stability of neighbouring water molecules.We used Grand Canonical Monte Carlo simulations and alchemical free energy calculations to retrospectively rationalise the trends in potency observed in a set of B-cell Lymphoma 6 (BCL6) inhibitors [2]. As part of the drug design process, the inhibitor molecular structure was modified to displace water molecules that were part of a network within the protein binding site.Using the BCL6 project as an example, this talk will demonstrate how computational approaches can shine a light into an aspect that is often overlooked but essential to guide the design of better compounds when working with hydrated protein pockets. We show the power of these methods and encourage their use more widely, particularly in prospective applications on drug discovery projects.

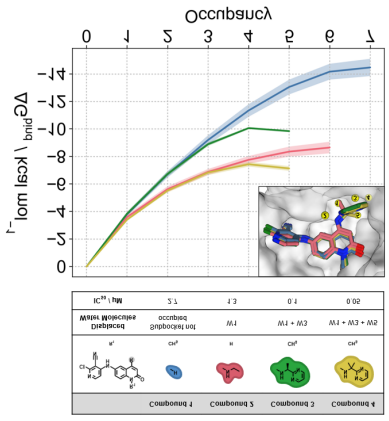


Figure 1: *Four compounds selected from Lloyd et al. [2] that sequentially displace up to three water molecules from a network within a subpocket of the BCL6 binding site. The IC_{50} values were calculated using TR-FRET. Using GCMC simulations performed at different chemical potentials, the binding free energy of the water molecules in the subpocket compared to bulk solvent can be calculated through the Grand Canonical Integration method [3].*

References

[1] Mobley, D. L. and Dill K. A. (2009). *Structure*, **17** (4), 489–498.

[2] Lloyd, M. G., Huckvale, R., Cheung, K.-M. J., Rodrigues, M. J., Collie, G. W., Pierrat, O. A., Gatti Iou, M., Carter, M., Davis, O. A., McAndrew, P. C., Gunnell, E., Le Bihan, Y.-V., Talbot, R., Henley, A. T., Johnson, L. D., Hayes, A., Bright, M. D., Raynaud, F. I., Meniconi, M., Burke, R., van Montfort, R. L. M., Rossanese, O. W., Bellenie, B. R., Hoelder, S. (2021). *J. Med. Chem.*, **64** (23), 17079–17097.

[3] Ross, G. A., Bodnarchuk, M. S. and Essex, J. W. (2015). *J. Am. Chem. Soc.*, **137** (47), 14930–14943.

Data-driven modelling of the intrinsically disordered proteome

Giulio Tesei[1]; Anna Ida Trolle[1]; Nicolas Jonsson[1]; Johannes Betz[1]; Frederik E. Knudsen[1]; Francesco Pesce[1]; Kristoffer E. Johansson[1]; Kresten Lindorff-Larsen[1]

[1]*Structural Biology and NMR Laboratory, Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Copenhagen, Denmark*

Corresponding author: giulio.tesei@bio.ku.dk

Abstract

Intrinsically disordered regions (IDRs) constitute about one third of the human proteome and play important roles in biological processes. While lacking well-defined 3D structures, IDRs adopt heterogeneous conformational ensembles which are determined by the amino acid sequence, yet not captured by existing methods for structure prediction. I will first present CALVADOS, an efficient coarse-grained molecular model of IDRs [1,2]. The model was trained on experimental data reporting on conformational properties and validated extensively against both conformational and phase properties of IDRs with diverse sequences. Second, I will describe how we investigated relationships between sequence, conformational ensembles, and biological functions using molecular simulations of 28,058 human IDRs performed with CALVADOS [3]. Through bioinformatics analyses, we explored the relationship between the conformational properties of the IDRs, which we obtained from simulations, and the biological function and cellular localization of the full-length proteins. Further, we looked at how sequence features relate to conformational properties and analysed the evolutionary conservation of compaction across homologous IDRs. Our work exemplifies how a proteome-wide database of conformational ensembles can be used to gain insights into sequence–ensemble relationships and to formulate hypotheses on the biological role and evolution of human IDRs.

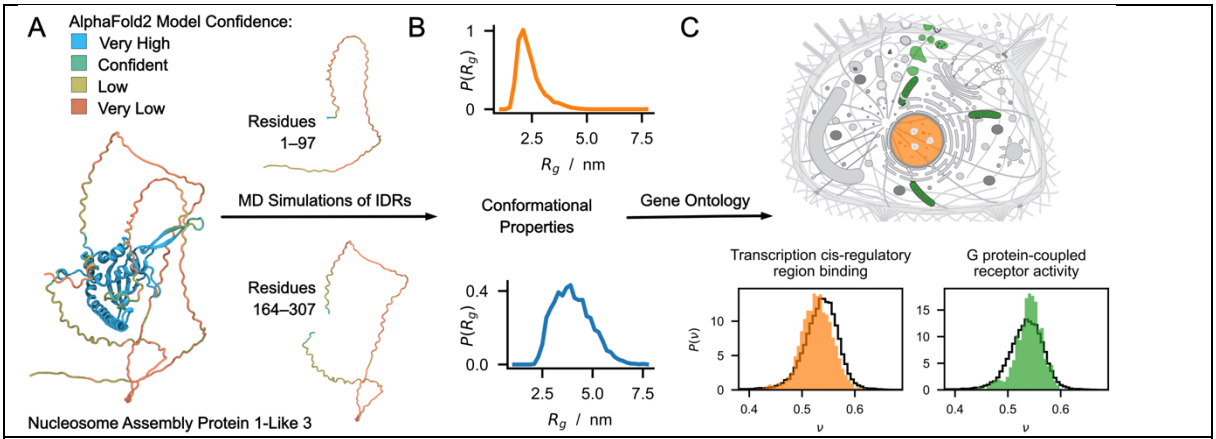


Figure 1: *Approach for proteome-wide study of human IDRs. (A) Selection of IDRs using AlfaFold2 predictions. (B) Generation of conformational ensembles via CALVADOS simulations. (C) Linking IDR conformational properties to the cellular localization and function of the full-length protein.*

References

[1] Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. (2021). *PNAS* **118**(44), e2111696118.
[2] Tesei, G.; Lindorff-Larsen, K. (2023). *Open Research Europe* **2**, 94.
[3] Tesei, G.; Trolle, A. I.; Jonsson, N. et al. (2024). *Nature* **626**, 897–904.

Lessons learned from the first CACHE challenge.

Matthieu Schapira[1]

[1]*University of Toronto*

The CACHE challenges are prospective benchmarking exercises where computationally predicted compounds are tested experimentally. In its first iteration, 23 computational teams collectively predicted 1955 compounds targeting the Parkinson's Disease target LRRK2. The seven successful computational workflows varied in their screening strategies and techniques. Three used molecular dynamics to produce a conformational ensemble of the targeted site, three included a fragment docking step, three implemented a generative design strategy and five used one or more deep learning steps. CACHE #1 reflects a highly exploratory phase in computational drug design where participants sometimes adopted strikingly diverging screening strategies. Machine-learning accelerated methods achieved similar results to brute force docking. First-in-class, experimentally confirmed compounds were rare and weakly potent, indicating that recent advances are not sufficient to effectively address challenging targets.

The ribosome lowers the entropic penalty of protein folding

Julian O. Streit1*, Ivana V. Bukvin[1]*, Sammy H.S. Chan[1]*, Shahzad Bashir[1], Lauren F. Woodburn[1], Tomasz Włodarski[1], Angelo Miguel Figueiredo[1], Gabija Jurkeviciute[1], Haneesh K. Sidhu[1], Charity R. Hornby1, Christopher A. Waudby[1], Lisa D. Cabrita[1], Anaïs M.E. Cassaignau[1] & John Christodoulou[1◊]

[1]*Institute of Structural & Molecular Biology, Department of Structural & Molecular Biology, University College*

London, London WC1E 6BT, UK and Department of Biological Sciences, Birkbeck College, London WC1E 7HX, UK

**These authors contributed equally to this work.*

◊*To whom correspondence should be addressed: j.christodoulou@ucl.ac.uk*

Abstract

Most proteins fold during biosynthesis on the ribosome, and co-translational folding energetics, pathways, and outcomes of many proteins have been found to differ considerably from those in refolding studies. The origin of this folding modulation by the ribosome has remained elusive. Here, we have exploited long-timescale all-atom MD simulations in explicit solvent of ribosome-nascent chain complexes. These simulations were integrated with restraints from paramagnetic relaxation enhancement NMR experiments, allowing us to determine atomistic structures of the unfolded state of a model protein on and off the ribosome. The ensembles reveal that the ribosome structurally expands the unfolded nascent chain and increases its solvation, resulting in its entropic destabilisation relative to in isolation. Quantitative 19F NMR experiments confirm that this destabilisation reduces the entropic penalty of folding by up to 30 kcal mol-1 and promotes formation of partially folded intermediates on the ribosome, an observation that extends to other protein domains and is obligate for some proteins to acquire their active conformation. The thermodynamic effects also contribute to the ribosome protecting the nascent chain from mutation-induced unfolding, which suggests a crucial role of the ribosome in supporting protein evolution. By correlating nascent chain structure and dynamics to their folding energetics and post-translational outcomes, our findings establish the physical basis of the distinct thermodynamics of co-translational protein folding.

Structure-guided disulfide engineering restricts antibody conformation and flexibility to elicit TNFR agonism in anti-cancer therapeutics

Isabel Elliott[1,2,3], Hayden Fisher[1,2,3], Christian M. Orr[4], Claude H.T. Chan[1], Christine A. Penfold[1], Patrick J. Duriez[1], Tatyana Inzhelevskaya[1], C. Ian Mockridge[1], Ivo Tews[2,5], Mark S. Cragg[1,5], Jonathan W. Essex[3,5].

[1]*University of Southampton, School of Cancer Sciences, Centre for Cancer Immunology; Southampton, SO16 6YD, UK.*

[2]*University of Southampton, School of Biological Sciences; Southampton SO17 1BJ, UK.*

[3]*University of Southampton, School of Chemistry; Southampton SO17 1BJ, UK.*

[4]*Diamond Light Source, Didcot OX11 0FA, UK.*

[5]*University of Southampton, Institute for Life Sciences; Southampton SO17 1BJ, UK.*

Corresponding author: I.G.Elliott@soton.ac.uk

Abstract

Immunostimulatory antibodies (ISAs) represent a promising strategy for cancer immunotherapy. By activating co-stimulatory molecules expressed on immune cells, such as tumour necrosis factor receptors (TNFRs), ISAs can enhance the immune response towards tumours, resulting in powerful anti-cancer effects [1]. Antibodies comprise two antigen-binding domains linked to an effector domain through a disulfide-containing hinge. There are four isotypes of human (h)IgG, and previous work has shown that the hIgG2 isotype can deliver strong agonistic activity for ISAs, due to its unique hinge disulfide arrangement [2]. Here we employ an integrative approach to understand how structure and conformational dynamics affect agonistic activity of hIgG2 antibodies. We use cellular assays to ascertain agonistic activity, small angle X-ray scattering (SAXS) to assess flexibility and conformation, X-ray crystallography to determine protein structure and disulfide position, and molecular dynamics simulations with SAXS-guided ensemble reweighting to probe antibody conformational dynamics. By modifying hinge disulfide patterns using cysteine-to-serine exchange mutations, we show that strong agonistic activity is associated with restricted global antibody flexibility and reduced conformational dynamics due to the presence of a cross-over of disulfides in the hinge [3]. This has been shown for antibodies targeting two different co-stimulatory receptors, CD40 and 4-1BB. We then use structure-guided approaches to design new hIgG2 antibody variants with novel disulfide patterns, to further restrict flexibility and enhance biological activity. Together, these results demonstrate the importance of structure and conformational dynamics in ISAs and provide a strategy for the rational design of more powerful antibody therapeutics, and thus more effective anti-cancer treatments.

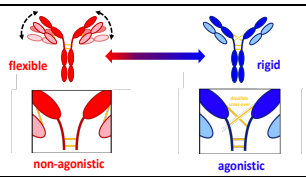


Figure 1: A crossover of disulfides in the hinge region of hIgG2 antibodies restricts conformational dynamics and flexibility, providing stronger receptor agonism and more powerful immunostimulatory activity to anti-cancer therapeutics.

References

[1] Mayes, P. A., Hance, K. W., and Hoos, A. (2018). *Nat. Rev. Drug Discov.* **17**, 509-207.

[2] White, A. L., Chan, H. T. C., French, R. R., Willoughby, J., Mockridge, C. I., Roghanian, A., Penfold, C. A., Booth, S. G., Dodhy, A., Polak, M. E., Potter, E. A., Ardern-Jones, M. R., Verbeek, J. S., Johnson, P. W. M., Al-Shamkhani, A., Cragg, M. S., Beers, S. A., and Glennie, M. J. (2015). *Cancer Cell.* **27**, 138-148.

[3] Orr, C. M., Fisher, H., Yu, X., Chan, C. H., Gao, Y., Duriez, P. J., Booth, S. G., Elliott, I., Inzhelevskaya, T., Mockridge, C. I., Penfold, C. A., Wagner, A., Glennie, M. J., White, A. L., Essex, J. W., Pearson, A. R., Cragg, M. S., and Tews, I. (2022). *Sci. Immunol.* **7**, eabm3723.

Improving continuous heterogeneity landscapes and interpretations using atomic structures: a case study of SARS-CoV-2 spike variants

James Krieger[1], Carlos Perez Mata, David Herreros, Wai Shing Tang, Xing Zhu, Carlos Oscar Sanchez Sorzano, Jose Maria Carazo, Pilar Cossio, Erik Thiede, Sriram Subramaniam
[1]*National Centre for Biotechnology, Madrid*

Abstract

There are now a large number of image processing methods capable of generating landscapes representing the continuous heterogeneity present in a Cryo-EM dataset along with volumes corresponding to particular landscape regions and particle images. These landscapes usually have many dimensions (e.g. 10 or 39) and it is common to visualize them by dimensionality reduction methods (such as PCA or UMAP) applied either directly to the multi-dimensional coordinates of the particles or to the output volumes, but the resulting 2D or 3D landscapes are still difficult to interpret and often vary between methods, executions and datasets.

We have, therefore, come up with a pipeline based around atomic models flexibly fitted to the volumes generated by continuous heterogeneity methods including both our deformation- and density-based approaches. This approach uses ProDy for PCA of the atomic positions after superposition and then projection of the atomic structures onto the principal components (PCs). The same PCA can be easily used for structures coming from different datasets as well as those already deposited in the PDB or derived from molecular simulations, enabling their comparison in a common space. As atomic PCA is a common approach in computational biophysics, the validity and interpretation of the coordinates is much clearer. Each axis corresponds to a mode of variation based on motion of the atoms away from the average structure that can be visualized in the normal mode wizard (NMWiz) in VMD. There is also a related method called logistic regression analysis (LRA), which can be used to generate modes of motion related to differences between different subsets of structures. We have also developed a connection with the Bayesian ensemble reweighting framework for comparing these atomic structures back to the particle images, enabling the full landscapes to represented in this space, providing a better idea of populations and free energies. These tools are implemented within plugins of the Scipion workflow engine, providing easy access to users downstream of continuous heterogeneity analyses.

We have extensively tested this pipeline on two datasets corresponding to the spike glycoprotein from two different SARS-CoV-2 variants, enabling us to draw conclusions about the overall variation within both datasets together as well as the differences between them that may underlie variant effects. This enabled us to compare these variations to published structures of these variants and better understand the effects of different experimental conditions such as stabilizing mutations.

From Active Learning to Zinc: adventures with alchemical free energy calculations

Antonia S. J. S. Mey[1]; Rohan Gorantla[1,2]; J. Jasmin Güven[1]
[1] *EaStCHEM School of Chemistry, University of Edinburgh, Edinburgh EH9 3FJ, U.K.*
[2] *School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, U.K.*

Corresponding author: Antonia.mey@ed.ac.uk

Abstract

Relative free energy (RFE) calculations are widely used in academia and industry to compute free energies of binding between ligands and proteins as accurately as possible. While alchemical methods do provide good accuracy for many systems some challenges remain. One challenge is how to effectively choose which ligands to screen with RFE calculations when the ligand pool for testing is larger than the computational budget allows. A second challenge involves ligands that interact with metals in the binding site. Both have unique sets of challenges.

I will highlight how we have used active learning (AL), an iterative process that allows for learning of labels of an unlabelled dataset, to identify which compounds to select next for RFE screening from a ligand pool. We benchmarked the influence of different surrogate models (Chemprop and Gaussian processes), AL strategies such as exploration and exploitation cycles, batch sizes, and the effect of noise on the selection efficiency of the data.

To address how we can better model ligands interacting with coordinating metal ions in metalloproteins and devise robust free energy protocols for such systems, I will share our ongoing efforts in this direction. Using different β-lactamases, divalent zinc metallo-β-lactamases, and serine β-lactamases, we can show that conventional RFE methods give good binding affinities with respect to experimental measures of known congeneric series. The same ligands showing activity against metallo-β-lactamases cannot be modelled to the same level of accuracy using naïve ligand and metal force field parameters as well as restraints for the effective coordination of the metal sites.

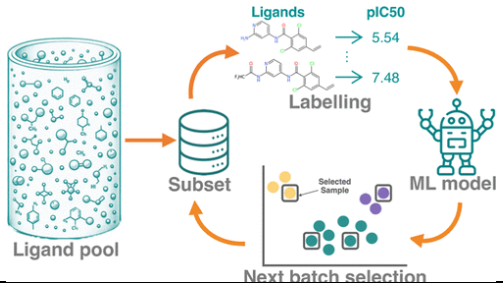


Figure 1: *Active learning allows for efficient selection of ligands in alchemical free energy calculations*

References
[1] Gorantla, Rohan, et al. (2024). *J. Chem. Inf. Model. 2024, 64, 6, 1955–1965.*
[2] E. Heid et al., *J. Chem. Inf. Model.*, 2023, 64, 9–17.

Atom Surface Site Interaction Mapping of Protein-Ligand Complexes

Katarzyna J. Zator[1]; Christopher A. Hunter[1]
[1]*Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, UK*

Corresponding author: kz265@cam.ac.uk

Abstract

Advances in theoretical chemistry allowed us to discover the equations that describe the electronic structure of the atoms, but also informed us that such accurate calculations are very expensive indeed. To look at larger systems, akin to protein-ligand complexes, simplifications need to be introduced which compromise the quality of result, and we ought to carefully consider errors that might henceforth arise. Furthermore, in a drug design campaign the aim is not merely in learning about a single system, but discovery of one with desirable properties, hence we need a reliable way of canvassing chemical space to find a ligand of interest. This process has been entrusted to scoring functions which evaluate the relative stability of complexes formed between ligands and proteins. We have used DFT to develop the Atom Surface Site Interaction Point (AIP) approach [1][2] to quantify all non-covalent interactions a molecule is capable of making with the environment, i.e. solvent or binding site. When two molecules interact, AIPs that are close in space are paired to calculate the free energy of binding by summing over all pairwise interactions. Each such interaction has an associated electrostatic and dispersion components. The method has been successfully applied to host-guest complexes in a variety of solvents by using SSIMPLE solvation energies [3]. We used the CASF-2016 dataset of diverse X-ray protein-ligand structures [4] to develop an AIP scoring function for use in drug discovery. The method is fast and produces interpretable and accurate binding free energies, as well as AIP Interaction Maps, as exemplified in Figure 1, that allow interpretation of the key contributions to binding. This knowledge can be used not only to judge promising ligand leads but to design better binding ligands overall.

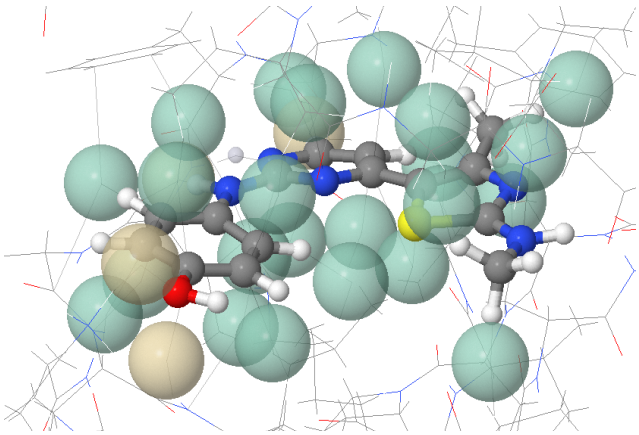


Figure 1: *AIP Interaction Map of protein-ligand interactions in the 1PXN complex: translucent green spheres represent attractive interactions, and yellow – repulsive.*

References

[1] Storer Maria Chiara, Zator Katarzyna J., Reynolds Derek P., and Hunter Christopher A. (2024). *Chem. Sci.*, **24**, 160-170.
[2] Hunter Christopher A. (2004) *Angewandte Chemie Intl. Ed.* **43** (40), 5310-5324.
[3] Hunter Christopher A. (2013) *Chem. Sci.*, **4**, 1687-1700.
[4] Su Minyi, *et.al.* (2019) J. *Chem. Inf. Model.* **59**, 895-913.

Computational Electrophysiology of Potassium and Chloride Channels

Wojciech Kopec[1,2,] Chenggong Hui[2], Reinier de Vries[2], Edward Mendez-Otalvaro[2], Ivan Manoza[1], Jakub Mróz[3], Artur Góra[3], Bert L. de Groot[2]

[1]*Department of Chemistry, Queen Mary University of London, 327 Mile End Road, E1 4NS London, UK*

[2]*Computational Biomolecular Dynamics Group, Max Planck Institute for Multidisciplinary Science, Am Fassberg 11, 37077 Göttingen, Germany*

[3]*Tunneling Group, Biotechnology Centre, ul. Krzywoustego 8, Silesian University of Technology, 44-100 Gliwice, Poland*

Ion channels are the third largest class of protein targets in drug discovery. Traditionally, ion channels have been studied by either structural biology or electrophysiology techniques. While both approaches are extremely powerful, a direct comparison of the "structural world" with the "functional" one is not always straightforward, due to resolution and time scales differences between these techniques. Here, I will present how Molecular Dynamics (MD) based simulations under applied voltage can be used as a bridging "Computational Electrophysiology" (CompEl) approach, to study ion permeation and gating at the atomistic scale, simultaneously enabling direct comparison with experimentally measured single channel currents. Progress in the force field development allowed us to record, for the first time, full I-V curves for several potassium channels, revealing the mechanism of inward rectification. We show how ion permeation in these channels, occurring via "full" direct knock-on mechanism, underlies their high permeation rates. Selectivity filter gating and its modulation by small molecules in the TREK-1 channel will be discussed as well. Finally, our work on the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) shows a bifurcated permeation pathway, with a hitherto unknown regulation mechanism. These examples of pharmacologically relevant ion channels suggest CompEl as an attractive technique in drug discovery of ion channel-targeting molecules.

Using dynamical-nonequilibrium MD simulations to understand drug resistance and allostery in proteins

A. Sofia F. Oliveira[1]
[1]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK*
Corresponding author: sofia.oliveira@bristol.ac.uk

Abstract

Proteins are neither static nor work in isolation in physiological conditions. In fact, it is the opposite; proteins are continuously moving and switching between many different conformations. Changes in the environment can shift the balance between this multitude of conformations, ultimately determining the protein's macroscopic behaviour. The binding of ions and small ligands and peptides, changes in voltage, pH, and temperature and light absorption by light-harvesting complexes are all good examples of external perturbations that can induce changes in protein structure and dynamics, thus affecting their function.

Several computational approaches have been developed over the years to study functional dynamics in proteins and understand how structural and dynamic changes shape function. Here, we will focus on the emerging dynamical non-equilibrium molecular dynamics (D-NEMD) simulations approach [1-3]. This method combines simulations in stationary (in particular, equilibrium) and non-equilibrium conditions and by doing so allows us to extract the evolution of the response of a protein when exposed to an external perturbation. This approach, which was initially proposed in the seventies, is currently undergoing a renaissance and having increasing impact on the study of biological systems.

Here, we will briefly discuss the essential features of the D-NEMD approach and how it can be used in a more general setting to study proteins. We also provide examples of different biomolecular systems and biological questions that can be addressed using this method. The examples, covering different proteins ranging from soluble enzymes to integral membrane receptors and the SARS-CoV2 spike demonstrate the versatility and general applicability of the D-NEMD approach and how it can provide a comprehensive and unbiased mapping of the structural responses and communication networks in proteins.

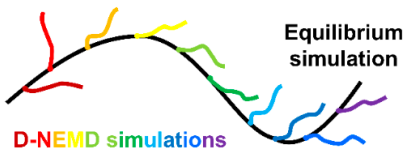


Figure 1: Schematic representation of the D-NEMD approach. The stationary (e.g., equilibrium) MD trajectory (black line) provides the initial distribution for the non-equilibrium simulations at time t=0. The individual non- equilibrium trajectories sample the dynamic evolution of the system after the introduction of a perturbation.

References

[1] Oliveira, A S F, Ciccotti, G, Haider, S and Mulholland, A J. (2021). *Eur. Phys. J. B*, **94** (144), 1-12
[2] Ciccotti, G, Jacucci, G and McDonald, I R. (1979). *J. Stat. Phys.*, **21** (1), 1-22
[3] Ciccotti, G and Jacucci, G. (1975). *Phys. Rev. Lett.*, **35** (12), 789-792

Targeting the conserved active site of splicing machines

Marco De Vivo

Laboratory of Molecular Modelling & Drug Discovery, Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genoa, Italy

In my contribution, I will illustrate our current efforts in targeting RNA, using small organic molecules. Specifically, I will explain our work on targeting the conserved active site of splicing machines. This work is based on our mechanistic investigations of the self-splicing group II introns, which are bacterial and organellar ancestors of the nuclear spliceosome. I will show how we integrate atomistic molecular simulations and alchemical free energy calculations with enzymatic and crystallographic studies to design potent splicing modulators. Furthermore, I will delve into the mechanism by which the evolutionarily conserved RNA core of splicing machinery exhibits specificity in recognizing small molecules. This capability paves the way for designing innovative splicing modulators potentially applicable to a broad spectrum of human diseases.

References

Targeting the conserved active site of splicing machines with specific and selective small molecule modulators
I. Silvestri, J. Manigrasso, A. Andreani, N. Brindani, M. De Vivo, M. Marcia https://doi.org/10.1101/2023.06.21.545906 - available in BioRxiv – 2024.

Automated Adaptive Absolute Binding Free Energy Calculations

Finlay Clark[1]; Graeme Robb[2]; Daniel J. Cole[3]; Julien Michel[1]
[1]*EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, United Kingdom.*
[2]*Oncology R&D, AstraZeneca, Cambridge CB4 0WG, United Kingdom.*
[3] *School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom.*

Corresponding author: finlay.clark@ed.ac.uk

Abstract

Alchemical absolute binding free energy (ABFE) calculations offer a rigorous method to calculate the binding affinities of varied molecules.[1] They can address problems such as optimising ligand selectivity or promiscuity, and predicting the functional response of ligand binding. They are also useful as a final, accurate filter in virtual screening. However, ABFE calculations are computationally expensive, demanding tens to hundreds of GPU hours for a single molecule or pair of molecules. To become a routine tool in industrial drug discovery, ABFE calculations must become more efficient and automated.

I will discuss algorithms to improve the efficiency of ABFE calculations by reducing both user and simulation time. In particular, I will discuss the automated selection of the number and spacing of simulations intermediate between the thermodynamic end states, the detection of equilibration, and the adaptive allocation of simulation time to achieve minimum variance of the final free energy estimate. I will address the automated selection of restraints between the receptor and ligand, which can significantly influence the performance of ABFE calculations.[2] Finally, I will discuss the implementation of these algorithms in an open-source software package, A3FE.[3] We hope that these approaches will reduce the barrier to the routine use of absolute binding free energy calculations in drug discovery.

References
[1] Cournia, Zoe, Allen, Bryce K., Beuming, Thijs, Pearlman, David A., Radak, Brian K. and Sherman, Woody. 2020. *J. Chem. Inf. Model.*, **60**, 4153–4169.
[2] Clark, Finlay, Robb, Graeme, Cole, Daniel J. and Michel, Julien. 2023. *J. Chem. Theory Comput.*, **19**, 3686–3704.
[3] Clark, Finlay. A3FE: Automated Adaptive Absolute alchemical Free Energy calculator. 2024. Available from: https://github.com/michellab/a3fe

Active Learning Driven Hit Mining and Optimization Based on Molecular Dynamics Simulated Free Energies

Maria G Kurnikova; Olexandr Isayev, Eugeny Gutkin, Phillip Gusev, Samuel Ben Koby, Chamali Narangoda

[1]*Chemistry Department, Carnegie Mellon University, Pittsburgh, PA 1513, USA*

Corresponding author: kurnikova@cmu.edu

Abstract

In silico identification of potent protein inhibitors requires prediction of a ligand binding free energy (BFE). Thermodynamics integration (TI) based on molecular dynamics (MD) simulations is a BFE calculation method capable of predicting accurate BFE, but it is computationally expensive and time-consuming. We have developed an efficient automated workflow for identifying compounds with the lowest BFE among thousands of congeneric ligands which requires only hundreds of TI calculations. Automated Machine Learning (AutoML) orchestrated by Active Learning (AL) in AL-AutoML workflow allows unbiased and efficient search for a small set of best performing molecules. We have applied this workflow to select inhibitors of the SARS-CoV-2 papain-like protease 1, as well as to predict hit and optimized hits for the Leucine-Rich Repeat Kinase 2 (LRRK2) WD40 Repeat (WDR) domain2, a Parkinson's disease target as a part of the Critical Assessment of Computational Hit-Finding experiments (CACHE) Challenge #1. We obtained hit rates that outperform a traditional expert medicinal chemist-guided campaigns, as well as other in silico schemes. Thus, we demonstrate that combination of AL and AutoML with free energy simulations provide at least 20x speedup relative to the naïve brute force approaches.

References

1. Gusev, F.; Gutkin, E.; Kurnikova, M. G.; Isayev, O., Active Learning Guided Drug Design Lead Optimization Based on Relative Binding Free Energy Modeling. Journal of Chemical Information and Modeling 2023, 63 (2), 583-594.
2. Gutkin, E.; Gusev, F.; Gentile, F.; Ban, F.; Koby, S. B.; Narangoda, C.; Isayev, O.; Cherkasov, A.; Kurnikova, M. G., In silico screening of LRRK2 WDR domain inhibitors using deep docking and free energy simulations. Chemical Science 2024.

Navigating protein landscapes with a machine-learned transferable coarse-grained model

Cecilia Clementi[1]
[1]*Freie Universität Berlin*

Abstract

The most popular and universally predictive protein simulation models employ all-atom molecular dynamics (MD), but they come at extreme computational cost. The development of a universal, computationally efficient coarse-grained (CG) model with similar prediction performance has been a long-standing challenge. By combining recent deep learning methods with a large and diverse training set of all-atom protein simulations, we have developed a bottom-up CG force field with chemical transferability, which can be used for extrapolative molecular dynamics on new sequences not used during model parametrization. We have demonstrated that the model successfully predicts folded structures, intermediates, metastable folded and unfolded basins, and the fluctuations of intrinsically disordered proteins while it is several orders of magnitude faster than an all-atom model. This showcases the feasibility of a universal and computationally efficient machine-learned CG model for proteins.

Electrostatic embedding of machine learned potentials for accurate and efficient simulation of enzyme catalysis

Elliot Chan[1], Kirill Zinojev[2], Lester Hedges[3], Christopher Woods[3], Marc van der Kamp[1]
[1] *School of Biochemistry, University of Bristol*
[2] *Departamento de Química Física, Universidad de Valencia*
[3] *Research Software Engineering, Advanced Computing Research Centre, University of Bristol,*

Corresponding author: marc.vanderkamp@bristol.ac.uk

Abstract

To simulate reactions in large biomolecular systems, Quantum Mechanics/Molecular Mechanics (QM/MM) is typically used, treating just the region where the reaction occurs with QM and the rest with MM. Whilst this is the state-of-the-art for simulating enzyme reactions, the speed is primarily determined by the QM method used, drastically limiting accessible timescales and sampling.

An emerging solution is to employ machine learning potentials (MLPs) that predict energies of atomic arrangements at a fraction of time needed for QM. Pure MLPs would still be slower than MM, but "ML/MM" simulations would offer significant speed-up over QM/MM. One cannot simply replace QM with an MLP: electrostatic embedding, crucial for capturing catalytic effects, would not work. Due to the absence of electronic density in MLPs, the influence of MM partial charges cannot be taken into account.

Here, we present the use of an electrostatic machine learning embedding scheme (EMLE)[1], in which generic in vacuo MLPs are implemented in an efficient ML/MM simulation[2], with EMLE predicting the response of the ML region to the MM region. We test EMLE on simulations of the Diels-Alder reaction catalysed by the natural Diels-Alderase AbyU. First, an Atomic Cluster Expansion (ACE) MLP was trained on the reaction in the gas phase, learning energies based on the M06-2X functional. Then, using this MLP in ML/MM simulations with EMLE, umbrella sampling simulations of the reaction were performed and compared to QM/MM simulations of the same reaction using M06-2X. We find that the QM/MM free energy profile is reproduced accurately by our ML/MM(EMLE) model, at a significant speed up (up to 285 times). Further, tests on different enzyme-substrate poses reveals that 'mechanical embedding' (where fixed point charges are used for the MLP to provided electrostatic interactions with the MM region) is not able to distinguish between the reactive pose and alternative non-reactive poses that were previously found in AbyU[3]. Therefore, we demonstrate that using EMLE in efficient ML/MM reaction modelling is crucial to capture enzyme catalysis accurately.

References

[1] Zinojev, K. (2023). *J. Chem. Theory Comput*, **19**(6), 1888-1897. DOI:/10.1021/acs.jctc.2c00914
[2] Zinojev, K, Hedges, L, Andreu, RM, Woods, C, Tuñón, I, van der Kamp, MW. (2023). *ChemRxiv*, DOI:10.26434/chemrxiv-2023-6rng3
[3]Maschio, L, Back, CR, Alnawah, J, Bowen, JI, Johns, ST, Mbatha, SZ, Han, L, Lees, NR, Zorn, J, Stach, JEM, Hayes, MA, van der Kamp, MW, Pudney, CR, Burston, SG, Willis, CL, Race, PR. (2024). *BioRxiv, DOI:*10.1101/2024.04.18.590041

Stable and Accurate Atomistic Simulations of Drug Molecules using Conformationally Generalisable Machine Learned Potentials

Christopher D. Williams[1]; Jas Kalayan[2], Neil A. Burton[3], Richard A. Bryce[1]
[1]*Division of Pharmacy and Optometry, School of Health Sciences, The University of Manchester, Manchester, UK.*
[2]*Science and Technologies Facilities Council, Warrington, UK.*
[3]*Department of Chemistry, School of Natural Sciences, The University of Manchester, Manchester, UK.*

Corresponding author: christopher.williams@manchester.ac.uk

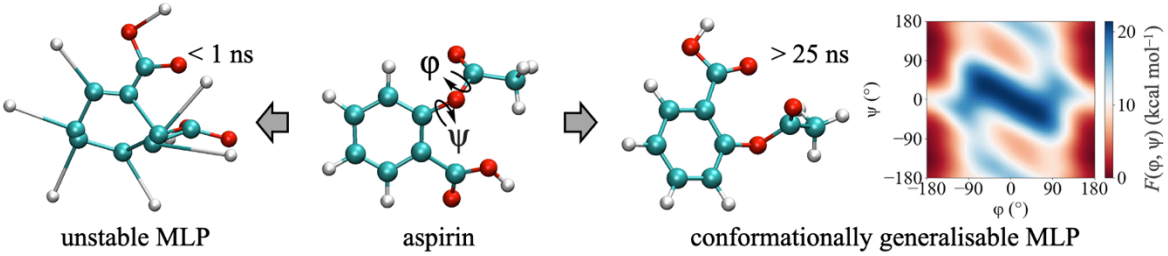
Abstract

Machine-learned interatomic potentials (MLPs) promise a transformative impact in computational drug design, which is a field hampered by the respective efficiency and accuracy limitations inherent to quantum mechanics and molecular mechanics techniques.

In this talk, we firstly show that feed-forward deep neural networks can be used to train molecule-specific MLPs according to the newly developed PairFE-Net scheme. PairFE-Net MLPs accurately predict forces and energies, competitive with the state-of-the-art, with MAEs << 1 kcal mol^{-1} (Å^{-1}). However, unlike most other schemes PairFE-Net uses global input features, ensuring a complete representation of the potential energy surface and an ability to account for non-local interactions, which are often crucial for correctly predicting the complex conformational landscapes of drug molecules.

Although much effort has been expended on obtaining MLPs with the primary objective of obtaining very low test errors on benchmark datasets, we will show that this approach can often be misleading and frequently produces MLPs that are dynamically unstable, producing unphysical structures and catastrophic instabilities over the course long molecular dynamics (MD) simulations. We demonstrate that this arises when a conformational change occurs and that the problem is particularly acute when MLPs are trained on some of the most popular literature benchmark datasets (e.g. rMD17), which are missing certain key conformers.

To address this short-coming we generated new reference datasets for several drug molecules, with torsional motions enhanced using metadynamics. Crucially, MLPs trained on these complete datasets are stable over long MD simulations and across conformation space, including barrier regions. This quality has enabled drug molecule conformational free energy surfaces to be accurately calculated, a task beyond the timescales accessible to quantum mechanical methods. In combining quantum mechanical accuracy with a computational cost close to molecular mechanics methods, PairFE-Net is a vital new tool in ligand-based drug design, enabling the conformational preferences of drug molecules to be rapidly evaluated.



References
CD Williams, J Kalayan, NA Burton and RA Bryce, 2024, ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv-2024-r75jz.

Enhancing Protein-Ligand Binding Affinity Predictions using Neural Network Potentials

Francesc Sabanés[1]; Gianni de Fabriitis[1]
[1]*Acellera Therapeutics*

Corresponding author. f.sabanes@acellera.com

Abstract

At Acellera Therapeutics we are developing several tools to combine AI-based methods with structural based drug discovery. Recently, we presented QuantumBind®, our new platform for automated small-molecule potency optimization via binding free energy calculations and Machine Learning (ML). With the combination of neural network potentials (NNPs) with relative binding free energy (RBFE) calculations we have been able to increase the accuracy of these predictions in comparison to other conventional force fields such as GAFF2.[1] We combine these results with active learning protocols that in combination with our state-of-the-art generative models [2] help us the best small molecules for experimental testing. We have recently applied the QuantumBind® platform to a target protein from our pipeline which helped us optimize the previously found hit in a short amount of time.

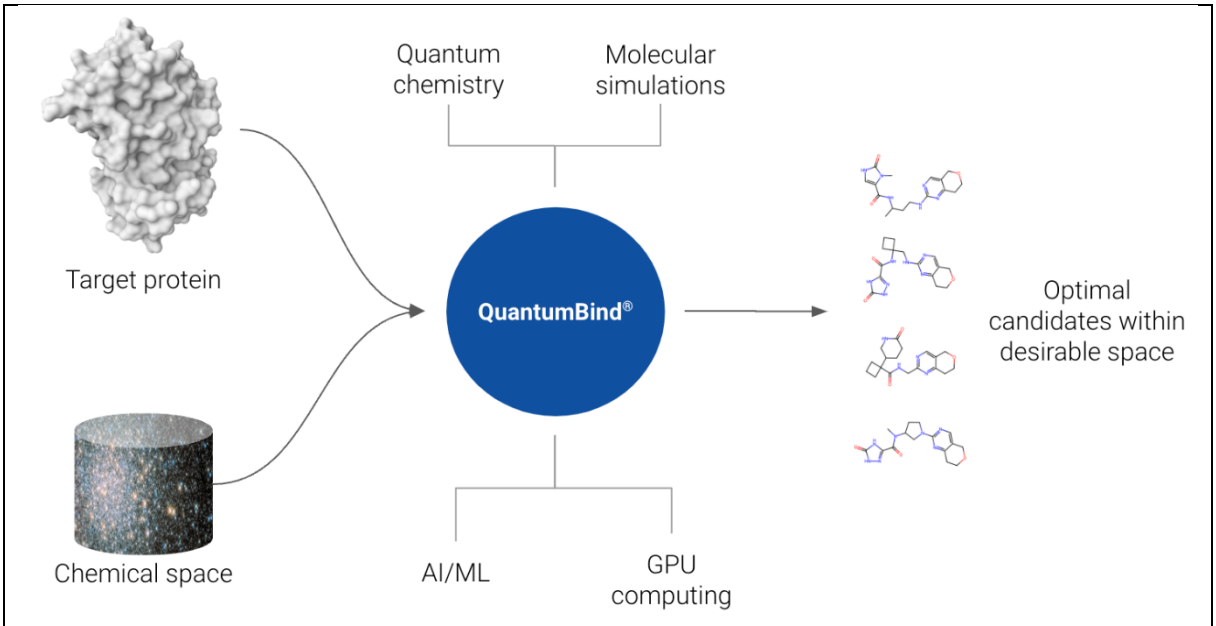


Figure 1: *Text*

References
[1] Sabanés, Francesc, Galvelis, Raimondas, et al, Enhancing Protein–Ligand Binding Affinity Predictions Using Neural Network Potentials (2024). *Journal Chem. Inf. Model*, 64(5), 1481-1485.
[2] Bou, A., Thomas, M., Dittert, S., "ACEGEN: Reinforcement Learning of Generative Chemical Agents for Drug Discovery." GEM workshop, ICLR 2024.

The State of Protein Structure Prediction and Friends

Mohammed AlQuraishi[1]

[1]*Columbia University*

Abstract:

AlphaFold2 revolutionized structural biology by accurately predicting protein structures from sequence. Its implementation however (i) lacks the code and data required to train models for new tasks, such as predicting alternate protein conformations or antibody structures, (ii) is unoptimized for commercially available computing hardware, making large- scale prediction campaigns impractical, and (iii) remains poorly understood with respect to how training data and regimen influence accuracy. Here we report OpenFold, an optimized and trainable version of AlphaFold2. We train OpenFold from scratch and demonstrate that it fully reproduces AlphaFold2's accuracy. By analyzing OpenFold training, we find new relationships between data size/diversity and prediction accuracy and gain insights into how OpenFold learns to fold proteins during its training process.

Real world comparison of dynamic ultra large library screening with commercial library screening against different target classes

Jonathan Heal[1]
[1]*RxCelerate*

Abstract:

A new tool RxNfinity is described which can hierarchically dock synthon libraries to target structures and adapt the chemical space during the search using a monte-carlo based scheme. This enables an effective structural context search in the billions to trillions of compound range. RxNfinity has been reduced practice on different targets and conventional virtual screens for novel compound identification were also conducted for comparison. Initial results generally demonstrate that ultra large library approaches yield higher hit rates and potencies but that conventional virtual screening is a useful complementary approach when considering scaffold diversity and hit rates.

Application of AI Augmented Design in Early Stage Drug Discovery

Will Gerrard[1]; Alex Punter[1], Raphael Chantreau Majerus[1], Jarryl D'Oyley[1], Georg Schusterisch[1], Charlie Adams[1], Anthony Nash[1], Parminder Ruprah[1], Vid Stojevic[1], David W. Wright[1]
[1]*Kuano, Hauxton House, Mill Scitech Park, Mill Lane, Cambridge, England, CB22 5HX*

Corresponding author: will.gerrard@kuano.ai

Abstract

Huge breakthroughs have been made in the design of methods for applying AI to drug discovery, however very few of these methods are actually used in industry, with most drugs still being discovered via high throughput screening [1]. Huge challenges remain in getting these new techniques and methods into real-world drug discovery pipelines. Kuano is an emerging biotech startup that has demonstrated how some of these challenges can be overcome, using expert medicinal chemists augmented by the latest AI tools.Kuano utilises some of the most popular recent advances in AI for small molecule drug development, applying them to enable our medicinal chemists to solve challenging design problems. Our approach is based on transition state drug design, and focuses on finding high quality, specific inhibitors for difficult or less-explored enzyme targets. Our pipeline relies on finding and developing a range of flexible AI tools informed by chemical knowledge, and the rapid iteration of design strategy between the developers and medicinal chemists.
Pipeline overview:

1. QMMM / MMML reaction simulations to identify ideal drug templates.
2. Development of bespoke similarity scorers based on interactions, electrostatics and pharmacophore features, driven by generative AI with broad exploratory power. This is done iteratively with feedback to and from expert medicinal chemists.
3. Tuning of similarity scorers into an optimisation function for obtaining candidate molecules through reaction-driven and local-optimisation-focused models
4. Wet lab selection guided by expert medicinal chemists informed by data from ML tools, classical scoring functions and explainable AI.

We have applied our approach to efficiently identify inhibitors of NOTUM and DNMT1, targets implicated in colorectal and pancreatic cancer respectively. For NOTUM, the pipeline designed a novel chemical series with nanomolar potency in biochemical assays, high ligand efficiency and was discovered by testing fewer than 30 compounds, an extremely high hit rate compared to alternative virtual screening campaigns. Our compounds show similar activity in disease relevant cell lines to toxic chemotherapy agents as well as excellent selectivity.

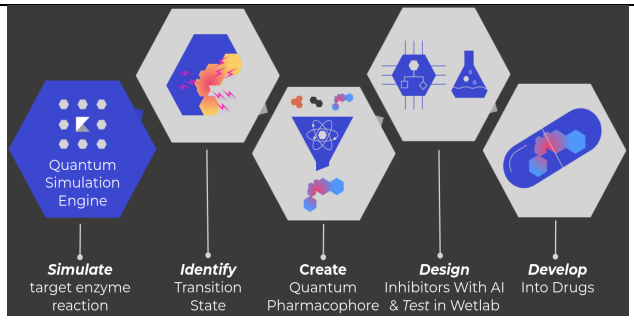


Figure 1: *Kuano Drug Design Pipeline*

References
[1] "AI is a viable alternative to high throughput screening: a 318-target study." *Scientific Reports* 14, no. 1 (2024): 7526.

AI-based identification of therapeutic agents targeting GPCRs: introducing ligand type classifiers and systems biology

Giulia Rossetti[1,2,3]

[1]*Institute of Neuroscience and Medicine (INM-9)/Institute for Advanced Simulation (IAS-5), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52425 Jülich, Germany*

[2]*Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52425 Jülich, Germany* [3]*Department of Neurology, University Hospital Aachen, RWTH Aachen University, Pauwelsstraße 30, 52074 Aachen*

Identifying ligands targeting G protein coupled receptors (GPCRs) with novel chemotypes other than the physiological ligands is a challenge for *in silico* screening campaigns. Here we present an approach that identifies novel chemotype ligands by combining structural data with an artificial inteligence - based classifier and a signal-transduction kinetic model. As a test case, we apply this approach to identify novel antagonists of the human adenosine transmembrane receptor type 2A, an attractive target against Parkinson's disease and cancer. The identified antagonists were tested here in a radio ligand binding assay. Among those, we found a promising ligand whose chemotype differs significantly from all so-far reported antagonists, with a binding affinity of 310 ± 23.4 nM. Thus, our protocol emerges as a powerful approach to identify promising ligand candidates with novel chemotypes while preserving antagonistic potential and affinity in the nanomolar range

References

10.1039/D3SC02352D (Edge Article) Chem. Sci. 2023, 14, 8651-8661.

Using SeeSAR to identify novel therapeutic starting points for Friedreich's Ataxia

Richard Heath[1], Ian Hope[1], Lanyu Wang[2], Rob Dawber[2], Tom McAllister[2]
Akane Kawamura[2], Agnieszka Bronowski[2], Wyatt W. Yue[1*]
[1]*Biosciences Institute, Catherine Cookson Building, Newcastle University Medical School, Framlington Place, Newcastle-upon-Tyne, NE2 4HH, UK*
[2]*School of Natural and Environmental Sciences, Bedson Building, Kings Road, Newcastle University, Newcastle-upon-Tyne, NE1 7RU, UK*

Corresponding author: Wyatt.Yue@newcastle.ac.uk

Abstract

Friedreich's Ataxia (FRDA) is an inherited neurodegenerative disorder caused by deficiency of Frataxin (FXN). FXN maintains cellular iron homeostasis by modulating the mitochondrial iron-sulphur cluster (ISC) assembly complex. This complex consists of a scaffolding subunit ISCU, a cysteine desulferase subunit NFS1, and accessory proteins. Our cryo-electron microscopy structure of the complex demonstrates that FXN binds at the interface between NFS1 and ISCU, bringing together catalytic loops to assist in the transfer of persulfide from NFS1 to ISCU for ISC synthesis. We demonstrated that peptidomimetics containing two FXN β-strands that surround Trp155, the essential residue at the FXN-NFS1-ISCU interface, are sufficient for binding and modulation of catalytic activity, suggesting that a frataxin mimic could bypass the cell's requirement of frataxin, and represent a potential therapeutic avenue for FRDA.

Utilising our structural data, we aim to identify novel druggable pockets at the NFS1-ISCU interface where FXN bind and carry out *in silico* screening to identify chemical starting points for frataxin mimics. The docking suite *SeeSAR* has revealed three pockets of potential for docking, including the FXN interface, a pocket neighbouring the NFS1 catalytic loop, and an allosteric pocket close to the ISCU site of activity. Our overall goal is to identify and develop small molecule/peptide binders to these pockets that stabilise the loops and improve catalytic activity.

As pilot, virtual screening of the Enamine fragment library (n=320) has yielded 16 fragment building blocks clustered in the pocket neighbouring the NSF1 loop. These compounds were validated using surface plasmon resonance, which demonstrated medium-to-high μM binding. We are in the process of expanding the chemical hit space through screening the dsi poised library (n=860), covalent library (n=11760) and peptidomimetic library (n=3600). Altogether, we have identified a druggable site within the ISC assembly complex ready for hit discovery and hit-to-lead optimisation of frataxin mimics.

Keywords: frataxin, drug discovery, iron-sulphur cluster assembly, virtual screening

Computational tools for the rational design of molecular glues

Jordi Juárez-Jiménez[1,2]; Marina Miñarro-Lleonar[2], Andrea Bertran-Mostazo[2], Jorge Duro[2] Xavier Barril,[1,2]

[1]*Universitat de Barcelona, Institut de Química Teòrica i Computacional, Barcelona, Spain.*

[2]*Universitat de Barcelona, Departament de Farmàcia i Tecnologia Farmacèutica, i Fisicoquímica Barcelona, Spain.*

Corresponding author: jordi.juarez@ub.edu

Abstract

Targeted protein degradation (TPD) is a very promising strategy to modulate protein activities in several diseases, spearheaded by anti-myeloma drugs lenalidomide and pomalidomide. It has been recently demonstrated that the mechanism of action of these drugs involves the increased degradation of several proteins. In particular, lenalidomide is able to stabilize the complex between the E3 ligase Cereblon (CRL4CRBN) and the transcription factors Ikaros and Aiolos as well as the enzyme Caseine Kinase 1a (CK1a), even though the stability of the native protein-protein interactions is very low. The structures of these complexes shed light into the molecular glue character of lenalidomide, although they did not reveal evident interactions that could account for the high efficiency of the ternary complex formation. In this work, we have leveraged Steered Molecular Dynamics to shed light into the molecular determinants underlying the stabilization effect exerted by lenalidomide in the complex between CRL4CRBN and CK1a. Furthermore, we evaluated the effect that different mutations of CK1a on the stability of the ternary complex CRL4CRBN- Lenalidomide-CK1a and provide a thermodynamic and kinetic rational for the stabilization of the PPI. These results pave the way to further understand cooperativity effects in drug-induced protein-protein complexes and could help in the future design of improved molecular glues.

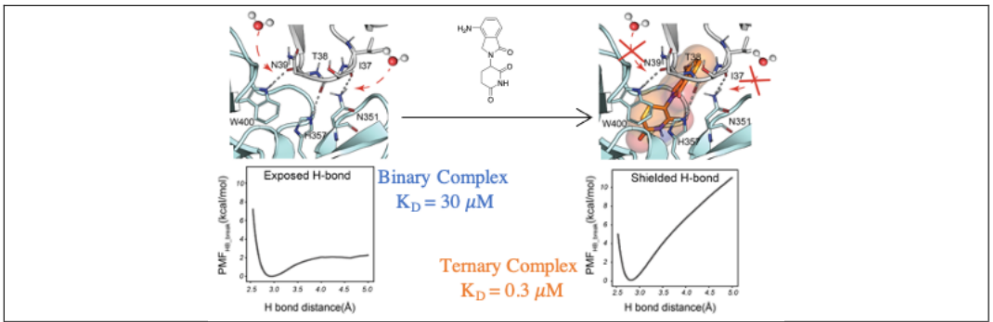


Figure 1: *The presence of lenalidomide stabilizes the CRBN-CK1α complex by reinforcing intermolecular H-Bonds.*

References

[1] Marina Miñarro-Lleonar, Andrea Bertran-Mostazo et. al. (2023). *J. Med. Chem*, **66**(9), 6037- 6046.

Modulation of Class B1 GPCRs by the plasma membrane environment

Kin Chao[1,3], George Hedger[1], Linda Wong[2], Affiong Oqua[2], Yusman Manchanda[2], Jas Kalayan[3], James Gebbie-Rayet[3], Alejandra Tomas[2], Sarah L Rouse[1]
[1]*Imperial College London, Life Sciences, London, United Kingdom*
[2]*Imperial College London, Metabolism, Digestion and Reproduction, London, United Kingdom*
[3]*Science and Technology Facilities Council, Scientific Computing, Warrington, United Kingdom*

Corresponding author: s.rouse@imperial.ac.uk

Abstract

Class B1 GPCRs, consisting of 15 receptors, are crucial drug targets for diseases such as obesity and type 2 diabetes. Current peptide drugs targeting these receptors have limitations, prompting the need for novel small molecule drugs, especially allosteric modulators with greater potential to fine-tune receptor activity. GPCR-lipid binding sites present a large, unexplored opportunity for allosteric drug development, yet our understanding of lipid modulation of Class B1 GPCRs remains limited.

We present a comprehensive coarse-grained molecular dynamics (cgMD) study of all 15 B1 family members in model mammalian plasma membranes, allowing us to determine patterns of specific lipid interactions across this subfamily. We used aiida-gromacs [1] to generate the full workflow to reproduce our simulations and encourage others to extend our simulation dataset. We also present a proof-of-principle case study of identifying cholesterol binding sites in GLP-1R using cgMD simulation, which informed the choices of a series of GLP-1R cholesterol-binding residues for further in-depth mutational analysis *in vitro*. In addition, we observe that GM3 plays a modulatory role in the dynamics of the extracellular domain (ECD), in a state-dependent manner. We further characterise *in vitro* the impact of GM3 on conformational dynamics of ECD in GLP-1R and GIPR using a TR-FRET GM3 inhibitor Eliglustat based assay.

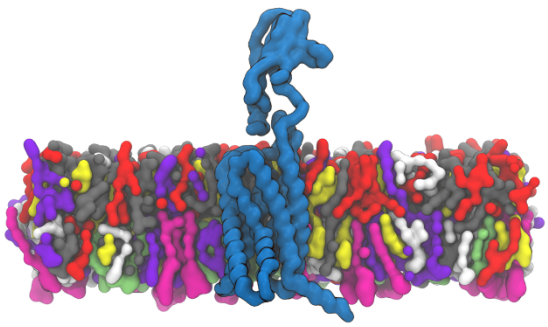


Figure 1: *GPCR in a asymmetric model mammalian plasma membrane with nine different lipid types. Each lipid is coloured differently.*

References

[1] https://aiida-gromacs.readthedocs.io

Prioritizing molecules for synthesis using in-silico tools

Joe Bluck[1]
[1]*Bayer*

Abstract:

Within the fast environment of active drug discovery projects there are often more compounds proposed than can feasibly be synthesized. Therefore, prioritizing molecules for synthesis is a key role of computational methods. Prioritization can be driven by modelled pharmacokinetic and physicochemical endpoints, but often on-target potency is the key determinant. Multiple tools exist for ranking molecular potency for a given target, from the cheap and popular molecular docking methods to more computationally expensive molecular-dynamics (MD)-based methods. Here we discuss how we embarked on a comparative study on multiple methods, their use in the drug discovery pipeline and whether the accuracy of the more rigorous methods justifies the higher computational cost and associated calculation time.

ASAP-Alchemy: A state-of-the-art open-source alchemical-free energy calculation pipeline with bespoke force fields

Joshua Horton[1], Jenke Scheen[2], Hugo MacDermott-Opeskin[2], David Dotson[3], Daniel Cole[1], John Chodera[3]

[1] *Newcastle University*

[2] *Open Molecular Software Foundation, 3 Sloan Kettering Institute New York*

Corresponding author: Josh.Horton@newcastle.ac.uk

Abstract

State-of-the-art computational free energy calculations offer a rigorous estimate of the binding affinity between a lead compound and its therapeutic target. Such calculations have become key in modern drug discovery and enable more efficient use of synthetic medicinal chemistry resources. While there exists a wide range of software tools that cover all aspects of the typical execution of a free energy calculation, to date there is no open-source end-to-end pipeline which unifies these packages into a single reproducible, customisable and extensible workflow suitable for large-scale drug discovery programs.

Here we present ASAP-Alchemy, a modelling pipeline that aids in the setup, bespoke force field parameterisation, execution and analysis of free energy calculations at scale and is created with reproducibility in mind. ASAP-Alchemy has a modular workflow composed of state-of-the-art open-source software, such as FEGrow[1], OpenFE, OpenFF[2], BespokeFit[3], Cinnabar and Alchemiscale, which can be customised to meet the specific needs of a project or trivially extended to interface with new software. The pipeline exposes a command line interface for each part of the workflow to simplify its use for routine high-throughput applications while reducing the barrier to entry for non-expert practitioners. ASAP-Alchemy has optional interfaces with modern drug discovery tools, like Manifold from Postera and the CDD vault, thereby enabling automated comparisons with experimental data and the ability to publish results on a project workspace. To aid the dissemination of the predicted results amongst interdisciplinary teams, interactive dashboard-style reports are provided to quickly rank and select the most promising compounds whilst giving an overview of the calculation accuracy where relevant experimental data is available.

We demonstrate the ease-of-use of ASAP-Alchemy by performing free energy calculations for a benchmark dataset of hundreds of ligand transformations, with bespoke force field parameters derived from machine learning predicted reference energy surfaces.

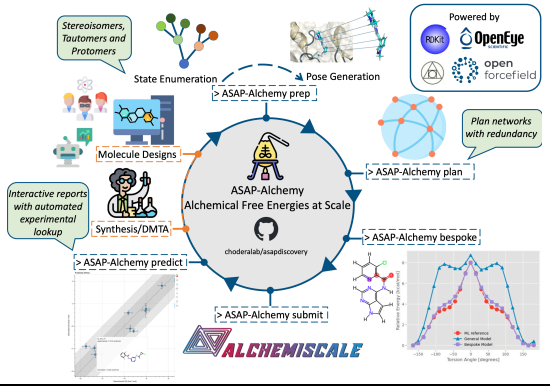


Figure 1: *ASAP-Alchemy makes it trivial to apply state-of-the-art binding free energy calculations at scale via its automated end-to end workflow.*

References

[1] Bieniek, M.K., Cree, B., Pirie, R., Horton, J.T., Tatum, N.J. and Cole, D.J., 2022. An open-source molecular builder and free energy preparation workflow. Communications chemistry, 5(1), p.136.

[2] Boothroyd, S., Behara, P.K., Madin, O.C., Hahn, D.F., Jang, H., Gapsys, V., Wagner, J.R., Horton, J.T., Dotson, D.L., Thompson, M.W. and Maat, J., 2023. Development and Benchmarking of Open Force Field 2.0. 0: The Sage Small Molecule Force Field. Journal of Chemical Theory and Computation.

[3] Horton, J.T., Boothroyd, S., Wagner, J., Mitchell, J.A., Gokey, T., Dotson, D.L., Behara, P.K., Ramaswamy, V.K., Mackey, M., Chodera, J.D. and Anwar, J., 2022. Open Force Field BespokeFit: Automating Bespoke Torsion Parametrization At Scale. Journal of Chemical Information and Modeling, 62(22), pp.5622-5633.

MD simulations map the havoc wreaked by one mutation in [Hsp60]$_{14}$

Stefano A. Serapian[1]; Luca Torielli[1]; Federica Guarra[1]; Giorgio Colombo[1]
[1]*Department of Chemistry, University of Pavia, Via T. Taramelli 12, 27100 Pavia, Italy*

Corresponding author: stefanoartin.serapian@unipv.it

Abstract

Chaperone proteins Hsp60 and Hsp10 cooperatively fold other "client" proteins into their biologically functional form.[1] The Hsp60 monomer **M** features 3 domains:[2] apical, middle, and equatorial (the latter contains the active site; Fig. 1a). Units of **M** self-assemble into a heptameric single ring **S** which, upon binding of ATP, Mg^{2+}, and K^{+}, pairs equatorially to form a 14-meric double ring **D**, whose lumen accommodates unfolded clients. Recruitment of 14 Hsp10 units by **D** apical domains (7 *per* pole) seals clients inside a resulting football-shaped complex **F** (Fig. 1b).[2] ATP cleavage breaks **F** apart, driving client folding and release. A hallmark of certain spastic paraplegia forms,[1] mutation V72I in **M** (Fig. 1a) disrupts this cycle: client misfolding is confirmed *in vitro*,[1] but with unidentified molecular causes. We here report findings from molecular dynamics simulations (32 μs in total) of solvated wild-type and V72I variants of **M** and **S** (*apo*), and **D** and **F** (ATP-bound and with 2 active site aspartates—Fig. 1a—modelled in 3 protonation states). V72I is found to pervasively disrupt allostery: the shortest path map[3] and distance fluctuation analyses[4] indicate communicational shutdown between equatorial and apical domains. Principal component analysis reveals **M** rigidification and reactivity alterations. Our results identify molecularly targetable hotspots whence to counter the effects of V72I.

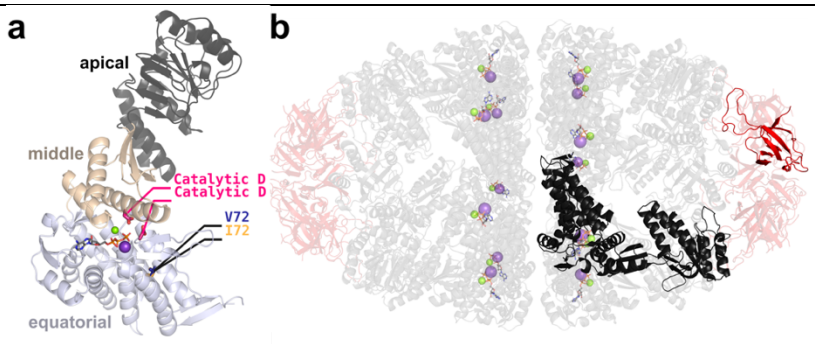


Figure 1: *(a) View of **M** (Hsp60 monomer) with domains and key residues labelled (ATP: sticks; K^{+}: purple; Mg^{2+}: green). (b) Overview of **F** highlighting one Hsp10 unit (red) and one **M** (black).*

References
[1] Chen, Lingling, Syed, Aiza and Balaji, Adhitya. (2022). *Sci. Rep.*, **12**(1), 18321.
[2] Braxton, Julian R., Shao, Hao, Tse, Eric, Gestwicki, Jason E. and Southworth, Daniel R. (2023). *bioRxiv*, 2023.05.15.540872.
[3] Osuna, Sílvia. (2021). *WIREs Computational Molecular Science*, **11**(3), e1502.
[4] Castelli, Matteo, Bhattacharya, Kaushik, Abboud, Ernest, Serapian, Stefano A., Picard, Didier and Colombo, Giorgio. (2023). *J. Mol. Biol.*, **435**(3), 167931.

Unravelling the Energetics of a Small Molecule-Induced Disorder-to-Order Transition

Salomé Llabrés[1,2]; Cesar Mendoza-Martinez[3], Michail Papadourakis[3], Arun Gupta[3], Paul N Barlow[3], Julien Michel[3]

[1]*Departament de Farmàcia, Tecnologia Farmacèutica i Físicoquímica, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Spain*

[2]*Institut de Química Teòrica i Computacional (IQTCUB), Spain*

[3] *EaStCHEM School of Chemistry, University of Edinburgh, United Kingdom*

Corresponding author: salome.llabres@ub.edu

Abstract

Many proteins recognize other proteins through mechanisms that involve the folding of intrinsically disordered regions upon complex formation [1]. We have explored how the selectivity of a drug-like small molecule arises from its modulation of a protein disorder-to-order transition [2]. The compound AM-7209 induces order in the intrinsically disordered 'lid' region of the oncoprotein MDM2. Calorimetric measurements revealed that truncating the lid region of MDM2 increases the apparent dissociation constant of AM-7209 by 250-fold. This effect is not reported for other families of MDM2 inhibitors. We have performed a comprehensive thermodynamic analysis, including a combination of adaptive absolute alchemical free energy of binding calculations and enhanced-sampling molecular dynamics simulations, to obtain insights into the differential binding energetics of MDM2 inhibitors. Our results show that in the absence of a ligand, the ordered lid state of MDM2 is energetically unfavorable. AM-7209 has a significant energetic preference for ordered lid conformations, thus promoting the ordering of the lid in the AM-7209/MDM2 complex. We have also applied this method to identify the molecular driving forces underpinning the selective ordering of the 'lid' region upon binding of AM-7209 [3]. The methodology presented here should facilitate medicinal chemistry efforts targeting of intrinsically disordered regions.

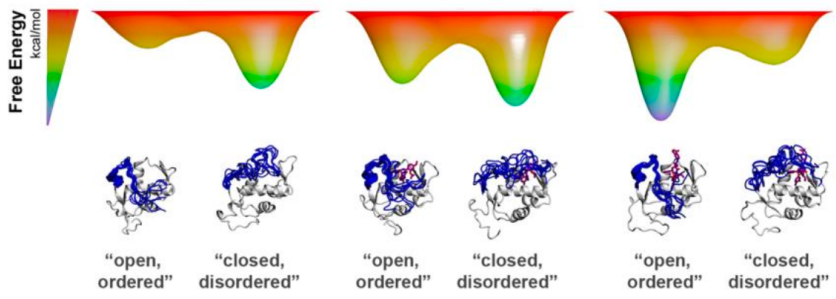


Figure 1: *Mechanism of ligand-dependent disorder–order transition of the MDM2 lid IDR.*

References

[1] Holehouse, Alex S, and Kragelund, BirtheB. (2024). *Nat Rev Mol Cell Biol*, **25**, 187-211.
[2] Mendoza-Martinez, Cesar; Papadourakis, Michail; Llabrés, Salomé; Gupta, Arun; Barlow, Paul N and Michel, Julien. (2022). *Chem. Sci*, **13**, 5220-5229.
[3] Mendoza-Martinez, Cesar; Gupta, Arun; Llabrés, Salomé; Barlow, Paul N and Michel, Julien. (2024).*ChemRxiv*, chemrxiv-2024-m15qm-v2

Exploring Electrostatic Properties Generated by DFT Functionals

Charlie Adams[1]; Daniel J. Cole[1], Josh Horton[1], David Wright[2]
[1] *Newcastle University, NE1 7RU*
[2] *Kuano, Hauxton House, Mill Scitech Park, Mill Lane, Cambridge, England, CB22 5HX*

Corresponding author: daniel.cole@ncl.ac.uk

Abstract

Quantum chemistry (QC) datasets form the backbone of many new machine learning (ML) partial charge/electrostatics models [1]. Legacy charge models like AM1-BCC were parameterized with the HF/6-31G* method due to having some fortunate over-polarisation, which is desirable for modelling organic compounds in the condensed phase. Nevertheless, in general the HF method is a poor choice for producing accurate electronic results due to its neglect of correlation effects. The choice of basis set has also shown to strongly influence the accuracy of the dipole moment, regardless of the method employed

In order to improve on the accuracy of the method whilst also using a method which is partially polarised, a combination of a density functional theory (DFT) functional and an implicit solvent model can be used [2]. This study will explore the influence of the DFT method, basis set, and implicit solvent model on electrostatic properties using the psi4 QC package, QCFractal, and in-house software. The accuracy of each method will be evaluated by comparing the partial charges, dipole moments, and electrostatic potential surfaces to a CCSD ground truth. Additionally, the effect of polarising charges with different dielectric constants will be explored in relation to HF derived properties.

References

[1] Y. Wang et al., 'EspalomaCharge: Machine learning-enabled ultra-fast partial charge assignment'. arXiv, Feb. 16, 2023. Accessed: Oct. 06, 2023. [Online]. Available: http://arxiv.org/abs/2302.06758
[2] J. T. Horton, A. E. A. Allen, L. S. Dodda, and D. J. Cole, 'QUBEKit: Automating the Derivation of Force Field Parameters from Quantum Mechanics', J. Chem. Inf. Model., vol. 59, no. 4, pp. 1366–1381, Apr. 2019, doi: 10.1021/acs.jcim.8b00767.

Theoretical Prediction of Metal-Biomolecule Interactions

Thuraya Alhabradi[1], Prof. Jamie Platts[1]
[1] *School of Chemistry, Cardiff University*

Abstract

The assembly kinetics of Aβ fibrils are significantly influenced by pH, with the sensitivity of first nucleation observed at a midway (pKa) of 7.0, corresponding to the pKa of Aβ histidine imidazole. Surprisingly, as the pH approaches the isoelectric point (pI) of Aβ, which is 5.3, the solubility of Aβ diminishes. This decrease in solubility encourages the self-association of Aβ and the development of amyloid aggregates. Alterations in pH have little impact on the rates of secondary nucleation and elongation, showing that the charge of Aβ, namely the protonation of histidine, illuminates the mechanisms by which Aβ is assembled, facilitating the development of inhibitors, and deepens our understanding of the pathogenesis of Alzheimer's disease.

Molecular dynamics (MD) is a suitable way to study how this protein's shape changes and interacts with itself or other molecules as a function of pH. We have used constant pH molecular dynamics simulations to investigate the impacts of pH change on Aβ monomers and dimers. We used MD simulations across a range of pH values (6, 7, and 8) to examine various pH settings and better understand how pH influences secondary structure and stability.

Utilizing constant pH settings, we conducted molecular dynamics (MD) simulations to compare Aβ. These simulations allowed us to observe the stability and equilibration of system configurations, as indicated by the RMSD and Rg data. These findings are consistent with our initial hypothesis. The cluster analysis identified a wide range of structures in different systems, such as β-strands and α-helices, providing vital information about changes in secondary structure. Studying salt bridges and other interactions helps us get insight into how pH affects the structure of Aβ.

Interrogation of Allosteric Signalling in SOAT enzymes using Dynamical Non-Equilibrium Molecular Dynamics

Saleh O. Alyemni[1]; Shozeb Haider [1]
[1] *School of Pharmacy, University College London,29-39 Brunswick Square, London WC1N 1AX*
Corresponding author: saleh.alyemni.19@ucl.ac.uk

Abstract

Sterol-o-acyltransferases (SOAT) enzymes are involved in different disorders such as atherosclerosis, neurodegenerative diseases, and certain types of cancer. Although the two SOAT isoforms SOAT1 and SOAT2 catalyse the same reaction and share similar structures they still behave differently in response to the same inhibitors (Long et al.,2020,2021). We believe that studying allosteric sites found in the structures of those two enzymes could help us in answering the question concerning the nature of their unique behaviour.

Extensive non-equilibrium simulations (9 μs each) were used to interrogate the allosteric signals in SOAT enzymes. Specifically, we used a method called dynamical non-equilibrium molecular dynamics (D-NEMD) (Oliveria et al.,2021) that allows us to follow any system in time after the introduction of a perturbation and compare it with the unperturbed simulation to calculate any observables. We calculated the Cα-deviation between the two conditions, and we traced the allosteric signal from the origin to its destination.

The allosteric signal in the identified major path starts from the allosteric site then travels to the FYXDWWN motif then to the TM8-TM9 loop from there to the TM8 helix to reach the active site. Two minor allosteric paths branching towards regions distant from the active site, but which might be involved in SOAT gating mechanisms are observed as well. We believe that allosteric signalling paths revealed here could help in clarifying the distinct behaviour of the two SOAT isoforms. Furthermore, our findings could provide drug hunters with a blueprint to look for cryptic pockets along the identified allosteric paths to design isoform specific drugs.

References

[1]Long, T.; Sun, Y.; Hassan, A.; Qi, X.; Li, X. Structure of Nevanimibe-Bound Tetrameric Human ACAT1. Nature 2020, 581 (7808), 339–343.
[2]Long, T.; Liu, Y.; Li, X. Molecular Structures of Human ACAT2 Disclose Mechanism for Selective Inhibition. Structure 2021, 29 (12), 1410-1418.e4.
[3]Oliveira, A. S. F.; Ciccotti, G.; Haider, S.; Mulholland, A. J. Dynamical Nonequilibrium Molecular Dynamics Reveals the Structural Basis for Allostery and Signal Propagation in Biomolecular Systems. Eur. Phys. J. B 2021, 94 (7), 144.

Deciphering the mechanism of avibactam against KPC β-Lactamases through QM/MM computational assays

Jaida Begum[1]; Adrian J. Mulholland[2]; James Spencer[3]; Marc W. van der Kamp[1]
[1]*School of Biochemistry, University of Bristol, Bristol, BS8 1TD, UK*
[2]*School of Chemistry, University of Bristol, Bristol, BS8 1TS, UK*
[3]*School of Cellular and Molecular Medicine, University of Bristol, Bristol, BS8 1TD, UK*

Corresponding author: Jaida.begum@bristol.ac.uk

Abstract

Diazabicyclooctanes (DBOs), such as avibactam, represent a crucial class of covalent β-lactamase inhibitors with clinical significance. Together with β-lactam antibiotics, they provide powerful combination therapies to treat complicated infections, as exemplified by avibactam-ceftazidime. However, the global rise of novel KPC β-lactamases threatens the effectiveness of DBO inhibitors in potentiating β-lactam antibiotics. It remains elusive how the KPC-variants alter the reaction rates of DBOs as a possible path towards resistance.

Using multiscale QM/MM simulations, we aim to explain the molecular basis behind the experimentally observed differences between DBOs and different enzyme variants. Using the clinically relevant β-lactamase KPC-2 with avibactam, MM MD simulations were first used to establish protonation states of key active site residues. Then, the potential energy profile of avibactam acylation was modelled using QM/MM with climbing image-nudged elastic band (CI-NEB) calculations. The rate-limiting step of acylation was revealed and corroborated by DFT/MM CI-NEB calculations (at the ωB97x/def2-TZVP//ff14SB level). To enable fast reaction simulations, 2D potential energy surfaces using the semi-empirical PDDG-PM3/ff14SB showed excellent agreement to the CI-NEB benchmarking studies and this method was then used successfully to determine the minimum free energy pathway using two-dimensional umbrella sampling (2D US).

2D US was used as an "avibactam acylation assay" to investigate different KPC variants of clinical interest bearing single amino acid substitutions. Good agreement between experimental and predicted activity amongst different enzymes was obtained. For example, our protocol was able to clearly differentiate between avibactam activity of KPC-2 vs. KPC-3 (His274Tyr mutation). Work is currently underway to establish the origins for the variation of free energy barriers amongst multiple KPC variants. Our computational assays will help to dissect the reaction mechanism of DBOs on an atomistic level and rationalize experimental differences. This can help anticipate variants with reduced DBO susceptibility and may aid design of new, efficient inhibitors or antibiotics to treat life-threating infections.

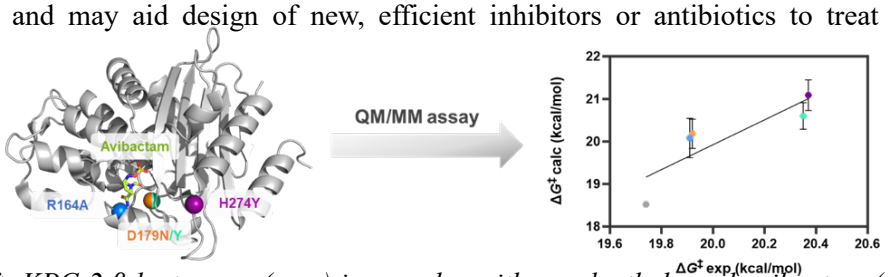


Figure 1. Left: KPC-2 β-lactamase (grey) in complex with covalently bound avibactam (green). Single amino acid substitutions of KPC-2 shown as spheres, Right: Experimental activation energies versus calculated free energy barriers for rate-limiting step of acylation. Error bars represent the standard deviation from three replicas.

Development of Novel Non-Peptidic VHL Binders Using a Fragment-Based Approach

Patricia Blanco-Gabella[1], Kevin Haubrich[2], Ryan Casement[2], Valentina Spiteri[2], Carles Galdeano[1], Jordi Juárez-Jiménez[1], Alessio Ciulli[2], Xavier Barril[1]

[1]*Department of Physical Chemistry, Faculty of Pharmacy, University of Barcelona, Joan XXIII s/n 08028, Barcelona, Spain*

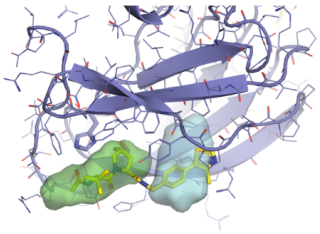
[2]*Centre for Targeted Protein Degradation, University of Dundee,1 James Lindsay Place, Dundee, DD1 5JJ, United Kingdom*

Corresponding author: patricia.blanco@ub.edu

Abstract

Targeted Protein Degradation has emerged in recent years as a revolutionary strategy in drug discovery. The von Hippel–Lindau protein (VHL) is a well-validated E3 ligase that is recruited by many efficacious PROteolysis TArgeting Chimera (PROTACs) molecules [1]. However, all existing VHL warheads have been developed around a central hydroxyproline unit[2,3] and, owing to their peptidic nature, exploiting them pharmacologically is challenging due to their poor absorption, distribution, and metabolism[4,5]. Consequently, few VHL PROTACs have reached clinical stages so far[6].

Here, we apply a fragment-based approach combining computational techniques and ligand-observed NMR studies to discover new chemotypes for VHL ligands that could be developed into more effective drugs. A virtual screening of all the accessible fragment space followed by paramagnetic relaxation enhancement assays allowed us to identify two novel fragment hits for the hydroxyproline binding site. These fragments show a stronger response for VHL than the L-hydroxyproline core on its own. Subsequently, we performed µs-long molecular dynamics simulations of VHL and multiple copies of each ligand to refine the predicted poses of the fragment hits. Finally, these compounds will be used as a starting point for fragment-growing strategies to obtain more potent ligands.



References

[1] Békés, M., Langley, D. R., and Crews, C. M. (2022). *Nature Reviews Drug Discovery*, **21**(3), 181-200.
[2] Frost, J. et al. (2016). *Nature communications*, **7**(1), 13312.
[3] Krieger, J. et al. (2023). *ChemMedChem*, **18**(8), e202200615.

[4] Pike, A. et al. (2020). *Drug Discovery Today*, **25**(10), 1793-1800. [5] Castellani, B. et al. (2023). *Journal of Medicinal Chemistry*, **66**(18), 13148-13171. [6] Chirnomas, D., Hornberger, K. R., and Crews, C. M. (2023). *Nature Reviews Clinical Oncology*, ***20***(4), 265-278.

Alchemical Free Energy Calculations with SOMD2

Matthew Burman[1]; Lester Hedges[1,2], Christopher Woods[1,2] and Julien Michel[1,3]
[1]*OpenBioSim community interest company*
[2] *University of Bristol*
[3] *University of Edinburgh*

Corresponding author: matthew@openbiosim.org

Abstract

SOMD2 is an open-source tool for alchemical free energy calculations, leveraging Sire and OpenMM libraries. Accessible via a command line interface or a Python API, SOMD2 is designed for ease of use and automatic resource management, with low-level access to molecules and lambda levers offering flexibility and adaptability. SOMD2 also allows precise control over simulation parameters, supporting multiple restraints, automatic hydrogen mass repartitioning, and a custom lambda scheduling framework for complex perturbation pathways.

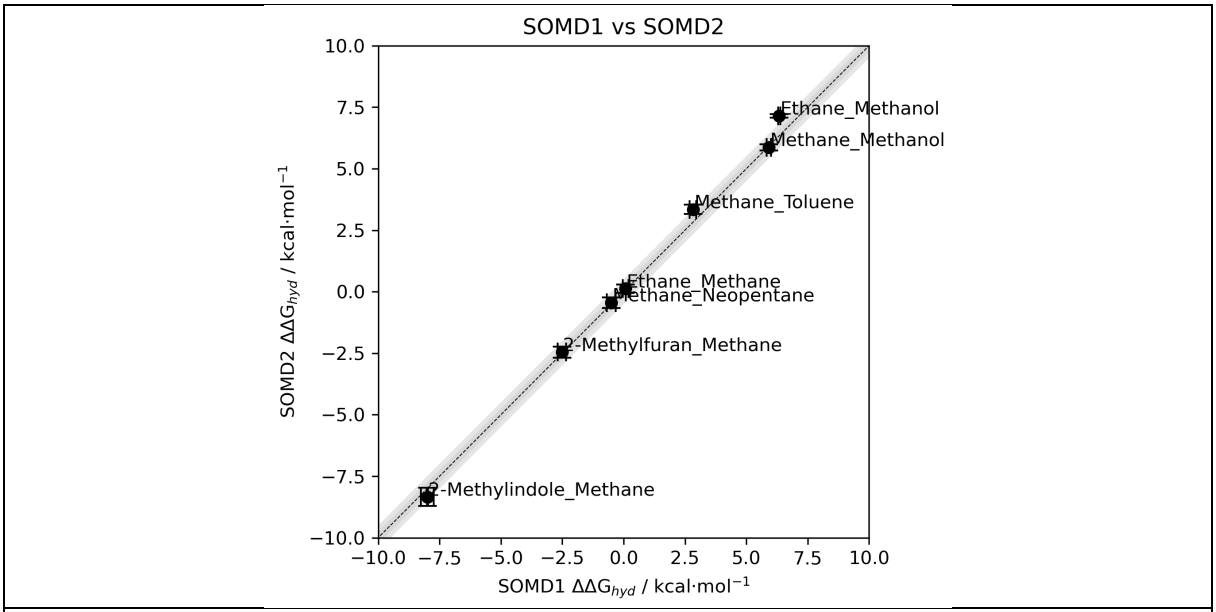


Figure 1: *Comparison of hydration free energy results between SOMD2 and its predecessor SOMD1*

References

[1] Hedges et al., (2019). BioSimSpace: An interoperable Python framework for biomolecular simulation. Journal of Open Source Software, 4(43), 1831, https://doi.org/10.21105/joss.01831

[2] Woods CJ, Hedges LO, Mulholland AJ, Malaisree M, Tosco P, Loeffler HH, Suruzhon M, Burman M, Bariami S, Bosisio S, Calabro G, Clark F, Mey ASJS, Michel J. Sire: An interoperability engine for prototyping algorithms and exchanging information between molecular simulation programs. J Chem Phys. 2024 May 28;160(20):202503. doi: 10.1063/5.0200458. PMID: 38814008.

Integrative Structural Studies of RNase E and the RNA Degradosome

Lisa Casteller[1]; Prof. Ben F. Luisi[1,2]
[1,2] Department of Biochemistry, University of Cambridge

Corresponding author: lc893@cam.ac.uk

Abstract

The aim of this project is to employ an integrative structural approach to investigate the molecular mechanisms underpinning spatial organisation, RNA capture and liquid-liquid phase separation of the RNA degradosome.

The RNA degradosome, responsible for selective RNA degradation and maturation in E. coli, is a highly dynamic multiprotein complex canonically consisting of RNAse E, RhlB, Enolase and PNPase. Spatial organisation of the degradosome machinery is believed to rely on endoribonuclease RNase E, who's unstructured C-terminal domain acts as a recruitment platform (Figure 1). Current structural understanding is limited to experimentally determined structures of the RNase E N-terminal Domain, Enolase and PNPase, however further characterisation by conventional means such as Cryo-EM or X-ray Crystallography is hindered by the dynamic nature of the RNA degradosome.

Molecular dynamics simulations provide an alternative investigative approach. Combining protein models from cryo-EM, X-ray Crystallography, and machine learning-based 3D predictions with subassembly dynamic information from biophysical experiments including DLS, SEC-MALS and SEC-SAXS, a model of the canonical RNA degradosome can be built. Coarse grained molecular dynamics simulations will be used to test the conformational ensemble and interrogate the molecular mechanisms of the complex.

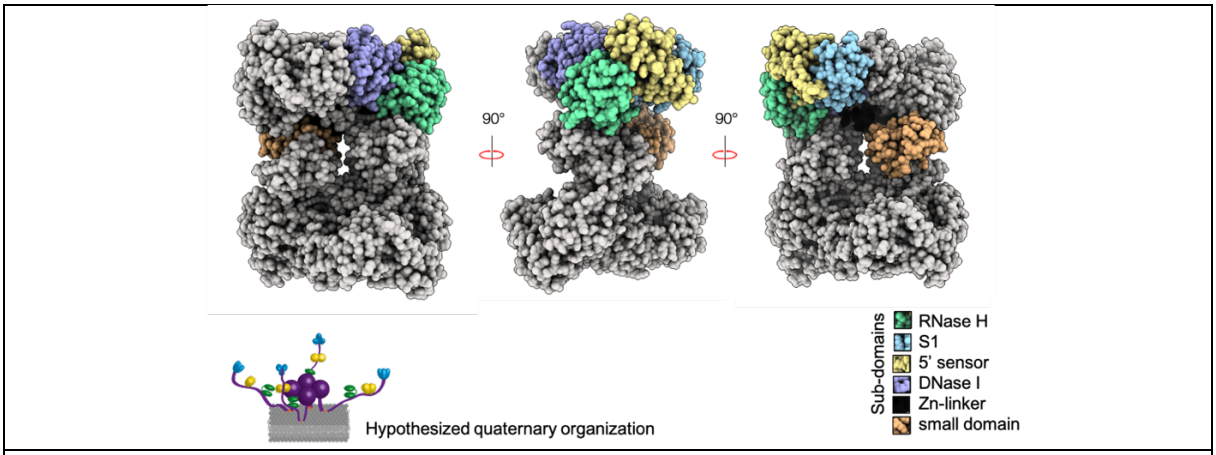


Figure 1. RNAse E N-terminal domain homotetramer adapted from PDB 6G63, featuring a single monomer coloured by subdomain. Hypothesised full-length quaternary conformation adapted from Dendooven, Paris et al. (2022).

References

[1] Dendooven, T., G. Paris, A. V. Shkumatov, M. S. Islam, A. Burt, M. A. Kubańska, T. Y. Yang, S. W. Hardwick and B. F. Luisi (2022). "Multi-scale ensemble properties of the Escherichia coli RNA degradosome." Mol Microbiol **117**(1): 102-120.

Neural Network Models for Prediction of Biological Activity Using Molecular Dynamics Data: A Case of Photoswitchable Peptides

Anton Cherednichenko[1,2]; Ihor Koleiev[1,3]; Serhii Starosyla[3]; Igor V. Komarov*[1,2]

[1]*Taras Shevchenko National University of Kyiv, Vul. Volodymyrska 60, 01601 Kyiv, Ukraine*
[2] *Enamine Ltd., Vul. Chervonotkatska 78, 02066 Kyiv, Ukraine*
[3] *ReceptorAI, Ground floor-20, United Kingdom, 20-22 Wenlock Road, London N1 7GU*

Corresponding author: igor.komarov@knu.ua

Abstract

Abstract: Molecular dynamics (MD) data contain information on structure and dynamics of compounds, which both determine the compounds' biological activity. One may hypothesize therefore that the features extracted from the MD trajectories may be used for prediction the biological activity of similar compounds using Neuronal Network (NN) models in target-independent manner, without involving the features describing the interaction of the compounds with their biological targets. One may also assume that if an NN model is trained using MD/activities of pairs of similar compounds, it may be used for prediction of differences in biological activities of compounds possessing entirely different biological action mechanism comparing to the training set. Here, verification of these hypotheses on the NN models aimed at prediction of biological activity of peptides and peptidomimetics is reported. Two NN models were constructed, and their performance was tested on sets of photo-isomerizable membrane-active peptides. The peptides contain a diarylethene "photoswitch" in their molecules, so they can exist in two isomeric forms termed open and closed, interconvertible into each other by irradiation with light of different wavelengths.

To predict how photoisomerization influences the biological activity, and if the light can be used to efficiently control it is important for a rapidly developing area of science called photopharmacology. One of the models, named "Activity Prediction Model" was trained on MD features of photoswitchable cytotoxic peptidomimetics, analogues of a natural peptidic antibiotic gramicidin S. It adequately predicts the cytotoxic activity for similar analogues, including the diarylethene-containing photoisomers. The other NN model, "Open-Closed Similarity Prediction", was trained by features derived from the MD trajectories of pairs of photoisomers of the same gramicidin S analogues which were used for the training of the Activity Prediction model. Optimised Open-Closed Similarity Prediction model reliably predicts the differences in the protease-inhibiting activity of the diarylethene photoisomers from the validation set.

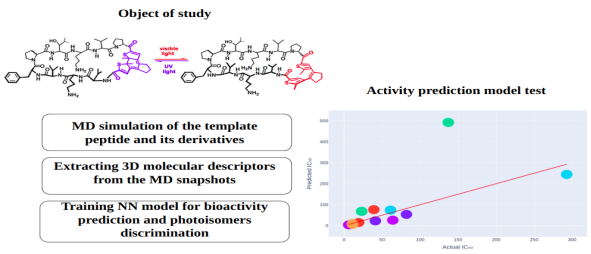


Figure 1

References

[1] O. Babii, S. Afonin, A.Yu. Ishchenko, T. Schober, A.O. Negelia, G.M. Tolstanova, L.V. Garmanchuk, L.I. Ostapchenko, I.V. Komarov and A.S. Ulrich, *J. Med. Chem.*, 2018, 61, 10793.

Exploring the Formation of Protein Cryptic Pockets with Ghost Probes

Joan Clark-Nicolas[1]
[1]*Independent Scientist*

Corresponding author: joan.clark88@gmail.com

Abstract

The detection of protein cryptic pockets often requires conformational rearrangements that cannot be detected in the time scales typically simulated by classical Molecular Dynamics (MD) simulations. A variety of MD workflows rely on the use of small chemical probes to induce the formation of and stabilise protein cavities [1-7]. Those methods, however, are limited to the region of the chemical space defined by the chosen probes, and the fact that the probes are initially placed in the bulk solvent can hinder the exploration of cavities that are buried deep within the target protein.

In this talk, I will present an approach that addresses those issues by using what will be referred to as Ghost Probes. Ghost Probes are not part of the simulated system, but 3-dimensional points associated with an activity score which represents the ability of the protein to bind a small chemical probe at their location (Figure 1). This score is used to construct a biasing potential capable of opening a cavity in the surrounding area and is paired with a search algorithm that enhances the sampling of different protein cavities.

Ghost probes are not constrained to a certain region of the chemical space, and they can be placed on any region of the protein. This increases the possibilities of sampling buried as well as shallow cryptic pockets and presents an opportunity to expand the druggable proteome with proteins previously thought undruggable.

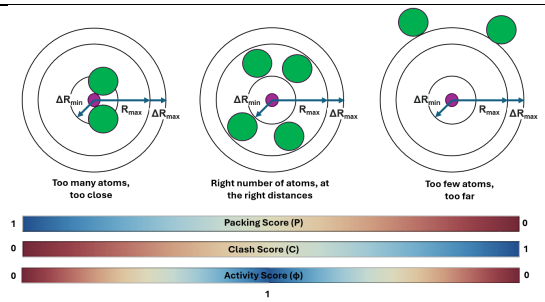


Figure 1: *Schematic view of the activity score of a ghost probe (purple), considering the number of surrounding atoms (green), and their location relative to the ghost probe.*

References

[1] Ghanakota, P. et al., J. Med. Chem., 2016, **59**, 10383–10399.
[2] Bissaro, M. et al., Drug Discov. Today, 2020, **25**, 1693.
[3] Bakan, A. et al., J. Chem. Theory Comput., 2012, **8**, 2435–2447.
[4] Ghanakota, P. et al., J. Phys. Chem. B, 2016, **120**, 8685–8695.
[5] Smith, R. D. et al., J. Chem. Inf. Model., 2021, **61**, 1287–1299.
[6] Schmidt, D. et al., J. Chem. Theory Comput., 2019, **15**, 3331–3343.
[7] Martinez-Rosell, G. et al., J. Chem. Inf. Model., 2020, **60**, 2314–2324.

GRK2 and Neurotensin Receptor: from PDB to MD

Hocine El Khaoudi Enyoury[1]; Peter Kolb[1]
[1]*Department of Pharmaceutical Chemistry, University of Marburg*

Corresponding author: peter.kolb@uni-marburg.de

Abstract

GPCRs constitute the largest protein family, and are associated with many normal as well as aberrant biological processes. They are also known as seven-domain transmembrane receptors (7TM) and approximately 40% of present-day drugs target this class of proteins. The functions of GPCRs are very diverse and they recognize a wide variety of ligands, including photons, small molecules, and peptides [1].

GRK2 is an important modulator of GPCRs as well as their signaling pathways, due to its role in phosphorylating the receptor after activation. Furthermore, GRK2 can interact with different signaling proteins related to cell migration [2,3]. Studies have shown that a higher or lower expression of GRK2 may be involved in numerous pathologies such as cardiovascular diseases, neurodegenerative diseases, metabolic syndromes, or cancer [4,5,6].

So far, the dynamics of the interactions between GRK2 and GPCRs are poorly understood due to the unavailability of experimental structures of such complexes. Recently, Duan *et al.* reported the first structure of a G protein-coupled receptor, the neurotensin receptor 1, in complex with GRK2 [7]. In our study, we explored the conformational space of the complex via molecular dynamics simulations, aiming to understand the interactions between GRK2, $G\alpha_q$, and the neurotensin receptor.

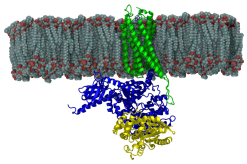


Figure 1: *Structure of NTSR1(yellow) - GRK2(blue) - $G\alpha_q$(yellow) complex.*

References

[1] Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B., Gloriam, D. E. (2017). *Nature Reviews Drug Discovery*, 16(12), 829–842.
[2] Penela, P., Ribas, C., Aymerich, I., Eijkelkamp, N., Barreiro, O., Heijnen, C. J., Kavelaars, A., Sánchez-Madrid, F., Mayor Jr, F. (2008). *The EMBO Journal*, 27(8), 1206-1218.
[3] Penela, P., Nogués, L., Mayor, F. (2014). *Current Opinion in Cell Biology*, 27C(1), 10-17.
[4] Oyama, J.-I., Node, K. (2018). *Hypertension Research*, 41(8), 575-577.
[5] Obrenovich ME, Palacios HH, Gasimov E, Leszek J, Aliev G. (2009). *Cardiovascular Psychiatry and Neurology*, 2009, 1-14.
[6] Nogués L, Reglero C, Rivas V, Salcedo A, Lafarga V, Neves M, Ramos P, Mendiola M, Berjón A, Stamatakis K, Zhou XZ, Lu KP, Hardisson D, Mayor F Jr, Penela P. (2016). *EBioMedicine*, 13, 132-145.
[7] Duan J, Liu H, Zhao F, Yuan Q, Ji Y, Cai X, He X, Li X, Li J, Wu K, Gao T, Zhu S, Lin S, Wang MW, Cheng X, Yin W, Jiang Y, Yang D, Xu HE. (2023). *Nature*, 620(7974), 676-681.

In silico design of small molecule probes of fibrosis using molecular dynamics simulations

Aya M. Emam[1,2]; Richard A. Bryce[1], Sam Butterworth[1], Holly Birchenough[3], Thomas A. Jowitt[3], Karen Piper Hanley[4]

1 Division of Pharmacy and Optometry, School of Health Sciences, University of Manchester, UK. 2 Department of Medicinal Chemistry, Faculty of Pharmacy, Zagazig University, Egypt. 3 Biomolecular Analysis Core Facility, University of Manchester, Manchester, UK. 4 Division of Diabetes, Endocrinology and Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, UK.

Corresponding author: richard.bryce@manchester.ac.uk

Abstract

Fibrosis is a progressive health disease that affects different organs of the body. In the liver it can develop into severe stages such as cirrhosis and cancer.(1) One of the key mediators which play an important role in signaling pathway during fibrosis are integrins.(2) In this project, we focus on integrin α11β1 which is overexpressed in myofibroblasts during fibrosis.(3) Antibodies targeting this integrin have been developed for treatment of fibrosis, joint disease, and cancer;(4) however, there is a lack of effective small molecule inhibitors to date.(5) Our aim is to identify small molecule allosteric inhibitors of this integrin using computational tools. Virtual screening of ZINC and FDA approved drug databases as well as molecular dynamics simulations were performed to identify potential small molecule binders to α11β1 for subsequent experimental testing. Based on the hits identified from *in silico* tools, selected small molecules were assessed for their binding affinities using the biophysical assays, microscale thermophoresis (MST) and surface plasmon resonance (SPR). From these assays, four small molecules which bind to integrin α11β1 with micromolar affinity were found (eg. **21** and **23** in Figure 1).

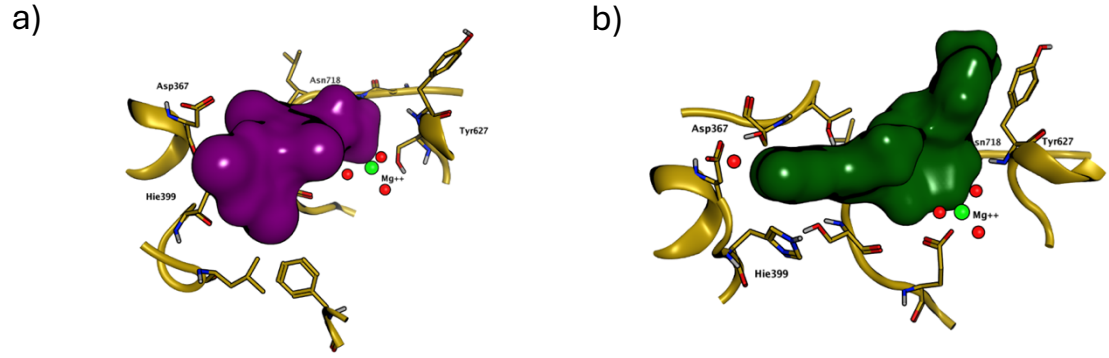


Figure 1: *Predicted interactions of hit compounds (a)* ***21*** *and (b)* ***23*** *respectively with* Mg^{2+} *and amino acid residues within allosteric pocket of integrin α11β1.*

References

[1] O'Rourke JM, et al. (2018). *World J Gastroenterol*, **24**(39), 4436-47. [2] Rahman SR, et al. (2022). *Liver International*, **42**(3), 507-521. [3] Martin K, et al. (2016). *Nat Commun*, **18**(7), 12502. [4] Gullberg E. (2019). WO2019154961A1. [5] Hunter EJ, et al. (2021). *Toxicol Appl Pharmacol*, **1**(428), 115669.

CodeEntropy Software Development

Sarah K. Fegan[1], Ioana A. Papa[1,2], Jas Kalayan[1], Richard H. Henchman[3], Sarah A. Harris[4]

[1] *Scientific Computing Department, STFC, Daresbury Lab, UK*
[2] *Department of Chemistry, University of Edinburgh, UK*
[3] *School of Medicine, University of Sydney, Australia*
[4] *Department of Physics, University of Sheffield, UK*

Corresponding author: sarah.fegan@stfc.ac.uk

Abstract

Entropy is an interesting thermodynamic quantity which contributes to the free energy of any system. The CodeEntropy software is an open source software program which implements the multiscale cell correlation method [1] for determining the entropy of a system from molecular dynamics trajectories. This method breaks the entropy into independent terms and calculates each at different length scales. Here we describe how the code works and give details of the latest release and future plans.

References
[1] Chakravorty, Arghya; Higham, Jonathan and Henchman, Richard H. (2020). *J. Chem. Inf. Model.*, **60**, 5540−5551.

Correct nucleotide selection is confined at the binding site of polymerase enzymes

David Ricardo Figueroa Blanco[1], Pietro Vidossich[1], Marco De Vivo[1]
[1] *Laboratory of Molecular Modelling & Drug Discovery, Istituto Italiano di Tecnologia, Via Morego 30, Genoa 16163, Italy*

Abstract

Polymerases (Pols) catalyze RNA or DNA synthesis by adding ribonucleoside triphosphates (NTPs) or deoxyribonucleoside triphosphates (dNTPs), respectively. The insertion of correct versus incorrect nucleotides relates to Pols' fidelity, which defines their ability to faithfully replicate DNA strands in a template-dependent manner. These enzymes play a central role in several crucial processes such as repair, translation and replication of genetic material. The wide range of processes in which they are involved makes these enzymes promising targets for drug discovery, as well as for biotechnological application.[1]

Previous studies, including ours, have demonstrated that the alignment of reactants and base pairing are crucial structural features for fidelity.[2] In this work, we used molecular dynamics simulations to show how the protein-DNA interface near the catalytic site differs when correct (dG:dCTP) versus incorrect (dG:dATP) dNTPs are bound in Pol Beta. The formation and dynamic stability of specific interatomic interactions around the binding site explain the differences in binding affinity for correct and incorrect dNTPs. We also show how changes in the interaction network in the mutation R283A lead to a lower capacity to differentiate between correct and incorrect nucleotides, which influence directly the Pol fidelity. Additionally, we used alchemical free energy calculations to quantify the relative binding free energy for the transformation from correct to incorrect (dCTP→dATP) in both WT and R283A, further validating our models and mechanistic insights. Finally, sequence and structural alignments with other Pols from the same family suggest that these findings can be rationalized in other enzymes.

References

(1) Geronimo, I.; Vidossich, P.; Donati, E.; De Vivo, M. Computational Investigations of Polymerase Enzymes: Structure, Function, Inhibition, and Biotechnology. *WIREs Comput. Mol. Sci.* **2021**, *11* (6), e1534. https://doi.org/10.1002/wcms.1534.
(2) Geronimo, I.; Vidossich, P.; De Vivo, M. Local Structural Dynamics at the Metal-Centered Catalytic Site of Polymerases Is Critical for Fidelity. *ACS Catal.* **2021**, 14110–14121. https://doi.org/10.1021/acscatal.1c03840.

A Deep Learning Approach to Bottom-Up Coarse-Grained Protein Forcefields

Andrea Guljas[1,2], Nick Charron[3], Felix Musil[1], Yaoyi Chen[2,4,5], Andreas Krämer[6], Frank Noé[4,5], Cecilia Clementi[1,2]

[1] *Department of Physics, Freie Universität Berlin, Arnimallee 12, Berlin 14195, Germany.*
[2] *Max Planck Institute for Molecular Genetics, Ihnestraße 63, Berlin 14195, Germany.*
[3] *Zuse Institute Berlin, Freie Universität Berlin, Takustraße 7, Berlin 14195, Germany.*
[4] *Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 12, Berlin 14195, Germany.*
[5] *AI4Science, Microsoft Research, Karl-Liebknecht Str. 32, Berlin 10178, Germany.*
[6] *D. E. Shaw Research, 120 W 45th St 39th Fl, New York, NY 10036, United States.*

Corresponding author: aguljas@zedat.fu-berlin.de

Abstract

Molecular coarse graining (CG) is an invaluable tool for understanding protein dynamics, particularly in the study of large biomolecular systems. While molecular dynamics (MD) yields the dynamics of a system's atomistic degrees of freedom, emergent macro-scale properties are often more relevant than the microscopic dynamics themselves. Furthermore, simulations of many biological processes cannot be performed on timescales required to reproduce experimental observables. Using CG models, we focus on slow collective atomic motions and simulate protein systems on scales inaccessible to MD.

Bottom-up CG incorporates information from higher resolution models to create a simplified representation of the system. However, this process leads to the emergence of multi-body effects which are difficult to approximate by conventional physically-informed functions. By leveraging state-of-the-art machine learning approaches, it is possible to develop a CG potential that effectively captures these multi-bodied effects and accurately reproduces the thermodynamics of the underlying system [1,2]. Here, we train deep graph neural networks on atomistic MD simulations using a force-matching approach to learn complex interactions and reproduce the probability density distribution of the atomistic model projected onto the CG mapping [3].

To ensure that the model is transferable across various systems, we fine-tune both the composition of the input training data and the physics informed correction terms that are applied to the model. The transferability of these models is assessed by simulating protein folding and aggregation, and successful models are eventually applied to study the dynamics of large systems across long timescales.

References

[1] Wang, J, Olsson, S, Wehmeyer C, Charron, N E, de Fabritiis, G, Noé, F, Clementi, C. (2019). *ACS Central Science*, **5**(5), 755–767.
[2] Husic, B E, Charron, N E, Lemm, D, Wang, J, Pérez, A, Majewski, M, Krämer, A, Chen, Y, Olsson, S, de Fabritiis, G, Noé, F, Clementi, C. (2020). *J. Chem. Phys*. **153**(19), 194101.
[3] Charron, N E, Musil, F, Guljas, A, Chen, Y, Bonneau, K, Pasos-Trejo, A S, Venturin, J, Gusew, D, Zaporozhets, I, Krämer, A, Templeton, C, Kelkar, A, Durumeric, A E P, Olsson, S Pérez, A, Majewski, M, Husic, B E, Patel, A, De Fabritiis, G, Noé, F, Clementi, C. (2023). arXiv:2310.18278

Can Protein-Peptide Docking Replicate Experimental Data?

Charlie Holdship[1]; Dr. Leonie Windeln;[1] Dr. Jack Sawdon;[1] Dr. Steven Turner;[1] Prof. Jonathan Essex;[1] Prof. Jeremy Frey;[1]
[1]*University of Southampton, University Road, Southampton SO17 1BJ*

Corresponding author: cjh2g17@soton.ac,uk

Abstract

Conotoxin peptides are a class of neurotoxin derived from marine cone snail venom and target human nicotinic acetylcholine receptors (nAChR's). While their solution conformations and nAChR binding sites are known, their precise binding interactions are not and to elucidate their mechanism of action, these must be identified. Although alanine scanning experiments can provide information on which residues are crucial for peptide binding, the exact nature of how they bind is left to crystallographic methods or protein-peptide docking.

To sample the full range of peptide conformations and to give docking the best chance of finding the interactions, we performed molecular dynamics simulations on the known peptide structures and their alanine scanning mutants, using enhanced sampling (replica exchange (REST2)). The most frequently seen conformations were identified by clustering on the torsion angle space and then taken forward to the docking studies.
In these docking studies, we identify the interacting residues from poses that are consistent with the alanine scanning data. We also separately dock the different alanine scanning mutants to see how well the docking scores correlate to the experimental data.[1] We find generally good agreement between the docking scores and interactions involving the crucial residues identified by alanine scanning and hope to compare with an experimental structure in the future.

References

[1] Ning, J., Ren, J., Xiong, Y., Wu, Y., Zhangsun, M., Zhangsun, D., Zhu, X., and Luo, S. (2019). *Toxins*, **11**(10), 603-617.

Modelling DNA in Complex Topologies: The Role of Gyrase

Katy Hollands[1]; Dmitry Ghilarov[2]; Anthony Maxwell[2]; Mark Leake[1,3]; Agnes Noy[1]
[1] *School of Physics, Engineering and Technology, University of York*
[2] *Department of Biology, John Innes Centre*
[3] *Department of Biology, University of York*

Corresponding author: katy.hollands@york.ac.uk

Abstract

DNA gyrase is a type IIA topoisomerase, which is an enzyme that can relax and induce supercoils in DNA through the strand passage of the double helix. It only occurs in bacteria and is key in the DNA replication and transcription processes, as it relaxes superhelical tension caused by DNA polymerase. It is also uniquely capable of inducing negative supercoils through the hydrolysis of adenosine triphosphate (ATP). Due to its importance in the survival of the cell, they are a common target for antibiotics.

The exact mechanisms of the enzyme, such as strand passage and energy coupling, haven't been observed with experiments, and so these will be investigated using Molecular Dynamics (MD) simulations. The impact of varying factors, such as length of DNA and mutations, have also been looked at. This allows us to see the system in a dynamic way, with the hope of finding new potential targets for antibiotics.

Computational Modeling for Laundry Powder and Porphyrin-Binding Proteins

William Houppy[1]; W. Singh[1]; G. Black[1]; N. Lant[2]
[1] Department of Applied Sciences, Northumbria University of Newcastle (UK)
[2] Procter & Gamble, Newcastle Innovation Center (UK)

William HOUPPY
- PhD Student -
Northumbria University
Newcastle Upon Tyne (UK)

COMPUTATIONAL MODELING FOR LAUNDRY POWDER & PORPHYRIN-BINDING PROTEINS

"Computational approaches to engineer enzymes for improved cold-water performance"
"Porphyrin-Binding Proteins modeling"

W. Houppy,* Dr W. Singh,* Prof G. Black,* Dr N. Lant†

Enzymes in laundry detergents

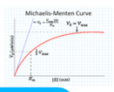
- 1913: 1st use of enzymes in laundry but became widespread in the 1960's
- CO_2 emissions of laundry are 80% due to heating the washing water
- Most enzymes from bacterial cultures are inactive <30°C, new enzymes needed for cold water
- Current approaches like directed evolution and rational design are slow & lab-time demanding
- ***Using computational modeling could speed up the discovery of cold-adapted enzymes***

Computational modeling

- Enzyme active site study with AutoDock
- Molecular dynamics using Amber
- Mechanistic with quantum chemistry

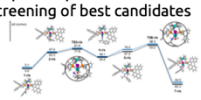
Laboratory testing

- Kinetics measured experimentally
- Industrial testing to show performance
- Validation of computational models



In silico engineering

- Mutagenesis design *in silico* & αFold
- Properties predictions of mutants
- Screening of best candidates



Enzyme synthesis

- Recombinant proteins synthesis
- Purification and characterization of proteins

1967 **2023**

ARIEL® launched in 1967 with protease, is a brand of the portfolio of Procter & Gamble who, at the forefront of innovation, introduced later other enzymes like amylase, lipase, cellulase and mannanase.

Prospectives

- A novel depolymerase is being investigated
- Initial simulations are in good agreement
- Quantum chemistry studies are planned
- Also deep learning algorithms are considered
- Industrial testing at Procter & Gamble

Porphyrin-Binding Proteins

- PBPs like HusA or GUN4 crucial in Heme , Chlorophyll metabolism
- New research lead with novel applications[1] and developments[2]
- Molecular dynamics performed with Protoporphyrin IX with X-ray structure of GUN4 (PDB ***1Y6I***) and NMR for HusA (PDB ***6PQ5***)
- GUN4 apoprotein simulation overcame closed conformation
- Showing different however similar promiscuous binding

Future work

- Further simulations with Heme and other porphyrins
- Free energy calculations like MMPBSA will be exploited
- Identification of residues important for binding site
- Quantum Chemistry deployed to reproduce absorption spectra
- Aim to utilize neural networks for *in silico* design of original PBPs

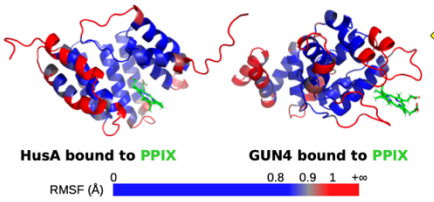
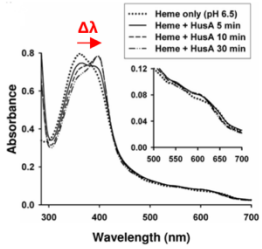


Figure 1: Root Mean Square Fluctuation (RMSF) of GUN4 & HusA bound to Protoporphyrin IX. (Amber MD)

Figure 2: Observed red shift (Δλ) of UV-Visible absorption spectrum of Heme with HusA. (data from Gao[3])



Northumbria University NEWCASTLE

Biotechnology and Biological Sciences Research Council

BiSCoP Collaborative Training Partnership

P&G

* Department of Applied Sciences, *Northumbria University at Newcastle* (UK)
† Procter & Gamble, *Newcastle Innovation Center* (UK)

[1] Gao, J.L. et al. (2018). *Nat Comm*, **9**(1), 4097
[2] Rockwell, N. C. & Lagarias, J. C. (2023). *PNAS Nexus*, **2**(5), pgad131
[3] Gao. J.-L. *et al. Journal of Bioloaical Chemistry* **285.** 40028–40038 (2010).

Counterions affect the self-assembly of virus-like nanoparticles

Ya-Wen Hsiao[1]; Mara Chiricotto[1], David Bray[1], Maxim Ryadnov[2], Jason Crain[3,4]
[1]*The Hartree Centre, STFC Daresbury Laboratory, Warrington, WA4 4AD, U.K.*
[2]*National Physical Laboratory, Hampton Road, Teddington TW11 0LW, U.K.*
[3]*Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, U.K.*
[4]*IBM Research Europe, Hartree Centre, Warrington WA4 4AD, U.K.*

Corresponding author: Ya-Wen.Hsiao@stfc.ac.uk

Abstract

A wide range of nanoparticles (NP) consisting of liposomes, synthetic and natural polymers, and inorganic-based carriers have been developed for tumour imaging and therapy. Among them, the protein-based NPs have the advantages of uniform structure, controllable assembly, biocompatibility, biodegradability, and suitable sizes. Virus-like nanoparticles (VLP) are one category of protein-based NPs. They can encapsulate a wide range of therapeutic and diagnostic agents, and therefore are full of potential in cancer therapy.

Understanding the driving forces that dominate VLP self-assembly can facilitate the design of VLPs with broader functionality. Electrostatic interactions are the strongest of noncovalent forces and are also long-range. Thus, charges and counterions play an important role in self-assembly processes: One of the strategies to carry/release cargo using VLPs is self-assembly/disassembly around cargo by altering pH and buffer conditions. It is also relevant to consider the Hofmeister series on the tendencies of ions to precipitate or solubilize proteins.

Here we tested the assembly of cyclic peptides (cyc-QdLRdLRdLRdL) with counterions PO_4^{3-}, $H_2PO_4^-$, F^- and Cl^- using MD simulation. The order of these anions of the Hofmeister series (salting-out): $Cl^- < F^- < PO_4^{3-}$; $H_2PO_4^-$. We find that the systems with PO_4^{3-} or $H_2PO_4^-$ form stable packings, while the other anions do not. We therefore investigate what makes PO_4^{3-} or $H_2PO_4^-$ different from F^- and Cl^-. The results may also shed lights on the origin of the Hofmeister series.

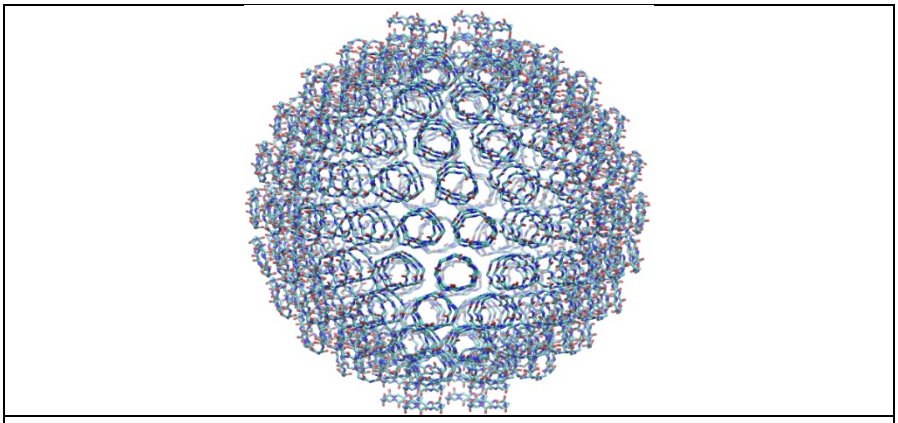


Figure 1: *VLP consists of cyclic peptides*

Computational binding mode elucidation of Fbw7 E3 ligase fragment hits through Multiple-copies Association Studies (MAS)

Varbina Ivanova [1,2]; Roger Castaño [1,3]; Carles Galdeano [1,3]; Jordi Juárez-Jiménez [1,2]; Xavier Barril[1,2,3,4]

[1] *Faculty of Pharmacy and Food Sciences, University of Barcelona*

[2] *Institute of Theoretic and Computational Chemistry (IQTC)*

[3] *Biomedicine Institute of University of Barcelona (IBUB)*

[4] *Catalan Institution for Research and Advanced Studies (ICREA)*

Corresponding author: varbina.ivanova@ub.edu

Abstract

Targeted Protein Degradation is a promising therapeutic approach for the regulation of proteins involved in cancer, neuropsychiatric diseases and other disorders. [1] [2] At the core of this approach, lies the development of small molecules able to engage E3 ubiquitin ligases. To date, however, many E3 ligases have been deemed "undruggable". In this context, the combination of fragment-based drug discovery and allosteric regulation provide an appealing path towards targeting undruggable proteins. [3]

In this work we develop a structure-based computational approach to elucidate fragment binding modes, aimed at the rational design of novel ligands targeting the Fbw7 E3 ligase. Fbw7 acts as a crucial tumor suppressor by facilitating the ubiquitination of key oncogenes such as Cyclin E, C-Myc, and Notch1, along with other vital proteins like DISC1. [4] [5] Previously, our research group had identified 10 potent Fbw7 fragment hits, but crystallography efforts to determine the complex structure have not been successful. Here we present a computational workflow for the elucidation of fragment binding modes through Multiple-copies Association Studies (MAS). MAS combines classical molecular dynamics (MD) simulations with elevated ligand concentrations and employs a post-MD clustering algorithm to rank the probable binding modes and determine the correct one. We have developed a novel LJ potentials adjustment approach to mitigate small molecule aggregation typically observed in water simulations with high concentrations of organic molecules, enabling MAS to be conducted at high concentrations of ligand. To optimize and validate MAS, we are utilizing a benchmarking set comprising protein-ligand systems with crystallographic and binding affinity data. This methodology holds promise for uncovering unknown binding modes for known small molecule binders and has already been applied to elucidate the binding poses of the Fbw7 E3 ligase fragment hits.

References

[1] Saravanan KM, Kannan M, Meera P, Bharathkumar N, Anand T. E3 ligases: a potential multi-drug target for different types of cancers and neurological disorders. (2022) *Future Med Chem.*,**14**(3), 187-201.

[2] Békés M, Langley DR, Crews CM. PROTAC targeted protein degraders: the past is prologue. (2022) *Nat Rev Drug Discov.*, **21**(3), 181-200.

[3] Xie X, Yu T, Li X, et al. Recent advances in targeting the "undruggable" proteins: from drug discovery to clinical trials. (2023) *Signal Transduct Target Ther.*, **8**(1), 335.

[4] Welcker M, Clurman BE. FBW7 ubiquitin ligase: a tumor suppressor at the crossroads of cell division, growth and differentiation. (2008) *Nat Rev Cancer.*, **8**(2), 83-93.

[5] Yalla K, et al. FBXW7 regulates DISC1 stability via the ubiquitin-proteosome system. (2018) *Molecular Psychiatry 23.5*, 1278-1286.

Curate and Share Complex Biomolecular Simulation Workflows

Jas Kalayan[1]; Kin Chao[1,2], Sarah Rouse[2], James Gebbie-Rayet[1]
[1]*Scientific Computing Department, Science and Technology Facilities Council, Daresbury Laboratory, Warrington, United Kingdom*
[2]*Department of Life Sciences, Imperial College London, London, United Kingdom*

Corresponding author: jas.kalayan@stfc.ac.uk

Abstract

Can interesting simulation methods presented in academic papers be applied to your scientific problem? What stops you from doing so? One reason is the high barrier to reproduce results from the information provided. Biomolecular simulation protocols are multifaceted and tailored to the biological system of interest, with a variety of software used to build and prepare molecular systems for simulations. Explicitly describing all simulation steps is unrealistic, time consuming and often inaccurate if written after the fact. Currently, no tools are available to track and collect steps as they are performed for biomolecular simulations, due to the complexity of workflows and the lack of standard practices in how to capture and disseminate them.

Here we present our software tools [1] to aid in full simulation data provenance collection with minimal impact to research time. Our aim is to develop user infrastructure [2], guided by the biomolecular simulation community, to promote sharing and reuse of data according to FAIR principles [3]. Examples of how simulation metadata can be tracked, saved, queried and shared is presented, using coarse-grained MD simulations of GPCRs as a case study.

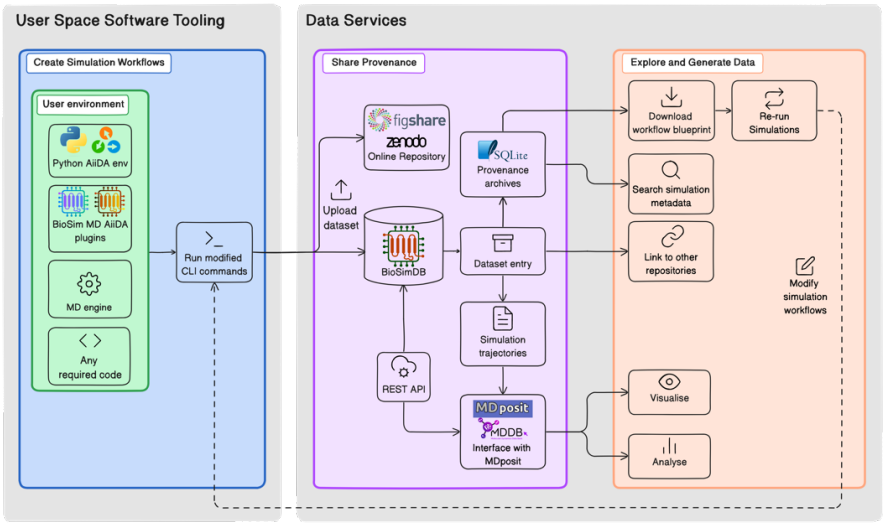


Figure 1: *Vision for Biomolecular Simulation Data Provenance Collection and Storage*

References
[1] https://aiida-gromacs.readthedocs.io
[2] https://www.psdi.ac.uk
[3] Wilkinson, Mark D. et al. (2016). Scientific Data, **3**(1), 160018.

L1 Metallo- β -Lactamase Antimicrobial Resistance Enzyme: A Computational Reaction Mechanism Study

Papu Kalita[1]; Marko Hanzevacki[1], Carlos A. Ramos-Guzmán[1], Philip Hinchliffe[2], James Spencer[2], and Adrian J. Mulholland[1]

[1] *School of Chemistry; University of Bristol; UK*

[2] *School of Cellular and Molecular Medicine, Infection, and Immunity; University of Bristol; UK*

Corresponding author: adrian.mulholland@bristol.ac.uk

Abstract

Antimicrobial resistance (AMR) poses a global health threat as the resistance of pathogenic bacteria to vital drugs. Particularly concerning is the widespread bacterial resistance to carbapenems, a class of "last resort" β-lactam antibiotics. An important and growing cause of resistance is L1, a metallo-β-lactamase (MBL) enzyme, which can break down nearly all types of β-lactam antibiotics, including carbapenems and penems. Notably, no clinical inhibitors have been discovered thus far that can effectively inhibit the hydrolysis process of L1. So, understanding the mechanisms through which these bacterial enzymes hydrolyse β-lactam antibiotics is crucial in combating this resistance.

To gain insights into the hydrolysis mechanism of faropenem (a penem) by L1 MBL, we employed combined quantum mechanics/molecular mechanics (QM/MM) simulations using DFTB3 and B3LYP/6-31G(d) QM methods, with umbrella sampling (US) and the adaptive string method (ASM). The simulations revealed that a zinc-bridging nucleophile, likely in the form of a hydroxide ion, plays a pivotal role in the hydrolysis process. The free energy curves analysed using the DFTB3 1D-US method indicated the presence of a single transition state. Furthermore, the ASM analysis suggested a concerted pathway for faropenem hydrolysis by L1. Calculations at the DFT level indicate that the protonation of anionic nitrogen of the ring most likely involves another water molecule that orients towards the nitrogen which facilitates the proton transfer. These simulations provide an atomic level analysis of the reaction mechanism of antibiotic breakdown. These findings contribute to our understanding of how β-lactam antibiotics are hydrolysed by L1 and may aid in the development of strategies to prevent and combat AMR, including the development of MBL inhibitors.

References

[1] Lucic, P. Hinchliffe, T. R. Malla, C. L. Tooke, J. Brem, K. Calvopiña, C. T. Lohans, P. Rabe, M. A. McDonough, T. Armistead, A. M. Orville, J. Spencer and C. J. Schofield, *Eur. J. Med. Chem.*, 2021, **215**, 113257.

[2] M. Twidale, P. Hinchliffe, J. Spencer and A. J. Mulholland, *J. Chem. Inf. Model.*, 2021, **61**, 5988-5999.

Developing Free Energy Calculation Protocols for Intrinsically Disordered Protein Regions

Audrius Kalpokas[1]; Mark Mackey[2]; Julien Michel[1]
[1]*EaStCHEM School of Chemistry, University of Edinburgh, Edinburgh EH9 3FJ, UK*
[2]*Cresset, New Cambridge House, Bassingbourn Road, Litlington, Cambridgeshire, SG8 0SS, UK*

Corresponding author: audrius.kalpokas@ed.ac.uk

Abstract

Free Energy Perturbation (FEP) methodologies are frequently used in industrial settings to optimise the binding affinity of small molecular ligands to protein binding sites. FEP protocols for modelling protein modifications, in particular side-chain mutations, are less established. In particular, protein FEP protocols are currently challenging to apply to proteins that undergo large conformational changes upon mutations. Here we focus on developing robust protein FEP protocols to reproduce experimental binding affinities that have been measured for a panel of mutants of the protein MDM2 against two ligands, AM-7209 and Nutlin-3a. We focus on mutations that occur in the intrinsically disordered lid region of MDM2, that is known to undergo ligand-dependent folding upon binding.[1,2]

We systematically assess the effectiveness of both equilibrium and non-equilibrium FEP protocols in reproducing experimental binding affinities in particular for mutations with slowly-varying degrees of freedom. Our approach paves the way for robust application of FEP protocols to model drug resistance and thermostability in conformational flexible protein regions.

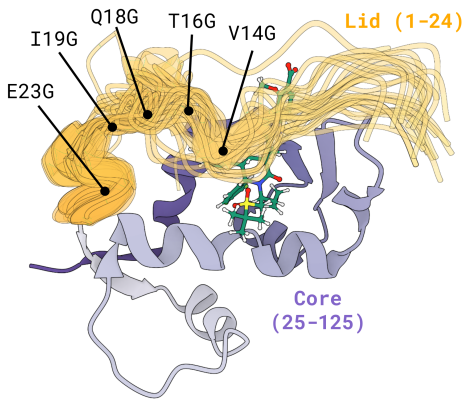


Figure 1: *MDM2 and the intrinsically disordered lid mutations explored in this work.*

References
[1] Bueren-Calabuig, Juan A., and Michel, Julien. (2015). *PLOS Computational Biology*, **11**(6).
[2] Mendoza-Martinez, Cesar, Gupta, Arun, and Michel, Julien. (2024) *Molecular Driving Force of a Small Molecule-Induced Protein Disorder-Order Transition. ChemRxiv.*

On a quest to improve enzymatic Diels – Alder reaction modelling – comparison of semi-empirical methods

Magdalena Kazmierczak[1]; Marc Van der Kamp[1]
[1]*School of Biochemistry, Faculty of Life Sciences, University of Bristol, Biomedical Sciences Building, University Walk, Bristol BS8 1TD*

Corresponding author: al23945@bristol.ac.uk

Abstract

AbyU is a powerful Diels-Alderase, which has been proven to be mutable, and also adaptable to changes within the substrate.[1–3] This highly active and stereoselective enzyme withstands harsh conditions, such as elevated temperatures and organic solvents, marking it as a potential biocatalyst for industrial applications.[4] Previously, for efficient Diels-Alder reaction modelling in AbyU, QM/MM umbrella sampling simulations were performed with DFTB as a QM method. The activation free energies obtained show acceptable consistency with higher level QM methods (DFT) and experiment.[1,2] Nevertheless, there is room for improvement in terms of the detailed reaction path followed. Here, the semi-empirical methods AM1 and PM6 were tested as the potential candidates for such an improvement.

In the Diels-Alder reaction, two new carbon-carbon bonds are forming. To allow efficient sampling, the modelling process was previously simplified by combining the two bond lengths with appropriate weights as a single reaction coordinate for umbrella sampling[1]. Here, a range of weights for 4 different substrate poses (with different reactivities) were used in an attempt to optimize weights specifically for AM1 and PM6. Although the resulting free energy profiles agree with the previous results to some extent, there are also clear differences between them, especially regarding the most probable active pose. 2D umbrella sampling (one reaction coordinate for each bond) does not resolve this disagreement. Therefore, alternative methods to AM1 and PM6 may need to be considered in this instance.

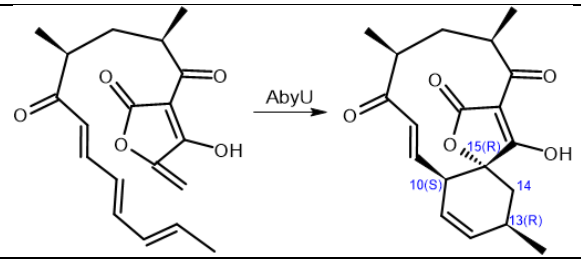


Figure 1: *Schematic representation of the Diels-Alder reaction catalysed by AbyU*

References
1. Johns, S. Towards simulation-led engineering of a natural Diels-Alderase. (2021).
2. Byrne, M. J. et al. The Catalytic Mechanism of a Natural Diels-Alderase Revealed in Molecular Detail. J Am Chem Soc 138, 6095–6098 (2016).
3. Maschio, L. et al. Delineation of the Complete Reaction Cycle of a Natural Diels-Alderase. doi:10.1101/2024.04.18.590041.
4. Marsh, C. O. et al. A Natural Diels-Alder Biocatalyst Enables Efficient [4+2] Cycloaddition Under Harsh Reaction Conditions. ChemCatChem 11, 5027–5031 (2019).

Structural Dynamics Analysis of Nitrite Reductase AniA Asal

Asal Khoshravan Azar[1]; Matteo Degiacomi[1]

[1] *Rochester Building, Lower Mountjoy, South Rd, Durham DH1 3LE*

Corresponding author:asal.khoshravan-azar@durham.ac.uk

Abstract

Metalloenzymes are crucial in numerous biological processes, yet the structural dynamics underpinning their function remain incompletely understood. Our research focuses on the Copper- containing nitrite reductase (nirK) where the Copper-loaded form is a trimer, and the apo form is a monomer. Trimer formation, dependent on copper loading, is a slow process in vitro and appears to be irreversible. However recent experimental results also show that, surprisingly, Copper can then be removed from a formed trimer without compromising its stability.

In this work, we gain insight into the trimerization process of nirK by characterizing the dynamics of its monomeric and multimeric states via molecular dynamics simulations. Our results confirm that the trimer does not require copper to be stable. We also find that in the microsecond timescale the unstructured C-terminal region of the monomer transiently occupies conformations that may make it binding-competent. We hypothesise that this region may act as a "fishing rod", ultimately latching to an adjacent subunit in the multimeric state. These insights enhance our understanding of the enzyme's activation mechanism, testable in future experimental studies via targeted mutagenesis.

De Novo Design of GALK1 Inhibitors in a Flexible Binding Pocket

Asma Feriel Khoualdi [1]; Matteo Degiacomi[2]; Wyatt Yue[3]; Daniel J. Cole[1]
[1] *School of Natural and Environmental Sciences, Newcastle University, Newcastle-upon-Tyne, NE1 7RU, UK*
[2] *Department of Physics, Durham University, DH1 3LE, UK*
[3] *Newcastle University Medical School, Newcastle-upon-Tyne, NE2 4HH, UK*

Corresponding author: a.f.khoualdi2@newcastle.ac.uk

Abstract

Galactosemia is a rare autosomal recessive disease caused by a defect in galactose metabolism. GALK1 is a kinase enzyme involved in galactose metabolism and whose inhibition can help alleviate galactosemia[1]. By making use of the GALK1's allosteric inhibitors which are crystalized[1], molecules with better binding properties can be designed, modified and optimized within the binding pocket using FEgrow tool[2]. Molecules can be designed interactively or automatically using substructures from databases, built and scored using FEgrow.

Pocket flexibility is addressed with the conformation prediction tool Molearn[3] which uses a convolutional neural network to learn from relatively short, example MD simulation trajectories to predict the different conformations the protein will have. Designing inhibitors that fit the predicted ensemble of pocket conformations means more potent inhibitors can be generated and tested.

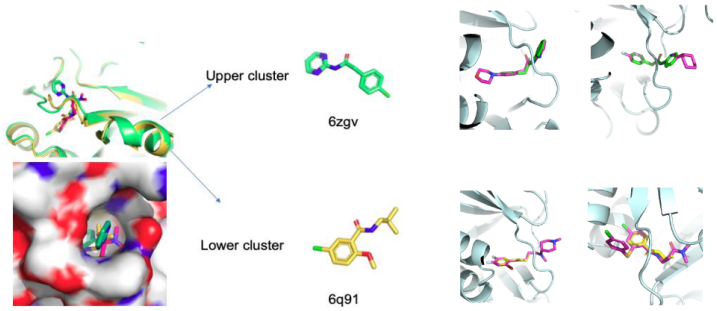


Figure 1: GALK1 upper and lower cluster starting molecules (left) from which the molecules shown in pink (right) are designed in the pocket

References

[1] Mackinnon, S.R., et al., Fragment Screening Reveals Starting Points for Rational Design of Galactokinase 1 Inhibitors to Treat Classic Galactosemia. ACS Chemical Biology, 2021. **16**(4): p. 586-595.
[2] Bieniek, M.K., et al., An open-source molecular builder and free energy preparation workflow. Communications Chemistry, 2022. **5**(1): p. 136.
[3] S.C. Musson and M.T. Degiacomi (2023). Molearn: a Python package streamlining the design of generative models of biomolecular dynamics. Journal of Open Source Software, **8**(89), 5523

Towards Predicting Conformational Ensembles of Intrinsically Disordered Proteins with Generative Neural Networks

Marco Mattia[1]; M. T. Degiacomi[2]; A. S. J. S. Mey[1]
[1]*EaStCHEM School of Chemistry, University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ, United Kingdom*
[2]*Department of Physics, Durham University, South Road, Durham, DH1 3LE, United Kingdom*

Marco Mattia: marco.mattia@ed.ac.uk

Abstract

Proteins are flexible biopolymers essential to all biological processes, with their collective conformational ensemble dictating their biological function. This dynamic nature is particularly pronounced in intrinsically disordered proteins (IDPs), which, despite exhibiting a higher degree of conformational variability compared to folded proteins, play crucial roles as promiscuous binders. This capacity allows them to participate in a wide range of cellular interactions, making them central to the molecular mechanisms of various diseases, including cancers and neurodegenerative disorders. Given the challenges associated with direct experimental observation of these dynamics, utilizing experimentally determined structures as starting points for molecular dynamics (MD) simulations has become a standard approach [1]. Still, sampling protein conformational dynamics at biological timescales with unbiased MD simulations alone is a highly computationally intensive task, especially for proteins featuring broad dynamics such as IDPs.

Our aim is to combine molecular simulations and generative neural networks (GNNs), a machine learning method, to predict the conformations of IDPs. To reach this goal we use molearn [2], a framework for training GNNs on protein conformational spaces. While a GNN implemented in molearn has previously been shown capable of learning the dynamics of folded proteins [3] it is unknown whether it will succeed in learning the dynamics of IDPs. To test this, we have trained the standard GNN in molearn with 4 μs of MD we produced for the IDP α-synuclein. We find that the GNN's performance with this dataset is worse than what previously observed for folded proteins, with lower performances obtained for extended conformations than for more compact (folded) ones. This suggests that additional strategies are needed to capture the conformational space of IDPs through machine learning approaches that go beyond the current GNN implementations available in molearn.

References

[1] Schor, Marieke, Mey, Antonia S. J. S. and MacPhee, Cait E. (2016). *Biophysical Reviews*, **8**(4), 429-439.
[2] Musson, Samuel C. and Degiacomi, Matteo T. (2023). *Journal of Open Source Software*, **8**(89), 5523.
[3] Ramaswamy, Venkata K., Musson, Samuel C., Willcocks, Chris G., and Degiacomi, Matteo T. (2021). *Physical Review X*, **11**(1), 011052.

Computing hydration free energies of small molecules with first principles accuracy

Harry Moore[1]; Daniel J. Cole[2], Gabor Csanyi[1,2]
[1]*Engineering Laboratory, University of Cambridge, Cambridge, CB2 1PZ, UK*
[2] *School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK*

Corresponding author: jhm72@cam.ac.uk

Abstract

Free energies play a central role in characterising the behaviour of chemical systems and are among the most important quantities that can be calculated by molecular dynamics simulations. The free energy of hydration in particular is a well studied physicochemical property of drug-like molecules and is commonly used to assess and optimise the accuracy of nonbonded parameters in empirical forcefields, and as a fast-to-compute surrogate of performance for protein-ligand binding free energy estimation. Machine learned potentials (MLPs) show great promise as more accurate alternatives to empirical forcefields, but are not readily decomposed into physically motivated functional forms, which has thus far rendered them incompatible with standard alchemical free energy methods that manipulate individual pairwise interaction terms. However, since the accuracy of free energy calculations is highly sensitive to the forcefield, this is a key area in which MLPs have the potential to address the shortcomings of empirical forcefields. In this work, we introduce an efficient alchemical free energy method compatible with MLPs, enabling, for the first time, calculations of biomolecular free energy with ab initio accuracy. Using a pretrained, transferrable, alchemically equipped MACE model, we demonstrate sub-chemical accuracy for the hydration free energies of organic molecules.

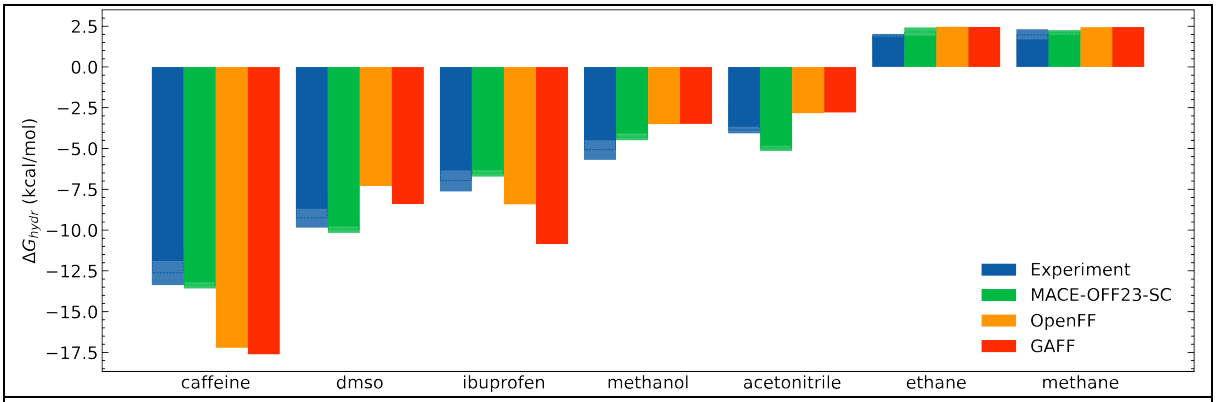


Figure 1: *Comparison of MACE-OFF23-SC hydration free energies with classical forcefields and experiment*

Electrostatically-Embedded Hybrid ML/MM Simulations

João Morado[1]; Lester Hedges[2]; Josh Horton[3]; Daniel J. Cole[3]; Julien Michel[1]

[1]*EaStCHEM School of Chemistry, University of Edinburgh, United Kingdom*
[2] *OpenBioSim Community Interest Company, United Kingdom*
[3] *School of Natural and Environmental Sciences, Newcastle University, United Kingdom*
Corresponding author: jmorado@ed.ac.uk

Abstract

The landscape of molecular models used in biomolecular simulations is rapidly evolving. While molecular mechanics (MM) simulations have traditionally relied on complex functional forms or optimised force field (FF) parameters for accuracy, the advent of machine learning potentials (MLPs) has opened new horizons in this field. MLPs offer the potential to achieve quantum-level accuracy at a computational cost near that of MM methods. As a result, new methodologies are emerging to seamlessly integrate MLPs into existing simulation frameworks.

In condensed-phase biomolecular simulations, mechanically-embedded MLP/MM schemes have been predominantly used [1]. These hybrid simulations leverage MLPs to accurately describe the intramolecular energetics of a specific region, while treating the remainder of the system and MLP/MM interactions at the MM level. To enhance this static formulation and include polarisation effects, various electrostatic embedding schemes have been recently developed. Among these, the electrostatic ML embedding (EMLE) stands out as a promising model for capturing the response of an MLP system to MM point charges [2]. EMLE can be used on top of existing MLPs, and it combines static and induced physics-informed potentials with an ML model that predicts the required EMLE parameters.

In this poster, we present absolute alchemical solvation free energy calculations for a set of small organic molecules using mechanically- and electrostatically-embedded MLP/MM simulations. We benchmark the accuracy of employing a new-generation MACE-OFF23[3] MLP in these settings and compare the results against pure MM approaches. Our findings offer insights into the additional physics captured by the EMLE potential and suggest potential research avenues to enhance the accuracy of hybrid MLP/MM simulations.

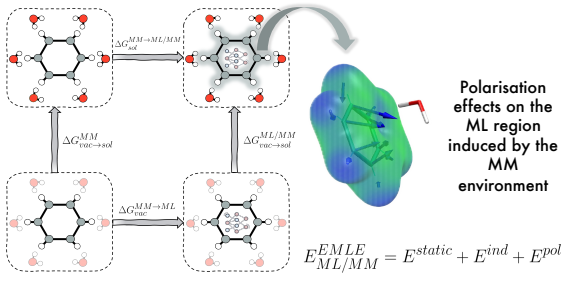


Figure 1: *Schematic of the thermodynamic cycle used to calculate electrostatically-embedded hydration free energies.*

References

[1] Rufa, D. A, Bruce Macdonald, H. E., Fass, J., Wieder, M., Grinaway, P. B., Roitberg, A. E., Isayev, O., Chodera, J. D. (2020). *bioRxiv.*
[2] Zinovjev, K. (2023). *J. Chem. Theory Comput*. 2023, **19**(6), 1888–1897.
[3] Kovács, D. P., Moore, J. H., Browning, N. J., Batatia, I., Horton, J. T., Kapil, V., Witt, W. C., Magdău, I.-B., Cole, D. J., Csányi, G. (2023). *arXiv.*

Untangling the Concentrations of Condensates in vitro

William Morton[1]; Marek Šebesta[1], Robert Vácha[1,2,3], Richard Štefl[1,3]

[1] *CEITEC—Central European Institute of Technology, Masaryk University, Brno 62500, Czech Republic*
[2] Department of Condensed Matter Physics, Faculty of Science, Masaryk University, Brno 62500, Czech Republic
[3] National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno 62500, Czech Republic
[4] National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czechia
Corresponding author: william.morton@ceitec.muni.cz

Abstract

The interaction between RNA polymerase II (Pol II) and transcription factors (TFs) is essential for proper cellular function. TFs act as controllers and aids for Pol II's transcription of RNA. Many TFs can work in unison with Pol II, forming condensates where the interaction between disordered regions precisely controls the size and makeup of these molecular hubs[1]. Our understanding of species' spatial organization and concentration within the condensates is limited by the microscopic resolution of densely packed proteins[2]. Here we focus on RECQL5, a TF known to be present during the elongation phase of transcription when Pol II is hyperphosphorylated[3]. *In vitro* mixtures of proteins were recreated *in silico*, giving insight to their interaction network. The disordered region of RECQL5 is responsible for localization in a homotypic condensate. The uMol affinity between phosphorylated serine and RECQL5's SRI domain increases the valency of the heterotypic condensate and may be responsible for recruiting Pol II.

Funding: This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101180586, in association with funding from GAČR.

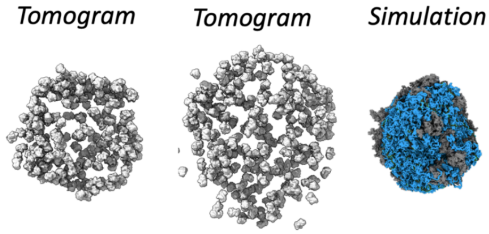


Figure 1: *Tomograms of Pol II in a heterotypic condensate with RECQL5, compared to simulation results of a condensate with 60 copies of Pol II.*

References

1. Shao, W., et al , *Nature Chem. Bio* **2022**, 18, 70-80
2. Ferrie, J. J., Karr, J. P., Tjian, R., & Darzacq, X. *Molecular Cell* **2022**. 82(21), 3970–3984 3. Saponaro, M.et al. *Cell* **2014,** 157(5), 1037-1049

Prediction of Protein Oligomerisation with Metadynamics Simulations

Evangelia Notari[1]; Christopher W. Wood[2]; Julien Michel[1]
[1] *EaStCHEM School of Chemistry, University of Edinburgh, Edinburgh EH9 3FJ, UK*
[2] *School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FF, UK*

Corresponding author: e.notari@sms.ed.ac.uk

Abstract

Proteins found in Nature have only sampled a small fraction of the available protein sequence space to solve problems faced during evolution. *De novo* protein design has so far mostly focused on design algorithms that maximise the free energy difference between a single conformational state and other plausible macrostates, resulting in rigid and hyper-stable folds that lack functionality. However, proteins often need to interconvert between different conformational states in order to execute their functions. Future progress in the field of *de novo* protein design depends on advances in designing multi-state proteins, i.e. proteins that can switch between different conformations upon an external perturbation. Coiled coils are an attractive scaffold for *de novo* protein design, given that they are one of the few folds in Nature whose structure can be accurately described with simple parametric equations. These clear sequence-to-structure rules can assist in the selection of sequences that can assemble into a desired oligomerisation state, e.g. dimer, trimer, tetramer etc. However, unexpected oligomerisation states can often be adopted upon external stimuli, e.g. a change in the pH, disagreeing with simple design rules.[1] It is therefore hypothesised that the fine balance of intermolecular interactions that dictates the adoption of multiple conformational states could be sufficiently captured with atomistic molecular dynamics simulations.

In our work, we are developing a computational pipeline for the prediction of the oligomerisation and conformational preferences of coiled coils. To this end, we couple parametric design and metadynamics simulations with a combination of conformational and orientational restraints, in order to estimate the free energy of aggregation of coiled coil designs. We show that our pipeline is able to distinguish between different topologies (e.g. parallel vs antiparallel) and oligomerisation states (e.g. dimer vs trimer), rank designs with different sequences, as well as model structural stability in different pH conditions, in agreement with experimental observations. Given the versatility of insights our protocol can provide, we propose it can be employed to compute the free energy of binding of other protein-protein or multiprotein complexes of interest to pharmaceutical applications.

References

[1] Rhys, Guto G., Wood, Christopher W., Beesley, Joseph L., Zaccai, Nathan R., Burton, Antony J., Brady, R. Leo, Thomson, Andrew R. and Woolfson, Derek N. (2019). *Journal of the American Chemical Society*, **141**(22), 8787-8797.

Discovery of Natural Products as M. tuberculosis DNA Gyrase ATPase Inhibitors

Bongkochawan Pakamwong[1], Paptawan Thongdee[1], Naruedon Phusi[1], Somjintana Taveepanich[1], Kampanart Chayarus[1], Jidapa sangswan[2], Auradee Punkvang[3], Khomson Suttisintong[4], Sanya Sureram[5], Prasat Kittakoop[5,6,7], Poonpilas Hongmanee[8], Pitak Santanirand[8], Jiraporn Leanpolchareanchai[9], James Spencer[10], Adrian Mulholland[11] and Pornpan Pungpo[1*]

[1]*Department of Chemistry and Center of Excellence for Innovation in Chemistry, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, 34190, Thailand*

[2]*Department of Biological Science, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, 34190, Thailand*

[3]*Division of Chemistry, Faculty of Science, Nakhon Phanom University, Nakhon Phanom, 48000, Thailand*

[4]*National Nanotechnology Center, NSTDA, Thailand Science Park, Klong Luang, Pathum Thani, 12120, Thailand*

[5]*Chulabhorn Research Institute, Bangkok, 10210, Thailand*

[6]*Chulabhorn Graduate Institute, Chemical Biology Program, Chulabhorn Royal Academy, Bangkok, 10210, Thailand*

[7]*Center of Excellence on Environmental Health and Toxicology (EHT), OPS, Ministry of Higher Education, Science, Research and Innovation, Bangkok, 10210, Thailand*

[8]*Division of Clinical Microbiology, Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand*

[9]*Department of Pharmacy, Faculty of Pharmacy, Mahidol University, Bangkok 10400, Thailand*

[10]*School of Cellular and Molecular Medicine, Biomedical Sciences Building, University of Bristol, Bristol BS8 1TD, United Kingdom*

[11]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*

Corresponding author: pornpan_ubu@yahoo.com

Abstract

Tuberculosis, caused by the *Mycobacterium tuberculosis* (*M. tuberculosis*), is a communicable disease and one of the leading causes of death worldwide. *M. tuberculosis* DNA gyrase is a validated target of fluoroquinolones (FQs), second-line antibiotics used for the treatment of multidrug-resistant tuberculosis [1-2]. Mutations in DNA gyrase confer resistance to fluoroquinolones. Thus, the identification of new agents that inhibit *M. tuberculosis* DNA gyrase ATPase activity is one strategy to overcome this. Here, structure-based virtual screening, subsequently validated by biological assays, was applied to identify candidate compounds from natural products in the Specs compound library as *M. tuberculosis* GyrB ATPase inhibitors. Eight compounds were identified as hit compounds for *in vitro* biological assays. Four of eight compounds (**1**, AE–562/12222575; **2**, AE–562/43459688; **6**, AO–103/41868968; and **7**, AA–504/20999031) inhibited the inhibition of *M. tuberculosis* DNA gyrase ATPase activity with IC_{50} values of 18.39, 19.86, 16.31, and 18.93 μM, respectively. These compounds showed nontoxicity to Caco-2 cells at concentrations up to 20-fold, 18-fold, 25-fold, and 10-fold higher than their IC_{50} value, respectively. Further, the binding modes and binding energy of compounds **1**, **2**, **6**, and **7** were investigated by molecular dynamics (MD) simulations. These compounds provide attractive starting templates for the optimization of anti-tuberculosis agents.

References

[1] Asif, Mohammad, Siddiqui, Anees A. and Husain, Asif. (2013). *Medicinal Chemistry Research*, **22**, 1029–1042.

[2] Liu, Kun-Lin, Teng, Fei, Xiong, Lu, Li, Xiao, Gao, Chao and Yu, Luo-Ting. (2021). *Royal Society of Chemistry*, **11**, 24095–24115.

Multiscale Cell Correlation Entropy of the Streptavidin-Biotin System

Ioana A. Papa[1,2], Sarah K. Fegan[1], Jas Kalayan[1], Richard H. Henchman[3], Sarah A. Harris[4]
[1] *Scientific Computing Department, STFC, Daresbury Lab, UK*
[2] *Department of Chemistry, University of Edinburgh, UK*
[3] *School of Medicine, University of Sydney, Australia*
[4] *Department of Physics, University of Sheffield, UK*

Corresponding author: I.Papa@sms.ed.ac.uk

Abstract

The accurate calculation of binding free energies is crucial in medicinal chemistry as it contributes to assessing and tuning properties of drug molecules such as selectivity, specificity and off-target effects. Approaches developed for their quantification include energy-entropy methods, which calculate binding free energies from the two components. There have been various computational methods developed for computing the entropy of a system, but this quantity is generally difficult to compute in a manner that exhibits a good balance between computational cost and accuracy. Furthermore, the accuracy of such approaches is difficult to assess due to limited experimental methods that can help validate computational methodologies.

Multiscale cell correlation [1], a strategy which relies on decomposing molecules into sets of beads at different levels of hierarchy, is able to provide a detailed breakdown of contributions to the binding entropy of protein-ligand systems. The current release of CodeEntropy, gave results that show a good agreement with quasielastic neutron scattering data for the streptavidin-biotin system [2]. An in-development version of CodeEntropy was found to be more suitable for the analysis of small organic molecules, better optimized for larger systems and able to yield more accurate conformational entropies.

References

[1] Chakravorty, Arghya; Higham, Jonathan and Henchman, Richard H. (2020). *J. Chem. Inf. Model.*, **60**, 5540−5551.
[2] Sarter, Mona; Niether, Doreen; Koenig, Bernd. W; Lohstroh, Wiebke; Zamponi, Michaela; Jalarvo, Nina H.; Wiegand, Simone; Fitter, Jörg and Stadler, Andreas M. (2020) J. Phys. Chem. B, 2020, **124**, 324–335.

MemPrO: Membrane Protein Orientation in lipid bilayers

Matyas Parrag[1]; Phillip J. Stansfeld[1],
[1]*University of Warwick, Coventry, CV4 7AL, UK*

Presenting author: M.Parrag@warwick.ac.uk

Abstract

Membrane proteins play an important role in many vital systems of a cell, such as transport of ions and raw materials, communication between adjacent cells, and antibiotic resistant behaviours. Correctly orienting membrane proteins is almost always the first step in the molecular simulation and analysis of membrane-protein systems. The method presented aims to orient a wide range of proteins that interact with the membrane. Many such programs already exist such as OPM [1] and MemEmbed [2], however these do not work for some situations such as multi-bilayer systems or peripheral membrane proteins. MemPrO also contains tools for further analysis of membrane-protein systems without the need for simulations to be run. Currently membrane deformation and localisation of negatively charged lipids can be predicted. Figure 1 shows deformation and localisation predictions compared to simulation and experimental data respectively. The core method consists of a minimisation in a mean field of potential constructed using Martini3 [4] CG parameters.

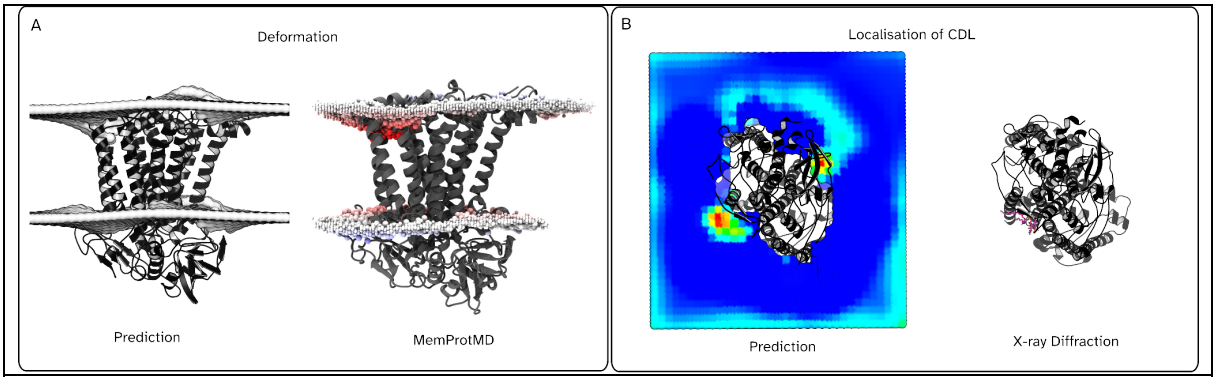


Figure 1: *The protein 7MHA was oriented using MemPrO. The resulting orientation was then used to predict the membrane deformation (A) and the localisation of cardiolipin (B). In A the deformation prediction is compared to MemProtMD simulation data [3] and in B the localisation prediction is compared to X-ray diffraction data from RCSB.*

References

[1] Lomize, Andrei et al. (2009). *Protein Science,* **Volume 15** (Issue 6) , 1318-1333
[2] Nugent, Timothy, Jones, David. (2013). *BMC Bioinformatics*, **Membrane protein orientation and refinement using a knowledge-based statistical potential.**
[3] D Newport, Thomas. (2018). *Nucleic Acids Research,* **Volume 47**, 390-397
[4] Souza, Paulo et al. (2021), *Nature Methods,* **Volume 18**, 382-388.

Psychedelic, Antipsychotic and Antidepressant effects of 5-HT2A modulation

Jordy Peeters[1]; Kenno Vanommeslaeghe[1]
[1] *Vrije Universiteit Brussel (VUB) Laarbeeklaan 103, 1090 Brussels, Belgium*

Corresponding author: jordy.peeters@vub.be

Abstract

The serotonin 2A-receptor (5-HT2AR) is best known for inducing hallucinations when activated by hallucinogenic ligands like LSD, psilocin or DMT.[1,2] However, it is also of high and increasing interest for the treatment of mentally linked afflictions such as depressions, ADHD and schizophrenia. The 5-HT2AR belongs to the family of the G protein-coupled receptors (GPCRs), but its two distinct effects suggest different intracellular signalling. This ability is called biased agonism (or functional selectivity) and is hypothesized to apply to 5-HT2AR.[3] Indeed, recent studies have shown that activating the receptor with certain agonists may give rise to antidepressant action without psychedelic side effects.[4,5]

Although a lot of research has been done on GPCRs in general, on an atomistic level the mechanisms behind the 5-HT2AR's differential activation remain unclear. In our study we performed multiple atomistic molecular dynamics (MD) simulations on 5-HT2AR bound with an antipsychotic, two different potential non-hallucinogenic antidepressants and a hallucinogen to identify the receptor's ligand dependent conformations. Additionally, we used an in-house variant of Adaptive Biasing Force (ABF), a free energy method, on these systems to energetically map these different conformational landscapes and hence explore their conformational freedom. Given the current gap in specialized pharmacological profiles for agonists, our ultimate goal is to develop new types of specific agonists to advance the design of novel therapeutic drugs.

References

[1] Nichols, David E. (2004). *Pharmacol Ther*, **101**(2), 131-181.
[2] Nichols, David E. (2016). *Pharmacol Ther*, **68**(2), 264-355.
[3] Duan, Wenwen, Cao, Dongmei, Wang, Sheng and Cheng, Jianjun. (2024). *Chem Rev*, **124**(1), 124-163.
[4] Kaplan, Anat Levit, Confair, Danielle N., et al, and Ellman, Jonathan A. (2022). *Nature*, **610**(7932), 582-591.
[5] Cao, Dongmei, Yu, Jing, et al, and Wang, Huan. (2022). *Science*, **375**(6579), 403-411.

Insight into the specific interactions between benzimidazole derivatives and InhA via ab initio fragment molecular orbital calculations and molecular dynamics simulations

Thimpika Pornprom[1], Naruedon Phusi[1], Paptawan Thongdee[1], Bongkochawan Pakamwong[1], Kampanart Chayarus[1], Jidapa sangswan[2], Auradee Punkvang[3], Khomson Suttisintong[4], Prasat Kittakoop[5,6,7], Noriyuki Kurita[8], James Spencer[9], Adrian Mulholland[10] and Pornpan Pungpo[1*]

[1]*Department of Chemistry and Center of Excellence for Innovation in Chemistry, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, 34190, Thailand*
[2]*Department of Biological Science, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, 34190, Thailand*
[3]*Division of Chemistry, Faculty of Science, Nakhon Phanom University, Nakhon Phanom, 48000, Thailand*
[4]*National Nanotechnology Center, NSTDA, Thailand Science Park, Klong Luang, Pathum Thani, 12120, Thailand*
[5]*Chulabhorn Research Institute, Bangkok, 10210, Thailand*
[6]*Chulabhorn Graduate Institute, Chemical Biology Program, Chulabhorn Royal Academy, Bangkok, 10210, Thailand*
[7]*Center of Excellence on Environmental Health and Toxicology (EHT), OPS, Ministry of Higher Education, Science, Research and Innovation, Bangkok, 10210, Thailand*
[8]*Department of Computer Science and Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi, Aichi, 441-8580, Japan*
[9]*School of Cellular and Molecular Medicine, Biomedical Sciences Building, University of Bristol, Bristol BS8 1TD, United Kingdom*
[10]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*

Corresponding author: pornpan_ubu@yahoo.com

Abstract

The *2-trans* enoyl-acyl carrier protein reductase (InhA) presents a compelling target for the development of novel chemotherapy agents to combat tuberculosis. Extensive research has been conducted to explore the inhibitory effects of these agents on InhA activity through various physicochemical experiments. In the present study, we first designed novel compounds based on ligand X-ray structure that obtained from our previous study (PDB code: 6R9W) as well as benzimidazole derivatives. To elucidate the binding energies and the specific interactions between InhA and our designed derivatives using *ab initio* fragment molecular orbital (FMO) and Waterswap calculations. The key interactions between InhA and the ligand X-ray structure were found with water molecule, Gln100, Ala157. Therefore, we designed novel compounds by replacing H atom at the R1, R2, or R3 site of the 2,3-dihydro-1*H*-indene ring of ligand X-ray structure to enhance specific interactions. The results evaluated by FMO highlight some key interactions between InhA and the derivatives, indicating that the most potent derivative has strong interaction with a water molecule, Gln100, Ala157, Ile215, and NAD^+. In addition, as compared with the conformation of ligand obtained from the X-ray crystallographic data, the enhancement of the size of R2 substitution results in the changes in total IFIE from −103.1 to −118.2 kcal/mol. Models of complexes of novel compound bound to the *M. tuberculosis* InhA-binding site, generated by molecular dynamics (MD) simulations. Moreover, MD simulations were conducted to explore the binding energies and interactions of novel compound within the InhA binding pocket. Finally, these findings offer valuable insights for the development of novel compounds with anti-tuberculosis properties by effectively inhibiting InhA activity.

References

[1] Kamsri, Pharit, Hanwarinroj, Chayanin, Phusi, Naruedon, Pornprom, Thimpika, Chayajarus, Kampanart, Punkvang, Auradee, Suttipanta, Nitima, Srimanote, Potjanee, Suttisintong, Khomson, Songsiriritthigul, Chomphunuch, Saparpakorn, Patchreenart, Hannongbua, Supa, Rattanabunyong, Siriluk, Seetaha, Supaporn, Choowongkomon, Kiattawee, Sureram, Sanya, Kittakoop, Prasat, Hongmanee, Poonpilas, Santanirand, Pitak, Chen, Zhaoqiang, Zhu, Weiliang, Blood, Rosemary A, Takebayashi, Yuiko, Hinchliffe, Philip, Mulholland, Adrian J, Spencer, James, Pungpo, Pornpan. (2020). *Journal of Chemical Information and Modeling*, **60(1)**, 226–234.

[2] Phusi, Naruedon, Hashimoto, Yuta, Otsubo, Naoki, Imai, Kyohei, Thongdee, Paptawan, Sukchit, Darunee, Kamsri, Pharit, Punkvang, Auradee, Suttisintong, Khomson, Pungpo, Pornpan, Kurita, Noriyuki. (2023). *Computers in Biology and Medicine*, **152**, 106434.

Drug Repurposing for Mammalian Heart Regeneration

Ignasi Puch-Giner[1]; Victor Guallar[1,2].

[1] *Barcelona Supercomputing Center, Plaça Eusebi Güell, 1-3, 08034 Barcelona*
[2] *Nostrum Biodiscovery, Avinguda Josep Tarradellas, 8-10, 3-2, 08028 Barcelona {ignasi.puchginer,victor.guallar}@bsc.es*

Abstract

The newborn mammalian heart has an inherent regenerative capacity through the replication of pre- existing cardiomyocytes. This capacity diminishes shortly after birth[1], leaving the adult mammalian heart with limited self-repair capacity following injury[2,3]. The transcription factors Meis1 and Hoxb13 cooperatively induce cell cycle arrest[4], resulting in the loss of this regenerative capacity. Structure-based drug repurposing screening identified FDA-approved drugs that can inhibit their transcriptional activity. These drugs, named Neomycin and Paromomycin, demonstrated a dose- dependent inhibition and significant proliferation of neonatal cardiomyocytes in vitro and in vivo[5]. In this study, we used HPC resources to explore how these small molecules disrupt the formation of the Meis1-Hoxb13-DNA trimer (Fig. 1, Image A), halting transcriptional activity. We proposed three hypotheses: (1) The ligands attach to the proteins' DNA-binding alpha helices, preventing DNA binding; (2) The ligands promote homodimerization of the proteins, halting transcription activity; (3) The ligands bind to DNA, preventing transcription factor attachment. The roadmap to prove or discard these hypotheses can be seen in Fig. 1, Image B. Our protein surface exploration revealed relatively small binding energies for all Paro/Neo-Meis1/Hoxb13 combinations. Neomycin and Meis1 exhibited the best binding energies. Simulations with PELE[6] and GROMACS[7] showed strong and lasting Neomycin-Meis1 interactions in the parallel alpha helices region, discarding the first hypothesis. Dimerization simulations indicated that Neomycin enhanced Meis1-Meis1 homodimer binding energies, but the DNA-binding alpha helices of both monomers remained available to bind DNA. DNA surface simulations showed that Neo and Paro increased the binding energy twofold compared to Meis1 or Hoxb13. The best binding-energy poses were located in the transcription factors region of binding to DNA. With these results, the DNA-blocking hypothesis is the only one that agrees with both in-silico and experimental observations.

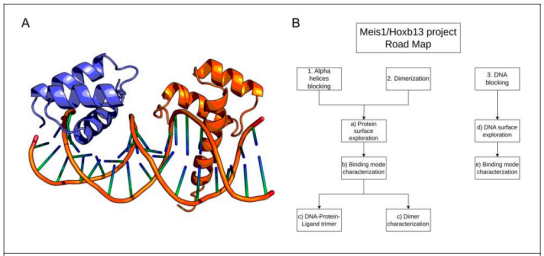


Figure 1: *Image A shows the structure of the Meis1-Hoxb13-DNA trimer (5EGO PDB[8]) from which we have based our in-silico simulations. Image B shows the road map we have followed to test out all the possible hypotheses.*

References

1. Soonpaa, M. H., & Field, L. J. (1997). Assessment of cardiomyocyte DNA synthesis in normal and injured adult mouse hearts.American Journal of Physiology-Heart and Circulatory Physiology H220-H226

2. Senyo, S. E., Steinhauser, M. L., Pizzimenti, C. L., Yang, V. K., Cai, L., Wang, M., ... & Lee, R. T. (2013). Mammalian heart renewal by pre-existing cardiomyocytes.Nature, 493(7432), 433-436.

3. Bergmann, O., Bhardwaj, R. D., Bernard, S., Zdunek, S., Barnabé-Heider, F., Walsh, S., ... & Frisén, J. (2009). Evidence for cardiomyocyte renewal in humans.Science, 324 (5923), 98-102.

4. Nguyen, N. U. N., Canseco, D. C., Xiao, F., Nakada, Y., Li, S., Lam, N. T., ... & Sadek, H. A. A calcineurin–Hoxb13 axis regulates growth mode of mammalian cardiomyocytes.Nature, 582(7811), 271-276.

5. Ahmed, M. S., Nguyen, N. U. N., Nakada, Y., Hsu, C. C., Farag, A., Lam, N. T., ... & Sadek, H. A. (2024) Identification of FDA-approved drugs that induce heart regeneration in mammals. Nature, Cardiovascular Research, 1-17

6. Borrelli, K. W., Vitalis, A., Alcantara, R., & Guallar, V. (2005). PELE: protein energy landscape exploration. A novel Monte Carlo based technique. Journal of chemical theory and computation, 1(6), 1304-1311.

7. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. Journal of computational chemistry

8. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., ... & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science, 26 (16), 1701-1718. Science 356 (6337), eaaj2239.

Identification of Natural products as potential M. Tuberculosis PknB inhibitors using structure-based virtual screening validated by biological assay

Pornpan Pungpo[1], Bongkochawan Pakamwong[1], Naruedon Phusi[1], Somjintana Taveepanich[1], Kampanart Chayarus[1], Jidapa sangswan[2], Auradee Punkvang[3], Khomson Suttisintong[4], Sanya Sureram[5], Prasat Kittakoop[5,6,7], Poonpilas Hongmanee[8], Pitak Santanirand[8], Jiraporn Leanpolchareanchai[9], James Spencer[10], Adrian Mulholland[11] and Paptawan Thongdee[1]

[1]*Department of Chemistry and Center of Excellence for Innovation in Chemistry, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, 34190, Thailand*

[2]*Department of Biological Science, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, 34190, Thailand*

[3]*Division of Chemistry, Faculty of Science, Nakhon Phanom University, Nakhon Phanom, 48000, Thailand*

[4]*National Nanotechnology Center, NSTDA, Thailand Science Park, Klong Luang, Pathum Thani, 12120, Thailand*

[5]*Chulabhorn Research Institute, Bangkok, 10210, Thailand*

[6]*Chulabhorn Graduate Institute, Chemical Biology Program, Chulabhorn Royal Academy, Bangkok, 10210, Thailand*

[7]*Center of Excellence on Environmental Health and Toxicology (EHT), OPS, Ministry of Higher Education, Science, Research and Innovation, Bangkok, 10210, Thailand*

[8]*Division of Clinical Microbiology, Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok 10400, Thailand*

[9]*Department of Pharmacy, Faculty of Pharmacy, Mahidol University, Bangkok 10400, Thailand*

[10]*School of Cellular and Molecular Medicine, Biomedical Sciences Building, University of Bristol, Bristol BS8 1TD, United Kingdom*

[11]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*

Corresponding author: pornpan_ubu@yahoo.com

Abstract

Protein kinase B (PknB) has emerged as a promising target for developing new drugs to combat tuberculosis (TB) [1]. This study utilized structure-based virtual screening and biological assays to identify candidate compounds from natural products in the Specs compound library targeting *M. tuberculosis* PknB ATPase. Among the six hit compounds, compound **6** (AO-103/41868968) was particularly effective, inhibiting *M. tuberculosis* PknB ATPase with an IC_{50} value of 9.96 µM. Importantly, this compound demonstrated no toxicity to Caco-2 cells at concentrations up to 40-fold higher than its IC_{50} value. Furthermore, the binding mode and binding energy of compound **6** were investigated using molecular dynamics (MD) simulations. The results revealed strong hydrogen bond interactions with Glu93 and Val95, along with hydrophobic interactions involving Leu17, Phe19, Val25, Val72, Met92, Ala142, Met145, and Met155. These findings underscore compound 6 as a promising starting point for optimizing anti-tuberculosis agents.

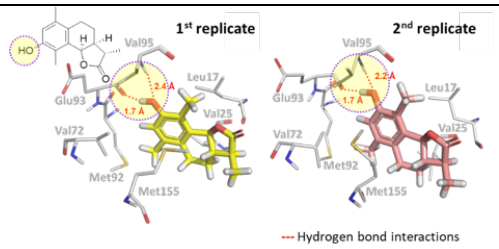


Figure 1: *Interaction structures of compound* ***6*** *and the key residues of M. tuberculosis PknB. Red dotted lines indicate hydrogen bonding interactions.*

References

[1] Mori, Matteo, Sammartino, José Camilla, Costantino, Luca, Gelain, Arianna, Meneghetti, Fiorella, Villa, Stefania, Chiarelli, Laurent Roberto. (2019). *Current Topics in Medicinal Chemistry*, **19**(9), 646-661.

Catalytic Residues pKa Modulation Impact on Reactivity, SARS-CoV-2 Mpro Evaluation

Carlos A. Ramos-Guzmán[1], Alexander Argyrou[1], Michael Beer[1], A. Sofia F. Oliveira[1], Adrian J. Mulholland[1]

[1] *Centre for Computational Chemistry, School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK*2

Corresponding author: carlos.ramos@bristol.ac.uk

Abstract

This study investigates the factors that shape the catalytic activity of the SARS-CoV-2 main protease (Mpro) in presence of three peptide substrates using extensive molecular dynamics (MD) simulations and the pKa prediction software PROPKA3. We focus on the differences in the rate of cleavage of the substrates by the enzyme. The S01 substrate is hydrolysed the fastest, followed by S02 and ending with a slow cleavage rate for the S05 substrate. Our results show a corelationship between His41 pKa values and substrate cleavage rates, indicating that substrates with poor cleavage rates spend more time in configurations with lower His41 pKa values than the fastest. In addition, an inverse behaviour is observed for the pKa values for the catalytic Cys145. A combination of high His pKa and low Cys pKa could be translated into a favourable proton transfer event between these two residues, a key step in the hydrolysis reaction catalysed by the enzyme.

We also investigated the role of the conserved crystallographic water molecule within the His41 - His164 - Asp187 (HHD) cavity on the pKa of His41. It was found that the hydrogen bond interactions formed by this water molecule with the residues in the HHD cavity correlate with the observed pKa shifts for His41. In particular, this water molecule shows stronger interactions with the HHD cavity in the presence of S01, followed by S02, and weaker interactions with S05. Overall, our study provides important insights into the catalytic mechanisms of natural peptide substrates of SARS-CoV-2 Mpro and lays the groundwork for future strategies in antiviral drug development.

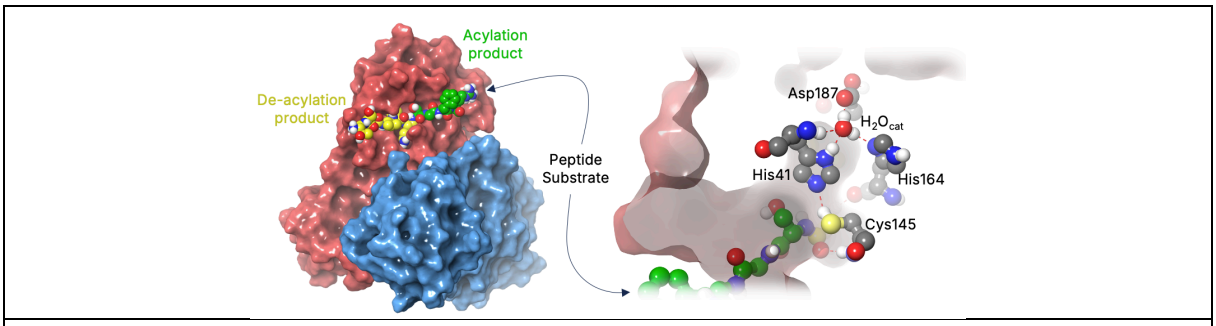


Figure 1: *SARS-CoV-2 M^{pro} Michaelis Complex and detail of the HHD cavity and the catalytic dyad.*

References

[1] Chan, H. T. H.; Moesser, M. A.; Walters, R. K. et al. (2021). *Chem. Sci*, **12**(41), 13686–13703.

[2] Hui Li, Andrew D. Robertson, and Jan H. Jensen. (2005). Proteins, **61**(4), 704-721.

Computational Enzyme Engineering for Lignin Valorization and Biofuel Production: Advancements in Aryl-O-Demethylation

Sonia F Santos 1[1,2]; Ewan Waters[1,2], Rajesh Reddy Bommareddy[1,2], Warispreet Singh [1,2]
[1] Hub for Biotechnology in Build Environment, Newcastle upon Tyne, United Kingdom, [2]Department of Applied Sciences, Northumbria University, Newcastle upon Tyne

Corresponding author: w.singhl@northumbria.ac.uk

Abstract

The significance of aryl-O-demethylation is increasingly recognized in lignin valorization for chemical production and biofuels. A key advancement is the discovery of the cytochrome P450 enzyme (GcoAB) from the CYP255A45 family, which converts guaiacol to catechol via aryl-O-demethylation. A single-site mutation (Phe169Ala) in the active site of GcoAP450 enables the oxidation of syringol, making the enzyme more versatile. Further mutations (Phe169Ser and The296Ser) have expanded its ability to demethylate o- and p-isomers of vanillin. Through recombinant engineering, GcoAP450 has been adapted to process diverse lignin-derived subunits, proving to be a versatile catalyst for lignin valorization.

In the biorefinery industry, a depolymerization method known as reducing catalytic fractionation (RCF) produces 4-alkylguaiacols. However, transforming these into valuable products poses a challenge, as GcoAP450 has limited activity against 4-alkylguaiacols with additional alkyl groups. A newly discovered enzyme, AgcAP450 from *Rhodococcus rhodochrous* EP4, effectively catalyzes the aryl-O-demethylation of 4-alkylguaiacols to catechol. Given the 64.4% sequence identity between AgcAP450 and GcoAP450 (Figure 1), a homology model based on GcoAP450's X-ray structure was developed. Using molecular docking, MD simulations, and QM/MM calculations, we elucidated AgcAP450's binding configuration and its ability to process bulkier substrates compared to GcoAP450.The structural insights from these enzymes, especially regarding 4-alkylguaiacols—lignin depolymerization by-products—are crucial for developing robust P450s for efficient lignocellulose-based biorefinery processes. Building on knowledge from P450 GcoA and AgcA, the goal is to design a versatile enzyme catalyst capable of aryl-O-demethylation of both small and bulky substrates. Using rational design and MD simulations with guaiacol and 4-propylguaiacol as model substrates has revealed interactions within the enzyme's active site, providing insights into the effects of engineered mutations. These simulations offer a detailed view of enzyme-substrate dynamics, enhancing understanding and predicting catalytic potential.

Figure 1: Sequence alignment of the $AgcA_{P450}$ and $GcoA_{P450}$ enzymes using Clustal-Omega webserver developed by the European Bioinformatics Institute. Marked in green we have the conserved residues between these two proteins. While in red we have the non-conserved residues, In the active centre we have an Isoleucine ($GcoA_{P450}$) replaced by a Leucine ($AgcA_{P450}$) and a threonine (GcoA) replaced by an Alanine ($AgcA_{P450}$). At the entrance of the main pocket in $AgcA_{P450}$ have a Serine, a much smaller residue than the Asparagine present in $GcoA_{P450}$.

Fragment-Guided Virtual Screening: Unveiling Promising Drug Candidates through Computational Chemistry

Natalia Serrano-Aparicio[1]; Ellie Stammers[1], Marc Cabry[1], Andrea Scarpino[1], Yann-Vaï Le Bihan[1], John Cadwell[1], Rob L.M. van Montfort[1], Zoran Rankovic[1]

[1] *Centre for Protein degradation, Centre for Cancer Drug Discovery at the Institute of Cancer Research, London SM2 5NG, UK*

Corresponding author: natalia.serranoaparicio@icr.ac.uk

Abstract

Modern drug discovery is a highly multidisciplinary area where computational chemistry enables the efficient exploration of vast chemical spaces. In this context, the use of computational techniques combined with structural information can lead to the identification of new binding sites and help finding new hit compounds for drug discovery purposes. In this study, we present a multi-stage virtual screening approach utilising crystallographic fragment information from XChem screening,[1] preliminary validated by LO-NMR[2]. Our aim is to identify potential hits binding to a functionally relevant site of Damage-specific DNA binding protein 1 (DDB1), with the possibility of hijacking it as ligase for protein degradation purposes.

Initially, we leveraged XChem crystallographic data of fragments bound to our target protein to define a pharmacophoric query including features shared by at least two fragments. These features guided a pharmacophore-based virtual screening of two Enamine compound libraries (3.1M compounds), yielding more than 150K virtual hits. Each hit was then subjected to docking simulations which allowed to prioritise compounds whose features matched the pharmacophore query also in the predicted binding conformations. Where the Molecular Operating Environment software (MOE)[3] was selected for performing the pharmacophoric search and the subsequent docking stage was done with Glide[4]. The resulting virtual hits were further classified into two distinct groups assigned, using cheminformatics approaches, depending on if structural information from the parent XChem fragments is contained or they correspond to novel structures. This classification informed the visual inspection stage and led to the selection of 60 compounds representing both categories to probe for the extent of structural diversity permitted in the pocket.

Our results highlight several promising candidates whose experimental validation is ongoing and demonstrate the effectiveness of integrating structural data from crystallographic fragment screening with structure-based computational methods. We believe the approach presented here not only enables and accelerates the identification of compounds with improved affinity but will also provide insights into the key features driving the binding to the site under investigation.

References

[1] Achieving Efficient Fragment Screening at XChem Facility at Diamond Light Source. JoVE, 2023 (JoVE).
[2] Almeida TB, Panova S, Walser R. (2021) SLAS DISCOVERY: Advancing the Science of Drug discovery. **26**(8),1020-1028.
[3] Molecular Operating Environment (MOE). Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2023.
[4] Glide. Schrödinger Release 2023-1: Glide, Schrödinger, LLC, New York, NY, 2023.

TORCphysics: A physical model of supercoiling mediated regulation in synthetic gene circuits

Victor Velasco-Berrelleza[1]; Aalap Mogre[2], Penn Faulkner Rainford[3], Charles J. Dorman[2], Carsten Kröger[2], Susan Stepney[3], Sarah Harris[1]
[1]*Department of Physics and Astronomy, University of Sheffield, Sheffield, UK*
[2] *Department of Microbiology, Trinity College Dublin, Dublin, Ireland*
[3] *Department of Computer Science, University of York, York, UK*

Corresponding author: V.VelascoBerrelleza@sheffield.ac.uk

Abstract

Current synthetic gene circuits are designed to perform logical functions, mimicking those in electronic circuits [1]. These designs ignore DNA supercoiling, which can affect gene expression as it is intimately related to transcription, capable of both up- and down-regulating the expression of distal genes [2]. Our goal is to understand, characterize, and harness the potential of supercoiling in regulating transcription in synthetic gene circuits as a novel and powerful component of the next generation of synthetic gene circuits.

In this work, we introduce TORCphysics, a fast and computationally cost-effective coarse-grained physical model designed to predict the expression profiles of gene circuits in bacteria by simulating transcription-dependent supercoiling under various biological conditions. These expression profiles are suitable for comparison with experimental data. Our results reveal the interplay between transcription and genomic architecture, highlighting the importance of factors such as gene orientation, topological barriers, and promoter sequences. Concurrently, wetlab experiments are underway to validate our findings.

The insights gained from our physical model, combined with experimental data, will contribute to the development of a computational toolkit called TORC. TORC will enable the design of genetic circuits that exploit transcription-dependent supercoiling. Once constructed, the toolkit's utility will extend to circular DNAs in complex environments, including eukaryotic cells. The practical implications are broad, potentially yielding gene circuits with enhanced functionalities and increased efficiency.

References

[1] Goñi-Moreno, Angel and Nikel Pablo I. (2019). *Frontiers in bioengineering and biotechnology*, **7**, 40.
[2] Dorman Charles J and Dorman Matthew J. (2016). *Biophysical Reviews*, **8**, 209-220.

Investigating Targeted Therapies for Cystic Fibrosis through D-NEMD Simulations of the CFTR Protein

Diana F. Veselu[1], Deborah K. Shoemark[1], David N. Sheppard[1] and Marc W. van der Kamp[1]
[1] *University of Bristol*

Corresponding author: Marc.VanderKamp@bristol.ac.uk

Cystic fibrosis (CF) is a debilitating genetic disorder caused by the malfunction of the cystic fibrosis transmembrane conductance regulator (CFTR) protein, an anion channel that regulates epithelial ion transport. The treatment of CF has been revolutionized by orally administered drugs that directly target defects in CFTR called correctors and potentiators. The potentiator ivacaftor partially restores function to CFTR channels with gating mutations, such as G551D. Ivacaftor binds CFTR at the protein-lipid interface in a cleft formed by transmembrane segments 4, 5, and 8,[1] but the conformational and dynamical changes it induces remain incompletely understood. Here, we investigate how the ivacaftor-binding site is allosterically coupled to other functionally important regions of CFTR using a new computational approach named dynamical-nonequilibrium molecular dynamics (D-NEMD) simulations.[2] The D-NEMD process combines molecular dynamics (MD) simulations under equilibrium and nonequilibrium conditions to map the evolving structural response of a protein to a perturbation.

We constructed a model of CFTR from the cryo-EM structure of phosphorylated, ATP-bound human CFTR in complex with ivacaftor.[1] Missing loops were filled in using AlphaFold.[3] The regulatory domain that is absent from the cryo-EM dataset was not included due to its considerable size and low-confidence AlphaFold prediction. The model was inserted into a lipid membrane composed of POPC, POPE, and cholesterol. The simulations were consistent with a 60 % lipid contribution to the ivacaftor binding site,[1] with interactions formed during a preliminary 20 ns "pre-production" stage. Consequently, ivacaftor remained in the cryo-EM-deduced binding site throughout all the subsequent unconstrained 500 ns runs (5 replicas for each system), unlike systems set up without prior equilibration of the lipid binding site. This allowed us to refine our understanding of ivacaftor-CFTR dynamics and its response to ivacaftor perturbation. These equilibrium MD simulations were used to generate a large set of short D-NEMD simulations (305 x 5 ns) where ivacaftor was removed as the perturbation, to identify communication pathways through the observed responses (ensuring the statistical significance of our findings).

Our work reveals that there are allosteric signals related to ivacaftor binding that are propagated from the binding pocket up towards the extracellular vestibule of the CFTR pore and, significantly, downwards via neighbouring transmembrane segments to both ATP-binding sites at the interface of the dimer (the nucleotide-binding domains). For example, there are three routes from the ivacaftor-binding site that reach ATP-binding site 2. One of these goes through transmembrane segment 5 and intracellular loop 2 to reach ATP-binding site 2. By studying disease-causing mutations that map to these allosteric pathways, such as G551D, we can gain a better understanding of how ivacaftor restores CFTR function, potentially guiding the rational design of future therapies for CF.

References

[1] Liu, F., et al. (2019). *Science*, **364**, 1184-1188.
[2] Oliveira, S., et al. (2021). *Eur Phys J B,* **94,** 144.
[3] Jumper, J., et al. (2021). *Nature,* **596,** 583-589.

A Machine Learning Approach to Identify Carbon Dioxide Binding Proteins for Sustainability and Health

George Weston[1]; Dr. Matteo Degiacomi[1], Prof. Martin Cann[1]
[1]*Durham University, Department of Biosciences, South Road, Durham, DH13LE*

Corresponding author: george.m.weston@durham.ac.uk

Abstract

Carbon dioxide (CO2) has a fundamental role in biology. It is involved in a huge variety of processes in many organisms, ranging from cellular homeostasis to photosynthesis [1, 2]. While much is known about the impact of CO2 on the overall physiology of an organism, much less is known about how its interaction with specific biomolecules may affect their function. Our work focusses on carbamylation, a non-enzymatic reversible protein post-translational modification (PTM) where CO2 binds onto the neutral lysine ε-amino groups [3]. A novel mass spectrometry-based experimental technique, TEO trapping, has been recently used to demonstrate that, for reasons yet unclear, CO2 does not bind to every lysine [2]. While informative, this experimental technique is laborious, which limits our ability to gain a clear and comprehensive view of the interactions between CO2 and proteins. For this reason, we are carrying out extensive molecular dynamics simulations and developing a computational method to predict which lysine, in any protein, may undergo carbamylation. We found that standard metrics (pKa, solvent accessible surface area, and amino acid depth) are insufficient to singlehandedly explain why some lysine may be modified, but not others. Therefore, we are now exploring the usage of atomic environment vectors, commonly used in machine learning, for this classification task.

References
[1] Linthwaite, V. L et al, Ubiquitin is a Carbon Dioxide Protein. *Science Advances* **7**, eabi5507
[2] Blake, L. I., Cann, M. J., Carbon Dioxide and the Carbamate Post-Translational Modification. *Frontiers in Molecular Biosciences,* **9**, 825706
[3] Meigh, L. I., CO_2 Carbamylation of Proteins as a Mechanism in Physiology. *Biochemical Society Transactions* **43**, 460-464

Multi-scale simulations provide insights for engineering product outcome of polyketide synthases

Elise White[1]; Marc W. van der Kamp[1]
[1]*School of Biochemistry, University of Bristol, Bristol, United Kingdom*
Corresponding author: ew17435@bristol.ac.uk

Abstract
Polyketides are a diverse group of natural compounds; many are biomedically active. For example, the antibiotic actinorhodin (act) is biosynthesised by a type-II polyketide synthase (PKS) multi-enzyme complex. A ketosynthase-chain-length-factor (KS-CLF) catalyses initial chain elongation reactions to iteratively build a polyketide backbone. This involves a repeating cycle of binding of an acyl carrier protein (ACP) with a phosphopantetheine (Ppant) arm attached to a malonyl-group, a condensation reaction to join the malonate units together, and unbinding of ACP. This occurs until the polyketide reaches the desired chain length, controlled by the substrate channel. Subsequent enzymes then tailor the polyketide to generate the final molecule. Due to the reactive nature of polyketide substrates and the transient nature of protein-protein interactions, these systems are difficult to study experimentally, and obtaining structural insight into the states relevant for reaction is a challenge. As such, the detailed mechanism of type-II PKS polyketide chain elongation is debated.

Here, we build models of actACP-actKS-CLF complexes for multiple stages of the chain elongation reaction cycle using protein-protein docking and iteratively refining Ppant and polyketide substrate positioning using molecular mechanical (MM) molecular dynamics (MD) simulations. We then use these optimised models in extensive combined quantum mechanical / MM (QM/MM) simulations. Semi-empirical QM methods were selected through benchmarking against small (QM-only) models at DFT level. Combined QM/MM CI-NEB (climbing image nudge elastic band) and umbrella sampling reaction simulations successfully revealed the mechanism of the first step in chain elongation and highlight an energy difference in penultimate and final rounds of elongation. Detailed mechanisms of the second and third steps of the reaction help generate a complete picture of the reaction. MM MD simulations further indicate that ACP-Ppant is released when the expected polyketide chain length is reached. This mechanistic insight into the chain elongation reaction, coupled with other work regarding chain length control, provides a platform for PKS engineering to develop novel polyketides with potential novel biomedical activities.

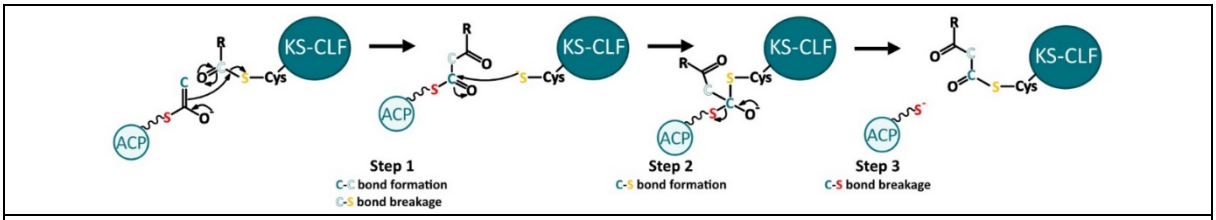


Figure 1: *Type-II PKS KS-CLF-catalysed chain elongation reaction steps. Step 1 is the transfer of the growing polyketide from the KS-CLF active-site cysteine to the malonyl-loaded-Ppant to elongate it by one ketide unit. Step 2 is the formation of the tetrahedral intermediate, before step 3 which is the complete transfer back to the active-site cysteine to yield an 'empty' ACP-Ppant which can unbind.*

References

[1] Bräuer, A., Zhou, Q., Grammbitter, G. L. C., Schmalhofer, M., Rühl, M., Kaila, V. R. I., Bode, H. B., and Groll, M. (2020). *Nat. Chem.,* **12**, 755–763.
[2] Chen, A., Jiang, Z., Burkart, M. D. (2022). *Chem. Sci.,* **13**, 4225-4238.

How do Autoencoders Help Explore the Conformational Space of MD Simulations of Cyclic peptides?

Leonie Windeln[1]; Charlie Holdship[1], Jeremy Frey[1], Jonathan Essex[1]
[1]*University of Southampton, School of Chemistry, B29 University Road, SO23 1BJ, United Kingdom*

Corresponding author: L.Windeln@soton.ac.uk

Abstract

α-Conotoxins are a class of disulfide-rich cyclic peptides produced by marine cone snails that target human nicotinic acetylcholine receptors (nAChRs). The specificity of these toxins against different isoforms of nAChRs make them attractive pharmacophore candidates.[1]

To elucidate the mechanism of action, the solution conformations of the conotoxins must be determined before modelling their interactions with the receptor. We have carried out molecular dynamics (MD) simulations of five conotoxins using enhanced sampling methods. In MD simulations, large datasets with high dimensionality (many variables) are generated. These variables are the cartesian coordinates of each atom for each time-step of the simulation. From this, other variables such as the backbone torsion angles can be derived. To extract meaningful movement of the system over time and thus obtain stable solution structures, the dimensionality of these data has to be reduced significantly.

A popular method for dimensionality reduction is principal component analysis (PCA) which uses a linear combination of all input variables to calculate the orthogonal collective variables which maximally capture the data covariance.[2] In the case of our conotoxin simulations, the first two principal components of the PCA captured less than 50 % of the variance. Here we compare the use of PCA to analyse the conformational space of conotoxins to using autoencoders for the same task. Standard autoencoders have a symmetric neural network architecture with a bottleneck in the center (encoder layer) whilst the network is trained to reduce the error between input and output (decoder) layer. We were able to perform reduction of the simulation data to two dimensions (2 encoder nodes) and extract diverse conformations with better separation than for the PCA workflow. Furthermore, we achieved high model accuracy (defined by low reproduction error) throughout the trajectory for autoencoders, which was not the case for PCA.

Going forward autoencoders will enable us to more accurately model interactions of conotoxins and their targets. Our findings in using autoencoders for analysing simulations of cyclic peptides will allow us to expand their use in structural biology and help identify more biologically relevant conformations of other cyclic peptides.

References

[1] Lebbe, Eline K.M., Peigneur, Steve, Wijesekara, Isuru, Tytgat, Jan. (2014). *Mar. Drugs*, **12**(5), 2970-3004.
[2] Sittel, Florian, Jain, Abhinav, Stock, Gerhard. (2014). *J. Chem. Phys.*, **141**(1), 014111.

DL_FIELD – a software workflow tool for force field models.

Chin Yong[1]
[1] *Computational Chemistry Group, Scientific Computing Department, STFC Daresbury Laboratory, Warrington WA4 4AD, UK*

Abstract

DL_FIELD is a software tool to set up molecular simulation force field (FF) models for DL_POLY and Gromacs. DL_FIELD contained a unified FF framework that enable smooth data transformation for a wide range of FF schemes implemented within DL_FIELD, such as CHARMM, AMBER, OPLS, PCFF, CVFF, TraPPE, and their variants. Within the similar framework, DL_FIELD also contained inorganic force fields for a wide range of materials such as zeolites, glass, and minerals. The software tool allows users to easily switch FFs, via a single-step process, and to construct complex mixed FF models, including those of bio-inorganic systems.

DL_FIELD program has the following functions:

(i) Force field model convertor. Reads user configurations and concurrent production of ready-to-run, self-contained force field files for DL_POLY and Gromacs simulation packages.
(ii) Force field editor: Allows users to edit or modify a particular FF scheme to produce a customised scheme that is specific to a particular molecular model.
(iii) Force field model repertoire: consistent file structure format for all FF schemes and molecular structure definitions.
(iv) Third-party FF file formats (CHARMM's pdb, rtf, psf and prm) conversion into DL_FIELD FF files for DL_POLY and Gromacs.
(v) Integration with pyChemShell for QM/MM calculations.
(vi) Chemical-sensitive atom typing and identification, by making use of the DL_F Notation [1], to facilitate rapid chemistry-based atomic interactions analysis including standardised, systematic annotation and quantification analytics using DL_ANALYSER [2].

DL_FIELD is designed to handle a wide range of molecular system of varying complexity: from simple ionic compounds, small covalent molecules to systems with complex topologies such as biomolecules, carbohydrates, drug molecules, organic cages, or a mixture of these components. In addition, the unification of file formats and data structures based on the DL_FIELD framework facilitates migration of one class of FF system model to another, with minimum learning curve. This encourages researchers to carry out calculations on novel classes of molecular systems spanning across multidisciplinary fields, from material sciences to biological and pharmaceutical areas.

References
[1] C. W. Yong, *J. Chem. Inf. Model.* vol.**56**, pp 1405-1409 (2016)
[2] C. W. Yong & I.T. Todorov, *Molecules* **23**, 36 (2018)

Between States: Assessing Allosteric Modulators Through Conformational Probabilities

Chenfeng Zhang[1]; Julien Michel[1]
[1]*EaStChem School of Chemistry, Joseph Black Building, University of Edinburgh, Edinburgh, EH9 3FJ, UK*

Corresponding author: julien.michel@ed.ac.uk

Abstract

The allosteric effect is a crucial mechanism for modulating protein functions. However, predicting whether a ligand binding at a non-orthosteric site can act as an allosteric modulator remains challenging. To tackle this issue, we propose a computational method that combines molecular dynamics simulations with Markov State Models (MSMs).[1] This method evaluates the modulation efficacy of allosteric binders based on the probabilities of observing a protein in different conformational states. In our workflow, after distinguishing between different states using conventional molecular dynamics simulations (cMD), state transitions are sampled through steered MD (sMD). Snapshots from the sMD trajectories are then used as seeds for further unbiased simulations, and their trajectory data are featurised to construct the MSMs. Through the model, probabilities of the protein in different states can be determined, enabling the assessment of the modulation potential of allosteric ligands. The potential and challenges of this method are illustrated through case studies on Phosphoinositide-dependent kinase-1 (PDK1). Several experimentally verified PDK1 allosteric modulators[2,3] were put through the workflow to validate the method. Our study confirms the method's effectiveness in assessing the modulation efficacy of allosteric ligands, highlighting its utility in the 'hit to lead' phase of allosteric drug discovery campaigns.

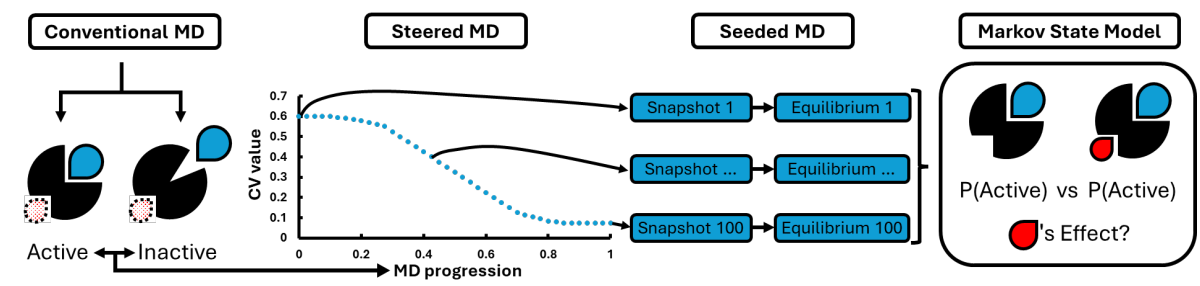


Figure 1: *Brief workflow of our method.*

References
[1] Hardie, A., Cossins, B. P., Lovera, S. and Michel, J. (2023). Communications Chemistry, 6(1), 125.
[2] Sadowsky, J. D., Burlingame, M. A., Wolan, D. W., McClendon, C. L., Jacobson, M. P., and Wells, J. A. (2011). Proceedings of the National Academy of Sciences, 108(15), 6056-6061.
[3] Busschots, K., Lopez-Garcia, L. A., Lammi, C., Stroba, A., Zeuzem, S., Piiper, A., Alzari, P.M., Neimanis, S., Arencibia, J.M., Engel, M. Schulze, J.O. and Biondi, R. M. (2012). Chemistry & biology, 19(9), 1152-1163.

Markov State Models from millisecond kinase MD trajectories

Ryan Zhu[1]; S. Singh[2]; M. T. Degiacomi[3]; A. S. J. S. Mey[1]
[1] *EaStCHEM School of Chemistry, University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ, United Kingdom*
[2] *Department of Biochemistry and MolecularBiophysics, Washington University School of Medicine, St.Louis, Missouri 63110, United States; Center for the Scienceand Engineering of Living Systems, WashingtonUniversity in St. Louis, St. Louis, Missouri 63130, United States*
[3] *Department of Physics, Durham University, South Road, Durham, DH1 3LE, United Kingdom*

Corresponding author: ryan.zhu@ed.ac.uk

Abstract

Protein kinases play crucial roles in cellular signalling by regulating various biological processes, including cell growth, differentiation, and division. Dysregulation of these kinases is often found in cancers, where mutations can lead to uncontrolled cell proliferation. Kinases, as other proteins, sample a large and diverse set of conformations that are key to their function and regulation. Understanding the structural features of these conformational states provides insights into the mechanisms underlying kinase activity and also facilitates inhibitor design. Computational methods, such as molecular dynamics (MD) simulations, enable us to study the dynamic behaviour of these proteins at an atomic resolution, capturing transitions between different conformations. Markov state models (MSMs) can then be used to recover information on kinetics even from sets of independent trajectories [1].

In this talk, I will present an MSMs analysis of MD simulations of the catalytic domains of EGFR and Abl1 kinases. Their catalytic activity is characterised by a kinetically slow transition between 'active' and 'inactive' states defined by a few key structural features, including the DFG-motif [2]. We analysed trajectories that reach a cumulative simulation time of 2.4 ms, which were started from a diverse set of starting conformations, collected on the Folding@home network. For the first time, we demonstrate that DFG-flipping, experimentally known to occur on timescales of tens of milliseconds [3], can be reconstructed from parallel simulations, with no single trajectory longer than 2 microseconds. This model gives atomic resolution insights into the transition between active and inactive conformations facilitating future design of novel inhibitors.

References

[1] Arbon, R. E., Zhu, Y., and Mey, A. S. J. S. (2024). *J. Chem. Theory Comput.*, **20** (2), 977–988.
[2] Modi, V., and Dunbrack, R. L. Jr. (2019). *Proc. Natl. Acad. Sci.*, **116** (14), 6818–6827.
[3] Xie, T., Sale, T., Rossi, P., and Kalodimos, C. G. (2020). *Science*, **370** (6513), 189.