11th Annual CCPBioSim 2025 Conference: Frontiers in Biomolecular Simulations

Conference Booklet

CCPBioSim

University of Southampton

14th-16th July, 2025

University of Southampton

Science and Technology Facilities Council

Engineering and Physical Sciences Research Council

CoSeC Computational Science Centre for Research Communities

SANDBOXAQ™

OPEN BIOSIM

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

Table of Contents

Dear Guests,

Hello and welcome to the 11th CCPBioSim 2025 conference in Southampton. This year we have a theme of exploring the cutting edge of biomolecular simulations and associated methods, and their real-world impacts. Biomolecular modelling has an ever increasing role in the creation of new therapies, including antibodies and peptides. Simulation is also widely used to help interpret experimental data and features in many structural biology pipelines. Machine learning and artificial intelligence are having an ever-increasing impact, analysing simulation data, improving the accuracy of intermolecular potentials, and designing novel biomolecules, often with the aid of physics-based approaches. This conference will showcase the latest advances in these broad areas, and how they are impacting industry and academia.

Topics of interest include:

- New methods: What new approaches have enabled faster or more accurate or larger-scale simulations of biomolecular systems?
- Pushing the envelope: Where have larger or longer simulations delivered new biological insights?
- Artificial intelligence: How is the combination of machine learning and artificial intelligence with biomolecular simulation advancing our knowledge of biological systems?
- Experimental interpretation and design: How can we best combine biomolecular simulation and experiment to deliver new biological understanding?

We are a highly multi-disciplinary community, and we encourage everyone to engage, network, make new friends and learn some interesting science. We hope you enjoy your time in Southampton. Should a problem occur please discuss with a member of the organizing committee, or emailing ccpbiosim2025@soton.ac.uk.

The CCPBioSim Conference Committee

The Conference Venue

The conference will be taking place on Highfield Campus at The University of Southampton. Talks will be held in the Building 2A lecture theatre and posters, lunch, snacks, drinks, dinner to be hosted in the 'Garden Court'. Registration will be at Garden Court.

Catering

On Monday 14th July, Tuesday 15th July and Wednesday 16th July there will be lunch available in the garden court. The lunch will be buffet style. Food allergies have been noted and will either be provided separately or be 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 Grand Cafe (see below for details). In the 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.

Southampton

The bus system in Southampton has regular buses covering a lot of the city; though traffic can be heavy at peak times. More information on the 'UniLink' buses can be found here: https://www.unilinkbus.co.uk/. In general, the final letter of the bus code indicates the direction of travel with almost all unilink buses stopping at Highfield campus along their route. For example, U1C takes you from the airport into the city via Highfield while the U1A is the reverse route. Southampton has two main train stations, an airport, and is connected to the M27 and M3. It is likely you will take the U1A from your hotel in town to the university campus (or the U1C if your hotel is by the airport).

By air: If you are flying to Southampton you will arrive at a small airport. There will be regular UniLink buses from outside the airport. The U1C will take you into the city center via the university campus. If you are flying to a London airport, we recommend Heathrow which is an hour and 15 by car, or two hours on public transport (Elizabeth Line to Reading followed by a train).

By train: If you are arriving by train, then you are likely going to arrive at either Southampton Central or Airport Parkway. The former is at the bottom of the city and is appox a 30 min bus (U2B) ride to the campus. Central station is likely closer to your hotel. Airport parkway is closer to campus and is better serviced by the U1C bus.

By car: Highfield campus can be accessed by exiting the M27 at junction 5. We are unable to offer parking for the conference however parking based on accessibility grounds can be arranged. Please contact the email address below.

Southampton is a port city which is well known for its maritime history with the fabled Titanic departing from here in 1912. You will also find medieval structures and Tudor buildings which remarkably survived the blitz. Known as Sunny Southampton, you can expect warm weather in July. The city boasts lots of green space, most notably Southampton Common which is close to Highfield Campus and may be a nice place to relax/have a BBQ in the evenings.

Southampton offers a diverse array of restaurants and pubs to suit various tastes. The main areas to consider are Portswood (down the road from campus), Bedford Place, West Quay and Ocean Village.

Posters

All posters should be put up during the registration on July 14th and they must be taken down on July 16th. The poster boards are to be located in the Garden Court. 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 landscape. Please have your poster with you when you come; it is not possible to print it on the day. Each poster session is scheduled for 90 minutes.

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

1) The visual clarity of the poster

2) The interaction of the poster presenter during poster sessions

3) The quality of the research

We will announce the **three** poster prizes at the conference dinner. The prizes are sponsored by OpenBioSim, CCPBioSim and Physical Sciences Data Infrastructure.

Conference dinner

The Grand Café – 19:30

South Western House, Southampton SO14 3AS

On Tuesday, July 15th we will be having a conference dinner at the Grand Café. After the poster session, there will be **two** coaches available to transit us to the venue should you wish. We have the venue until late. Please note, there is no return transit.

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. All registrants should have received a google poll requesting a pre-order on your food choices. If we have not heard back from you, unfortunately we have had to select a default option according to your dietary requirements.

IF FOR ANY REASON YOU ARE UNABLE TO ATTEND THE DINNER, PLEASE INFORM A MEMBER OF THE ORGANISING COMMITTEE AS SOON AS POSSIBLE.

Sponsors

As you may, or may not, have noticed, the registration fee for this years conference was particularly cheap by todays standards. This is in thanks due to the subsidiaries passed on by the CCPBioSim grant and our sponsors, OpenBioSim, Sandbox AQ and the Physical Sciences Data Infrastructure. We thank all of them for their support.

Organising Committee

Prof. Jonathan Essex, Southampton University
Dr. William Poole, Southampton University
Dr. James Gebbie-Rayet, Daresbury Laboratory @ STFC
Dr. Shivani Harshe, Sheffield University
Prof. Sarah Harris, Sheffield University
Dr. Daniel Cole, Newcastle University
Dr. Sofia Oliveira, Bristol University
Prof. Aditi Borkar, Nottingham University
Dr. Agnieszka Bronowska, Newcastle University
Ast. Prof. Warispreet Singh, Northumbria University

ccpbiosim2025@soton.ac.uk

CCPBioSim

CCPBioSim is an inclusive wide-ranging project, bringing together chemists, physicists and chemical engineers as well as researchers from all branches of ‘molecule-oriented’ biochemistry and biology. Our aim is to involve experimentalists and computational specialists, sharing the belief that the best science can be done when theory and experiment are closely integrated. CCPBioSim develops and provides training and tools to lower the barrier to non-experts becoming proficient and productive users of biomolecular simulation techniques. We also work to develop and apply advanced methods.

Programme

Day 1 – Monday 14th July

12:00-13:00	Arrival and Lunch	
13:00-13:05	Jon Essex	Welcome
13:05-13:15	Nicolas Foloppe	CCPBioSim
Session 1: Experimental Integration and Design Chair:		
13:15-13:55 **Invited** Talk 1	Tomek Wlodarski - Institute of Biochemistry and Biophysics, Warsaw	
13:55-14:20 Contributed Talk 1	Marko Hanzevacki – University of Bristol	Directed Evolution of Cytidine Deaminase for Biocatalysis: Simulation Insights for Enzyme Design and Engineering
14:20-14:45 Contributed Talk 2	Cameron Brown – University of Southampton	Integrating SAXS, MD and coarse-grained modelling for all-atom protein ensemble generation
14:45-15:15	Tea and Coffee	
15:15-15:35	UKRI	Words from UKRI
15:35-16:15 **Invited** Talk 2	Max Bonomi - Institut Pasteur	Atomic resolution ensembles of intrinsically disordered and multi-domain proteins with Alphafold
Session 2: Theory, Application and Design: part 1 Chair:		
16:15-16:40 Contributed Talk 3	Ioana Papa - University of Sheffield	Energy-entropy analysis of interactions between CLR/RAMP ectodomain complexes and small molecule antagonists
16:40-17:05 Contributed Talk 4	Frederick Powell - Heriot-Wat University	In Search of Enzyme Activation: A Markov state model approach to allosteric drug discovery
17:05-17:30 Contributed Talk 5	Alice Allen - Los Alamos National Laboratory	Multi-fidelity learning and meta-learning for interatomic potentials
17:30-19:00	Poster Session (Odd Numbers)	

Day 2 – Tuesday 15th July - Morning

Session 3: Pushing the Envelope Chair:		
09:00-09:40 **Invited** Talk 3	Syma Khalid - University of Oxford	
09:40-10:05 Contributed Talk 6	Chelsea Brown - University of Groningen	Integrative modelling of the mitochondrial cristae
10:05-10:45 **Invited** Talk 4	Pavel Buslaev – Astex Therapeutics	
10:45-11:15	Tea and Coffee	
11:15-11:55 **Invited** Talk 5	Kush Coshic - Max Planck Institute	
11:55-12:20 Contributed Talk 7	Keith Cassidy - University of Missouri	Modelling lipidated apoB-100 from human low-density lipoprotein
12:20-12:45 Contributed Talk 8	Chandra Verma - Agency for Science, Technology and Research (A*STAR)	Translating molecular simulations towards informing clinician decisions
12:45-13:45	Lunch	

Day 2 – Tuesday 15th July - Afternoon

Session 4: AI and Simulation Chair:		
13:45-14:25 **Invited** Talk 6	Greg Ross – Isomorphic Labs	
14:25-14:50 Contributed Talk 9	Joao Morado - University of Edinburgh	Enhancing Electrostatic Embedding for ML/MM Free Energy Calculations: Strategies for Accurate and Transferable Models
14:50-15:15 Contributed Talk 10	Ying-Chih Chiang - The Chinese University of Hong Kong Shenzhen	Toward AI-Assisted Antimicrobial Peptide Design
15:15-15:45	Tea and Coffee	
15:45-16:25 **Invited** Talk 7	Hannah Bruce Macdonald - CHARM Therapeutics	
16:25-17:05 **Invited** Talk 8	Natasja Brooijmans – Antares Therapeutics	
17:05-17:30 Contributed Talk 11	Yanchen Zhu – University of Edinburgh	A Deep Generative Model for Sampling Protein Transition States
17:30-19:00	Posters (Even Numbers)	
19:00-19:30	Travel to Grand Café (via coach)	
19:30-	Conference Dinner	

Day 3 – Wednesday 16th July

Session 5: Theory, Application and Design: part 2 Chair:		
09:00-09:40 **Invited** Talk 9	Anita De Ruiter - BOKU University	
09:40-10:05 Contributed Talk 12	Victor Sebastian Perez – Sandbox AQ	AI and LQM-Driven methodologies to accelerate drug discovery
10:05-10:45 **Invited** Talk 10	Gábor Csányi – Cambridge University	
10:45-11:15	Tea and Coffee	
11:15-11:55 **Invited** Talk 11	Michael Shirts - University of Colorado Boulder	
11:55-12:20 Contributed Talk 13	Jasmin Güven – University of Edinburgh	Potential β-lactamase inhibitors under the alchemical microscope
12:20-12:45 Contributed Talk 14	Edina Rosta – University College London	Enhanced Sampling Simulations of Biomolecular Systems
12:45-	Close and Lunch	

Poster Presenters

1	**Chenggong Hui**	Open-source Grand Canonical Monte Carlo Package for enhancing water sampling in free energy calculations
2	**Ka Fu Man**	Impact of pathogenic missense mutations on the dynamics of human skeletal myosin
3	**Venkat Ramaswamy**	Active learning FEP using 3D-QSAR for prioritizing bioisosteres in medicinal chemistry
4	**Juliana de Abrantes**	Nuclear Quantum Effects in Methylated DNA Base Pairs via NEODFT and Machine-Learned Potentials
5	**Katarina Blow**	CCD2MD: A suite of packages for ease of simulation using co-folding outputs
6	**Charlotte Lynch**	The MemProtMD Database for Molecular Dynamics Simulations of Membrane Proteins and the MDDB Project
7	**William Houppy**	Unveiling the Catalytic Mechanism of Alginate Lyases via QM/MM
8	**Chin Yong**	D_ATA (Atom Typer and Analyser) - a new software to investigate atomic interactions behaviour
9	**Finlay Clark**	Fast Parameterisation of Bespoke Classical Force Fields: Beyond Torsion Fitting
10	**Ying-Chih Chiang**	A Partition Function Estimator
11	**Barbara Abreu**	Multiscale simulations of styrene-maleic acid co-polymers (SMALPs)
12	**Breanna Voss**	Using Computational Methods to Identify Post-Translational Modifications on Lysine Amino Acids
13	**James Davies**	Increased functional unit flexibility and solvent accessibility favours oxygen capture in molluscan hemocyanin
14	**Wojciech Kopec**	Computational Electrophysiology of Potassium and Chloride Channels
15	**Justina Ratkeviciute**	Improving Alchemical Binding Free Energy Calculations Using Fully Adaptive Simulated Tempering (FAST)
16	**Peter Bond**	An Integrative Approach to Transforming Endogenous Molecules into Drugs
17	**Wenhao Deng**	Reaction Mechanism and Metal Selectivity of Human SAMHD1 Elucidated by QM/MM Calculations
18	**Pornpan Pungpo**	Investigation of key interactions between Thai natural products and the PknB binding pocket using molecular docking calculations
19	**Zhujun Liu**	Investigating CFTR-targeted therapies for cystic fibrosis using molecular docking and MD simulation
20	**Victoria Nathan-Maister**	Title

21	**Matthew Burman**	SOMD2: a modular and extensible open-source engine for GPU-accelerated free energy calculations
22	**Yu-Yuan (Stuart) Yang**	Holo-like conformation selection using a computer vision-based deep-learning model
23	**Lorenzo Tulli**	Capturing the dynamical responses of the epidermal growth factor receptor to reveal the impact of clinically relevant mutations in cancer
24	**Charlie Holdship**	Molecular Modelling of Conotoxin Peptides as a Route to Antitoxin Design
25	**Kazi Hossain**	When lipids embrace RNA: mechanistic insights to LNP-mediated delivery
26	**Kin Chao**	How Well Does REST2 Perform in Conformational Sampling? A Case Study on Short and Disordered Peptides
27	**Asma Feriel Khoualdi**	De Novo Design of GALK1 Inhibitors in a Flexible Binding Pocket
28	**Wenhao Deng**	Reaction Mechanism and Metal Selectivity of Human SAMHD1 Elucidated by QM/MM Calculations
29	**Thomas Osborne**	Molecular Dynamics Simulations Reveal Lipid and G Protein-Dependent Mechanisms of GPCR Dimerisation
30	**Konstantinos Tornesakis**	Title
31	**Bjarne Feddersen**	Glycine receptors contain state-dependent, drug-permeable fenestrations
32	**Jana Pavlikova Precechtelova**	Conformational Landscapes from Molecular Dynamics Simulations of Intrinsically Disordered Proteins
33	**James Robins**	Multiscale Modelling for RNA Therapeutics: Molecular Dynamics to Understand Polymer-RNA Interactions in Polymeric Nanoparticles
34	**Valeria Losasso**	Refining molecular simulations of lipid monolayers using neutron reflectivity: the critical role of area per lipid
35	**Sarah Fegan**	CodeEntropy: The Multiscale Cell Correlation Method
36	**Jas Kalayan**	BioSim Data Resources - Capture Full Simulation Provenance and Deposit Data to BioSimDB
37	**Ahmed Elgaziari**	Using ML and MMPBSA for free energy predictions for protein-protein complexes
38	**Abbie Lear**	Understanding evolutionary improvements in designer enzymes to inform computational design
39	**Benedict Tan**	Protein complex structural prediction for molecular glue design
40	**Krithik Shai Murali Padma Rekha**	Computational Investigation of Lgt activity in Bacterial Lipoprotein post translational modifications

41	**Chi Kit Ng**	Active Learning-Driven Workflow for Molecular Design: Integrating CReM and FEgrow
42	**Matilda Ymeraj**	QM/MM investigation of the catalytic mechanism of adenosine phosphosulphate reductase (APSR)
43	**Ziwei Pang**	Leveraging Multiple Short MD Simulations for Screening PepT2 Transporter Substrates
44	**Nga Man Cheng**	Impact of mRNA structures on the interaction with lipids and nanoparticle formulation properties
45	**Ziad Fakhoury**	Building Markov Models For Protein Folding with Graph Driven Search
46	**Hima Bindu Kolli**	Workflow for molecular dynamics simulations from CCPNmr Analysis to AMBER
47	**Asal Azar**	Structural Dynamics Analysis of Nitrite Reductase AniA
48	**Ivan Manoza**	Reducing Drug Cardiac Toxicity Using Molecular Simulations and Machine Learning
49	**Yixin He**	Human Cardiac Sodium Channel (hNaV1.5) Modelling for Cardiac Safety Prediction in Drug Discovery
50	**Harry Swift**	Empowering Biomolecular Research: Modernising CodeEntropy for Scalability and Usability
51	**Robert Welch**	Engineering Supercomputing Platforms for Biomolecular Applications
52	**Robert Clark**	Cooling fast and slow: Recovering equilibrium statistics from vitrified cryo-EM samples
53	**Chenfeng Zhang**	Assessing Mutations and Allosteric Ligands on the PKD2 Ion Channel through Conformational Probabilities
54	**Pedro J. Buigues**	Fast & Fourier Features for Transferable ML Potentials
55	**Leonardo Cirqueira**	Dynamical assessment of the MOR1-PKR Complex
56	**Max Cutler**	Development of a Mixed Resolution Protein-RNA LLPS Model
57	**Huong Vu**	Investigating structural ensembles of artificial RNA condensates using UNISIS model
58	**Bahar Alizadeh**	Distinct Binding Modes of Nitazene Analogues at the Œº-Opioid Receptor Revealed by Molecular Dynamics Simulations
59	**Vladimir Kozyrev**	Computational Insights into Ion-Mobility Mass Spectrometry for RNA Therapeutics
60		
61		
62		
63		
64		
65		

Speaker Abstracts – in order of appearance

Directed Evolution of Cytidine Deaminase for Biocatalysis: Simulation Insights for Enzyme Design and Engineering

Marko Hanzevacki[1]; Carlos A. Ramos-Guzmán[1]; Anthony P. Green[2]; Adrian J. Mulholland[1]
[1]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK*
[2]*Department of Chemistry, Manchester Institute of Biotechnology, University of Manchester, Manchester M1 7DN, UK*

Presenting author: marko.hanzevacki@bristol.ac.uk

Abstract

Cytidine deaminase (CD) is a zinc-dependent metalloenzyme that catalyses the hydrolytic deamination of cytidine to uridine. It has been engineered through directed evolution to utilise hydroxylamine as a substrate for the biosynthesis of *N*-hydroxy-cytidine (NHC), a key precursor for the COVID-19 drug molnupiravir [1]. We have combined experimental structural biology with biomolecular simulations to analyse the effects of directed evolution of this enzyme and elucidate the molecular mechanisms underlying the altered activity of a novel CD variant, CAT4, which carries eleven distal point mutations. CAT4 exhibits significantly reduced cytidine hydrolysis, and a markedly improved NHC-to-uridine product ratio, compared to wild-type CD. Analysis of the CAT4 crystal structure reveals substantial conformational rearrangements and a more negatively charged enzyme surface. Molecular dynamics simulations show that these mutations induce surface conformational changes that restrict water access to the active site, potentially modulating catalytic activity. Local electric fields (LEFs) are important in determining catalytic activity in many enzymes. Analysis of the LEF around the catalytic base Glu104 shows that it is significantly altered in CAT4, impacting the initial deprotonation of zinc-bound water or hydroxylamine. Extensive DFTB3/MM molecular dynamics simulations of the reaction, using the adaptive string method, provide mechanistic insights into how the distal mutations in CAT4 modulate the hydrolytic activity of CD. Altogether, the results help to explain and rationalise the effects of directed evolution, and show that simulations have the potential to contribute to practical programmes of enzyme design and engineering.

References

[1] Burke, A. J., Birmingham, W. R., Zhuo, Y., Thorpe, T. W., Zucoloto da Costa, B., Crawshaw, R., Rowles, I., Finnigan, J. D., Young, C., Holgate, G. M., Muldowney, M. P., Charnock, S. J., Lovelock, S. L., Turner, N. J., and Green, A. P. (2022). *JACS*, **144**(9), 3761-3765.

Integrating SAXS, MD and coarse-grained modelling for all-atom protein ensemble generation

Cameron Brown, Josh McKeown, Arron Bale, Hayden Fisher, Matteo Degiacomi, Christopher Prior, Robert P. Rambo, Jonathan W. Essex

[1] *School of Chemistry and Chemical Engineering, University of Southampton, Southampton, SO17 1BJ*
[2] *Diamond Light Source, Harwell Science and Innovation Campus, Didcot, OX11 0DE*
[3] School of Mathematical and Computer Science, Durham University, DH1 3LE

Corresponding author: cameron.brown@soton.ac.uk

Abstract

Small Angle X-ray Scattering (SAXS) is a low-resolution biophysical technique that reveals the shape and size of biological macromolecules in near-native conditions. In contrast, MD simulations provide atomistic resolution structural ensembles. The rise of ML-based structure prediction methods such as AlphaFold have revolutionised structural biology. Despite this, proteins exist as a dynamic ensemble of structures, with many ML predictions and crystal structures not agreeing with their corresponding ensemble averaged solution SAXS profiles. To this end, we apply an integrative, semi-automated modelling protocol utilising SAXS, MD and a novel coarse-grained structure optimisation algorithm to bridge the gap between static protein structure prediction and the dynamic reality.[1]

We demonstrate applicability through two examples. Firstly, demonstrating the intrinsic plasticity of the Rift Valley Fever Gn antigen, which is hypothesised to aid higher order assembly formation and viral cell entry. While an open conformation of the Gn crystal structure was not obtainable, MD simulations of the closed Gn structure explore domain flexibility, supported by the experimental SAXS. Our second use case builds on previous X-ray crystallographic and SAXS studies on hinge disulfide-engineered IgG2 $F(ab)_2$ fragments holding promise for cancer therapeutics.[2] MD simulations of crystal $F(ab)_2$ structures produced atomistic ensembles that did not cover the necessary radius of gyration (Rg) range from which accurate ensembles could be selected. Here, we recover the necessary ensemble Rg range through SAXS-driven structural optimisation. $F(ab)_2$ structures are restrained by interatomic distances between the C-alpha atoms of opposing disulfide bonding cysteine residues, while extending the F(ab) arms outwards in accordance with the experimental scattering data. Following all-atom modelling, secondary structure is largely maintained while the resultant reweighted MD ensembles exhibit an Rg range in line with that suggested by the experimental scattering.

References

[1] C. Prior, O. R. Davies, D. Bruce and E. Pohl, Journal of Chemical Theory and Computation, 2020, 16, 1985–2001.
[2] C. Orr, H. Fisher, X. Yu, C. Chan, Y. Gao, P. Duriez, S. Booth, I. Elliott, T. Inzhelevskaya, I. Mockridge, C. Penfold, A. Wagner, M. Glennie, A. White, J. Essex, A. Pearson, M. Cragg and I. Tews, Science Immunology, 2022, 7, 73–84

Atomic resolution ensembles of intrinsically disordered and multi-domain proteins with Alphafold

Max Bonomi 1[1]; Vincent Schnapka[1]
[1]*Institut Pasteur, Université Paris Cité, CNRS UMR 3528, Computational Structural Biology Unit, Paris, France*

Corresponding author: mbonomi@pasteur.fr

Abstract

Intrinsically disordered proteins are ubiquitous in biological systems and play essential roles in a wide range of biological processes and diseases. Despite recent advances in high-resolution structural biology techniques and breakthroughs in deep learning-based protein structure prediction, accurately determining structural ensembles of IDPs at atomic resolution remains a major challenge. Here we introduce bAIes [1], a Bayesian framework that integrates AlphaFold2 predictions with physico-chemical molecular mechanics force fields to generate accurate atomic-resolution ensembles of IDPs. We show that bAIes produces structural ensembles that match a wide range of high- and low-resolution experimental data across diverse systems, achieving accuracy comparable to atomistic molecular dynamics simulations but at a fraction of their computational cost. Furthermore, bAIes outperforms state-of-the-art IDP models based on coarse-grained potentials as well as deep-learning approaches. Our findings pave the way for integrating structural information from modern deep-learning approaches with molecular simulations, advancing ensemble-based understanding of disordered and multi-domain proteins.

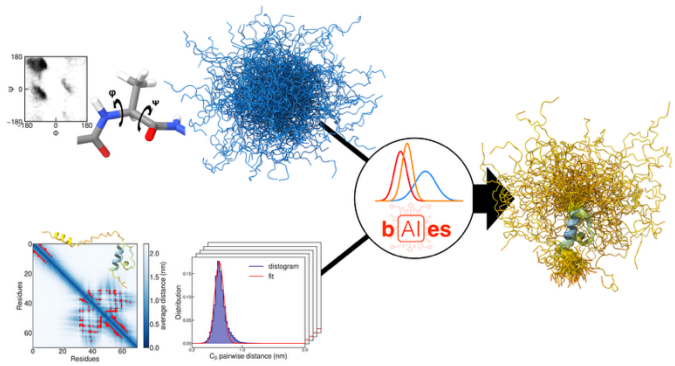

Figure 1: *Schematic overview of bAIes.*

References

[1] V. Schnapka, T. I. Morozova, S. Sen, M. Bonomi. Atomic resolution ensembles of intrinsically disordered and multi-domain proteins with Alphafold. bioRxiv (2025) doi:10.1101/2025.06.18.660298.

Energy-entropy analysis of interactions between CLR/RAMP ectodomain complexes and small molecule antagonists

Ioana A. Papa[1], Sarah K. Fegan[2], Harry Swift[2], Jas Kalayan[2], James Gebbie-Rayet[2], Richard H. Henchman[3], Joseph L. Egan[1,4], Georgia A. Taylor-Vine[1,4], Alistair Keys[4], Paris Avgoustou[4], Ameera B. A. Jailani[4], Joseph P. A. Harrity[1], Gareth O. Richards[4], Sarah A. Harris[1]

[1] *School of Mathematical and Physical Sciences, University of Sheffield, UK*
[2] *Scientific Computing Department, Science and Technology Facilities Council, Daresbury Laboratory, UK*
[3] *School of Medicine, University of Sydney, Australia*
[4] *Clinical Medicine, School of Medicine and Population Health, University of Sheffield, UK*

Corresponding author: iapapa1@sheffield.ac.uk

Abstract

As non-covalent association processes are ubiquitous in biology, their quantification is crucial. One approach for calculating binding free energies consists of energy-entropy methods. The energy of a system gives information concerning the strength of molecular interactions and can be easily computed, while entropy relates to dynamics within the system, but is often overlooked and more difficult to calculate in a computationally efficient and accurate manner. [1,2] However, assessing this term is also important for understanding these processes as protein and solvent dynamics have been established to play a significant role in governing the mechanisms of binding. [3] Multiscale cell correlation (MCC) yields the entropy of a system by discretising configuration space into different length scales, as well as vibrational and topographical terms, thus giving rise to lower order terms which are easier to compute. This approach allows for detailed insights into contributions to entropy changes occurring upon binding, as well as for scalability and fast convergence. Another advantage of this method is that it treats all molecules in a system equivalently and hence, can be used for a complete analysis of the system, including both solutes and solvent. [1,4]

G-protein coupled receptors (GPCRs) form the largest family of integral membrane proteins and are extremely important drug targets, as a significant number of diseases involve their malfunction. [5] One such class B GPCR is the calcitonin receptor-like receptor (CLR), which associates with one of the three human receptor activity-modifying proteins (RAMPs), giving rise to complexes with different pharmacological and physiological properties. These complexes form the calcitonin gene-related peptide (CGRP) receptor and two adrenomedullin (AM) receptors, referred to as AM_1 and AM_2. AM has an important role in controlling blood pressure and has been associated with tumor progression and sepsis. [6]

The binding of small molecule antagonists developed for the AM receptors to the CLR/RAMP complexes has been thermodynamically characterised. MCC has allowed a detailed breakdown of changes in configurational entropy terms of the protein-protein complex, ligands and solvation water molecules. Furthermore, the different length scales have allowed for assessing local entropy changes and gaining insights into residues' individual contributions to binding. Free energies and entropies calculated are discussed alongside experimental measures of potency and results obtained reflect expected binding trends for the ligands considered.

References

[1] Jas Kalayan, Arghya Chakravorty, Jim Warwicker, and Richard H. Henchman. Total free energy analysis of fully hydrated proteins. Proteins: Struct., Funct., Bioinf., 91(1):74–90, August 2022

[2] Hafiz Saqib Ali, Arghya Chakravorty, Jas Kalayan, Samuel P. de Visser, and Richard H. Henchman. Energy–entropy method using multiscale cell correlation to calculate binding free energies in the SAMPL8 host–guest challenge. J. Comput.-Aided Mol. Des., 35(8):911–921, July 2021

[3] Kendra King Frederick, Michael S Marlow, Kathleen G Valentine, and A Joshua Wand. Conformational entropy in molecular recognition by proteins. Nature, 448(7151):325–329, July 2007.

[4] Arghya Chakravorty, Jonathan Higham, and Richard H. Henchman. Entropy of proteins using multiscale cell correlation. Journal of Chemical Information and Modeling, 60(11):5540–5551, September 2020.

[5] Yang Zhang, Mark E DeVries, and Jeffrey Skolnick. Structure modeling of all identified g protein–coupled receptors in the human genome. PLoS Computational Biology, 2(2):e13, February 2006.

[6] Paris Avgoustou, Ameera B. A. Jailani, Jean-Olivier Zirimwabagabo, Matthew J. Tozer, Karl R. Gibson, Paul A. Glossop, James E. J. Mills, Roderick A. Porter, Paul Blaney, Peter J. Bungay, Ning Wang, Alice P. Shaw, Kamilla J. A. Bigos, Joseph L. Holmes, Jessica I. Warrington, Timothy M. Skerry, Joseph P. A. Harrity, and Gareth O. Richards. Discovery of a first-in-class potent small molecule antagonist against the adrenomedullin-2 receptor. ACS Pharmacology & Translational Science, 3(4):706–719, June 2020

In Search of Enzyme Activation: A Markov state model approach to allosteric drug discovery

Frederick G. Powell[1]; Adele Hardie[2], Julien Michel[2], Stephen J. Yarwood[3], Graeme Barker[1]
[1]*Institute of Chemical Sciences, Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS*
[2]*EaStCHEM School of Chemistry, University of Edinburgh, King's Buildings, Edinburgh, EH9 3FJ*
[3]*Institute of Biological Chemistry, Biophysics, and Bioengineering, Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS*

Corresponding author: F.Powell@hw.ac.uk

Abstract

Enzyme activation is an underexplored mode of pharmaceutical intervention, relying on exploiting delicate allosteric networks. We have developed a workflow using steered molecular dynamics (sMD) and Markov state models (MSMs) to assess the allosteric potential of small molecule modulators of enzyme activity (see Figure 1).[1] Our workflow uses sMD to explore the conformational space of the target system. By employing sMD, we can sample conformational space that is inaccessible under routine MD timescales. Subsequently, we utilise intermediate conformations arrived at *via* sMD as the starting point for multiple short, equilibrium MD simulations. The resulting data is pooled and used to construct MSMs, affording us insight into the metastable conformational states of the target protein, as well as the probability of the protein occupying said states with and without an allosteric modulator *in situ*.

We have applied our workflow to investigate small molecule activators of the cell signalling enzyme, Epac1, a key drug target for the treatment of chronic inflammation.[2] Using our model, we have demonstrated Epac1 activation by the endogenous activator, secondary messenger cyclic AMP (cAMP), as well as by the hit compound I942, a partial Epac1 activator.[3] This has afforded us mechanistic insight into the key interactions required for activation, guiding the development of a novel, efficacious, and selective Epac1-activating lead compound.

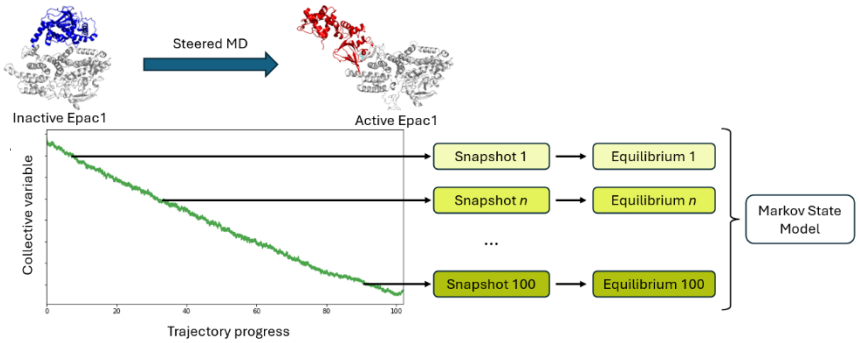

Figure 1: *A schematic overview of our sMD/MSM workflow.*

References

[1] A. Hardie, *et al.* (2023). *Commun. Chem.*, **6**(125), 1-9
[2] E. Parnell, *et al.* (2011). *Br. J. Pharmacol.*, **166**(2), 434-446
[3] E. Parnell, *et al.* (2017). *Sci. Rep.*, **7**(294), 1-12

Multi-fidelity learning and meta-learning for interatomic potentials

Alice Allen [1, 2]

1 Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545,USA
2 Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz,Germany

Corresponding author: allena@mpip-mainz.mpg.de

Abstract

Machine learning interatomic potentials offer a new route to model biomolecular systems at high accuracy. In recent years, many quantum mechanical datasets have been built to capture the chemical space explored in molecular simulations. However, combining quantum mechanical datasets from different sources can be problematic, particularly if the level of theory used differs. I will discuss two approaches to overcome this issue: meta-learning and multi-fidelity learning. These methods offer strategies to effectively combine multiple datasets and levels of theory, enabling more accurate modeling of complex molecular systems.

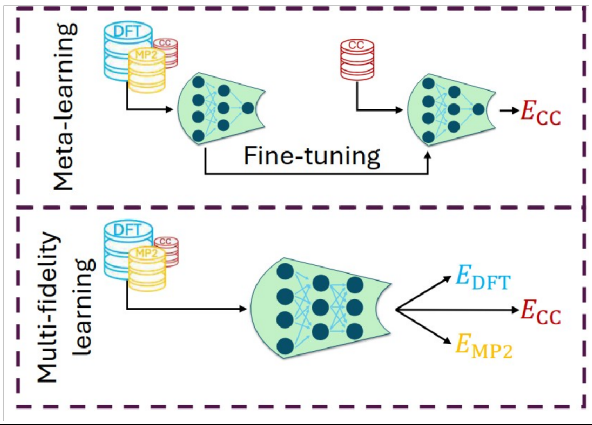

Figure 1: *A comparison of different methods capable of training to multiple datasets at different levels of theory.*

References

[1] Mitchell Messerly, Sakib Matin, Alice E. A. Allen, Benjamin Nebgen, Kipton Barros, Justin S. Smith, Nicholas Lubbers and Richard Messerly, Multi-fidelity learning for interatomic potentials: Low-level forces and high-level energies are all you need, 2025, https://arxiv.org/abs/2505.01590

[2]Alice E. A. Allen, Nicholas Lubbers, Sakib Matin, Justin Smith, Richard Messerly, Sergei Tretiak and Kipton Barros, Learning together: Towards foundation models for machine learning interatomic potentials with meta-learning, 2024, npj Computational Materials , 10, 154 (2024)

Integrative modelling of the mitochondrial cristae

Chelsea M. Brown[1]; M.S.S. Westendorp[1]; R. Zarmiento-Garcia[1]; J.A. Stevens[1]; B.M. Bruininks[1]; S.L. Rouse[2]; S.J. Marrink[1]; T.A. Wassenaar[1,3]

[1] *University of Groningen, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), Groningen, The Netherlands*
[2] *Imperial College London, Department of Life Sciences, London, United Kingdom*
[3] *Knowledge Center Biobased Economy, Hanze University for Applied Sciences, Groningen, The Netherlands*

Corresponding author: t.a.wassenaar@rug.nl

Abstract

The signature feature of the inner mitochondrial membrane is the curved morphology; the invaginations are called cristae, which house membrane proteins that are essential to mitochondrial function. Aspects of mitochondria have been studied *in silico*, but combining features into one representative model remains challenging. The objective of this work was to create an initial model of a human mitochondrial crista, including appropriate lipid composition and curvature, and the most relevant membrane protein complexes. With this objective in mind, we collated structures of 13 protein complexes, with 113 unique chains, using experimentally resolved structures in combination with AlphaFold models and biochemical data for validation. These structures were then embedded in membranes with lipid compositions and curvatures to reflect experimentally determined values. The Martini 3 forcefield was used to assemble and simulate this model, prior to performing initial analysis. Alongside the simulation results, this work provides a framework for combining experimental data from cryo-EM, biochemical studies, structural modelling, and molecular simulations to create a system that captures the complexity of biological membranes *in situ*. Overall, we present an initial 'living' model of a human mitochondrial crista, intended to be built upon and improved as our understanding, methodology and resources develop [1].

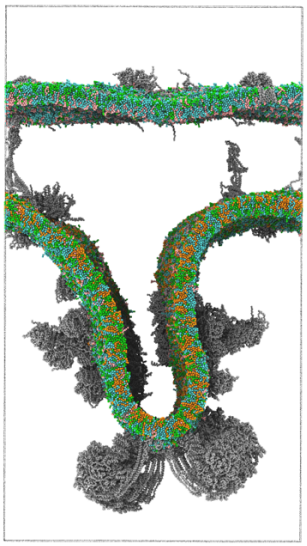

Figure 1: Crista junction model. Proteins are shown in grey, while the lipids are colored according to type.

References

[1] Brown, C. M., Westendorp, M. S., Zarmiento-Garcia, R., Stevens, J. A., Rouse, S. L., Marrink, S. J., & Wassenaar, T. A. (2024). An integrative modelling approach to the mitochondrial cristae. *BioRxiv*, 2024-09.

Modelling lipidated apoB-100 from human low-density lipoprotein

C. Keith Cassidy[1]; Robin A. Corey[2]; Zachary T. Berndsen[3]
[1] *Department of Physics, University of Missouri-Columbia, U.S.A.*
[2] *Department of Pharmacology, University of Bristol, U.K.*
[3] *Department of Biochemistry, University of Missouri-Columbia, U.S.A*

Corresponding author: ckcassidy@missouri.edu

Abstract: Low-density lipoprotein (LDL) plays a vital role in human health by transporting cholesterol and triglycerides to tissues, yet its accumulation and oxidation in arterial walls can trigger atherosclerosis and cardiovascular disease, a leading cause of death worldwide. LDL particles are highly heterogeneous with a hydrophobic core of cholesterol esters and triglycerides, encased in a phospholipid monolayer. A single copy of apolipoprotein B-100 (apoB-100) wraps around the particle to maintain cohesion and mediate clearance via the LDL-receptor. Despite its immense biological and clinical significance, apoB-100 has defied high-resolution structural determination for decades due to its massive size (4563 residues), flexibility, and complex lipid associations. Here, we report the first full-length structures of human apoB-100, derived for two different particle sizes, using an integrative approach of cryo-electron microscopy (cryo-EM), AlphaFold2, and molecular dynamics (MD)-based flexible refinement [1]. Our structures reveal a detailed picture of the conformational changes undergone by apoB-100 as it adapts to particles of different sizes and lipid compositions. Building on this structural framework, we have further developed a Martini-based coarse-grained (CG) self-assembly protocol to generate fully lipidated LDL particles based on multiple apoB-100 scaffolds. Together, these results set the stage for the use of multi-scale molecular simulations to systematically investigate how lipid composition and oxidation influence LDL core structure and phase, and how conformation-specific apoB-lipid interactions relate to disease-associated mutations, thereby advancing our molecular understanding of LDL's roles in both health and disease.

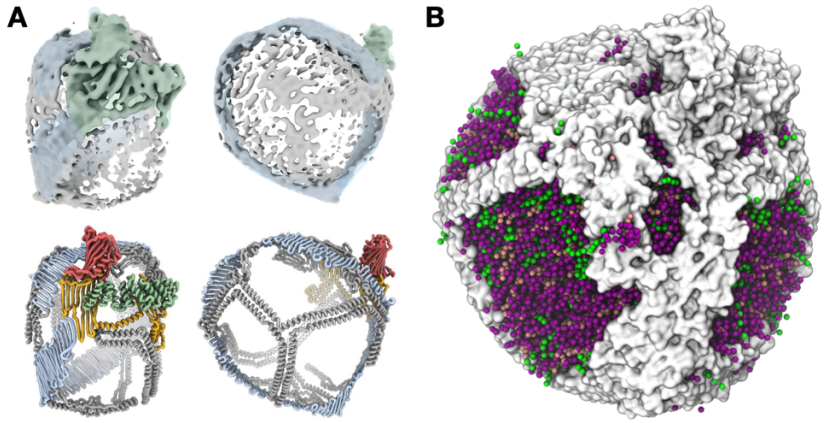

Figure 1: *(A) Cryo-EM density map of human low-density lipoprotein (top) and corresponding all-atom model of apoB-100 (bottom). (B) Model of lipidated apoB-100, constructed via coarse-grained self-assembly.*

References

[1] Berndsen, Z. T. and Cassidy, C. K. (2025). *Nature*, **638**(8051), 836-843.

Enhancing Electrostatic Embedding for ML/MM Free Energy Calculations: Strategies for Accurate and Transferable Models

João Morado[1]; Kirill Zinovjev[2]; Lester Hedges[3]; Daniel J. Cole[4]; Julien Michel[1]
[1]*EaStCHEM School of Chemistry, University of Edinburgh, United Kingdom*
[2]*Departament de Química Física, Universitat de València, Spain*
[3]*OpenBioSim Community Interest Company, United Kingdom*
[4]*School of Natural and Environmental Sciences, Newcastle University, United Kingdom*
Corresponding author: jmorado@ed.ac.uk

Abstract

Hybrid ML/MM approaches offer a promising balance between computational cost and accuracy, largely thanks to continued improvements in the performance of machine learning potentials (MLPs). These advancements are making MLPs increasingly viable for achieving quantum mechanical (QM)-level accuracy at a cost approaching that of molecular mechanics (MM). Compared to well-parametrised MM force fields, the fixed-charge, non-polarisable description of mechanical embedding ML/MM schemes has so far not led to significant improvements in the accuracy of free energy simulations. These schemes typically rely on Lennard-Jones (LJ) and Coulomb terms to model intermolecular interactions, lacking an explicit treatment of the polarisation response of an MLP to its MM environment.

Recently developed electrostatic embedding methods, such as the electrostatic embedding of machine learning potentials (EMLE)[1], have been specifically designed to account for these missing polarisation effects. EMLE holds the potential to improve the predictive power of ML/MM free energy simulations, a crucial step in both academic and industrial pipelines focused on the design and discovery of new molecules.
In this talk, we present our latest research on improving the accuracy of EMLE models and extending their applicability. Specifically, we showcase robust methodologies for training EMLE models using QM data, demonstrating their effectiveness through an absolute hydration free energy benchmark on a set of small organic molecules. We show that our training workflows can produce EMLE models that are competitive in accuracy with MM force fields and mechanical embedding ML/MM schemes, and that offer improvements for classes of drug-like molecules where the latter fall short.

Throughout the talk, we also provide insights into the development of polarisable force fields, highlighting important considerations to have in their design. We conclude by presenting ongoing efforts to derive sets of EMLE-compatible LJ parameters, along with our recent work on geometry-dependent EMLE-LJ parameters. These developments could have significant impact in scenarios where dynamically varying dispersion coefficients is essential, such as in reactive modelling.

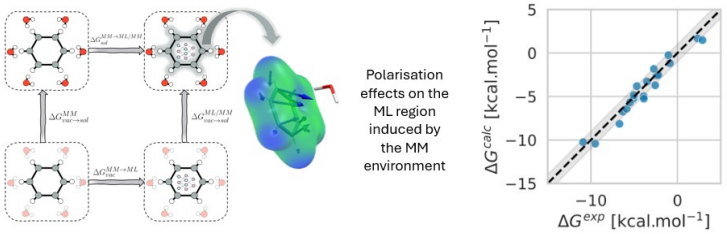

Figure 1: *Left: Schematic of the thermodynamic cycle used to calculate electrostatically embedded hydration free energies. Right: EMLE vs. experimental hydration free energy predictions.*

References

[1] Zinovjev, K. Electrostatic Embedding of Machine Learning Potentials. *J. Chem. Theory Comput.* **2023**, *19* (6), 1888–1897.

Toward AI-Assisted Antimicrobial Peptide Design

Lantian Yao[1,2]; **Jiahui Guan**[1]; **Peilin Xie**[1,2]; Ying-Chih Chiang[1,2]; Tzong-Yi Lee[3]
[1]*KIIDD, School of Medicine, The Chinese University of Hong Kong, Shenzhen, 518172, China*
[2]*School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, 518172, China*
[3]*Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan*

Corresponding author: chiangyc@cuhk.edu.cn; leetzongyi@nycu.edu.tw

Abstract
Antimicrobial peptides (AMPs) are peptides with antibacterial, antiviral, antifungal, or antiparasitic activities. They are often secreted by living organisms and present a promising alternative to existing antibiotics. Traditionally, their discovery through experimental methods was costly and time-consuming. We employ advanced artificial intelligence (AI) techniques to address the challenges in designing new antimicrobial peptides. These challenges include identifying AMP sequences, predicting activity levels, forecasting hemolytic toxicity, and optimizing other properties by modifying the sequence. Additionally, molecular dynamics simulations are utilized to explore the mechanism of action of AMPs, elucidating how a single residue can influence pore formation on the membrane.

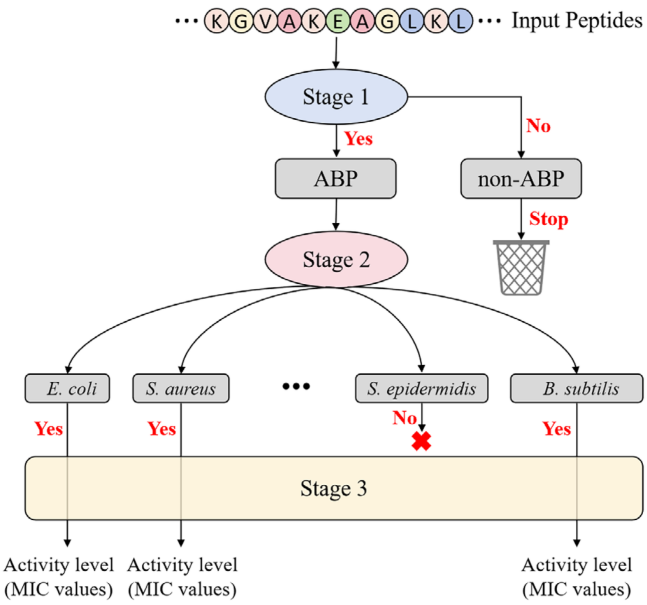

Figure 1: *The flowchart of AMPActiPred, a tool designed to identify an AMP sequence and predict its activity level against ten bacterial species.*

References

[1] Yao, Lantian, Guan, Jiahui, Xie, Peilin, Chung, Chia-Ru, Deng, Junyang, Huang, Yixian, Chiang, Ying-Chih, Lee, Tzong-Yi. (2024). *Protein Science*, **33**(6), e5006.
[2] Yao, Lantian, Guan, Jiahui, Xie, Peilin, Chung, Chia-Ru, Zhao, Zhihao, Dong, Danhong, Guo, Yilin, Zhang, Wenyang, Deng, Junyang, Pang, Yuxuan, Liu, Yulan, Peng,Yunlu, Horng, Jorng-Tzong, Chiang, Ying-Chih, and Lee, Tzong-Yi. (2025) Nucleic Acids Research, **53**(D1): D364-D376.

A Deep Generative Model for Sampling Protein Transition States

Y. Zhu[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] *School of Informatics, University of Edinburgh, Dugald Stewart Building, 10 Crichton Street, Edinburgh, EH8 9AB, United Kingdom*

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

Abstract

Proteins are dynamic molecules who continuously change conformations to interact with other biomolecules. Capturing these fleeting transitions between conformational states is crucial because they often encode the key mechanisms that drive biological processes. Experimental techniques, including X-ray crystallography, cryo-EM and NMR spectroscopy, have yielded an extensive library of high-resolution protein structures, but provide limited insight into dynamic pathways. Computation simulations, for example, molecular dynamics (MD) simulations, complement the experimental observations by providing atomistic movies of protein motions. However, the timescales can be achieved by such physics-based simulations are limited, often orders of magnitude shorter than timescales in which rare transitions occur. Moreover, MD is prone to be trapped in low-energy, metastable minima, spending vast computational resources on stable conformers while leaving transition states under-sampled.

Generative models, as a branch of deep learning, has emerged as a powerful tool in fields such as image and video generation and it offers a novel route to sampling protein transition states. We introduce Molearn [1], an autoencoder-based generative model that learns conformational ensemble of a protein and predicts its transition-state conformers. Trained on all-atom snapshots from MD simulations seeded with crystallographic data, Molearn generalises to unseen transition states by latent space sampling to generate 3D coordinates of transition-state conformers that obey physics constraints. After energy refinements, these predicted structures trace realistic paths between metastable states and can serve as seeding structures for focused MD sampling along the transition path. By running parallel seed-based simulations and applying Markov state modelling [2], one can reconstruct transition kinetics. Overall, Molearn enables more efficient exploration of protein conformational change.

References

[1] Musson, S. C. and Degiacomi, M. T., (2023). *J. Open Source Softw.*, **89**, 5523
[2] Arbon, R. E., Zhu, Y., and Mey, A. S. J. S. (2024). *J. Chem. Theory Comput.*, **20** (2), 977–988.

AI and LQM-Driven methodologies to accelerate drug discovery

Víctor Sebastián Pérez[1]; Andrea Bortolato[1]

[1] *SandboxAQ, Palo Alto, California, USA.*

Abstract

The evolution of drug discovery is increasingly driven by remarkable advances in AI, simulation, and data integration technologies. SandboxAQ generates proprietary data using physics-based methods, and trains Large Quantitative Models (LQMs) on that data, leading to new insights in areas, such as life sciences, energy, chemicals, and financial services. Structure-based methods in drug discovery have become an integral part of the modern drug discovery process. Recent advances in free energy binding prediction methods and generative AI solutions are transforming medicinal chemistry.

To address the challenges in drug discovery, we have developed a novel approach that integrates AI protein structure prediction, AI binding affinity predictions, and physics-based methodologies to better understand protein flexibility and ligand affinity. Additionally, our approach emphasizes the inclusion of diverse data types, enhancing the accuracy and relevance for projects at different stages.

We will highlight and present the application of this method to a set of systems, showing the performance of our methodology and the evaluation of novel drug discovery solutions for hit finding and lead optimization applied to the promising drug discovery targets for neurodegeneration. Examples will include i) active learning absolute free energy perturbation (AQFEP) virtual screening; ii) Alchemical Transfer Method (ATM - Tango) for the estimation of relative ligand binding free energies; iii) IDOLpro, a new generative AI solution combining deep diffusion with multi-objective optimization for structure-based drug design; and iv) knowledge-graph based solutions.

References

[1] Bryant, P., Pozzati, G., & Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. *Nature communications*, *13*(1), 1265.
[2] Kadan, A., Ryczko, K., Lloyd, E., Roitberg, A., Yamazaki, T. (2025) Guided multi-objective generative AI to enhance structure-based drug design. DOI: 10.1039/D5SC01778E Chemical Science
[3] Qiao, Z., Nie, W., Vahdat, A., Miller III, T. F., & Anandkumar, A. (2024). State-specific protein–ligand complex structure prediction with a multiscale deep generative model. *Nature Machine Intelligence*, *6*(2), 195-208.
[4] Crivelli-Decker, J. E., Beckwith, Z., Tom, G., Le, L., Khuttan, S., Salomon-Ferrer, R., ... & Bortolato, A. (2023). Machine Learning Guided AQFEP: A Fast and Efficient Absolute Free Energy Perturbation Solution for Virtual Screening. *Journal of Chemical Theory and Computation*.

Potential β-lactamase inhibitors under the alchemical microscope

J. Jasmin Güven[1]; Marko Hanževački[2]; Papu Kalita[2]; Adrian J. Mulholland[2] and Antonia S.J.S. Mey[1]

[1]*EaStCHEM School of Chemistry, University of Edinburgh, Edinburgh, United Kingdom*
[2]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol*

Corresponding author: jasmin.guven@ed.ac.uk

Abstract

While elative binding free energy (RBFE) calculations are widely used in drug discovery, challenges remain for using them for metalloproteins. Here, we use meze, an open-source, metalloenzyme-parameterization program to parameterize two metal models for VIM-2, a metallo-β-lactamase (MBL), with two active site zinc ions.[1] The first model is a restraint-based approach, while the second is an upgraded Amber force field (UAFF), taken from literature.[2,3] Our molecular dynamics simulations with VIM-2 and KPC-2, a serine-β-lactamase (SBL) with a set of phosphonate-based inhibitors with cross-class affinity for SBLs and MBLs[4], give us insight into the key interactions relevant to β-lactamase inhibition.[1] We also evaluate the performance of RBFE methods for KPC-2, and the two metal models for VIM-2, with this inhibitor set. Our KPC-2 RBFE calculations achieve a Pearson's correlation coefficient of 0.93. For VIM-2, the UAFF model improves correlation from 0.55 to 0.78, compared to the restraint approach. We find that simple metal models can provide predictive free energy estimates but leave room for improvement in their modelling accuracy.

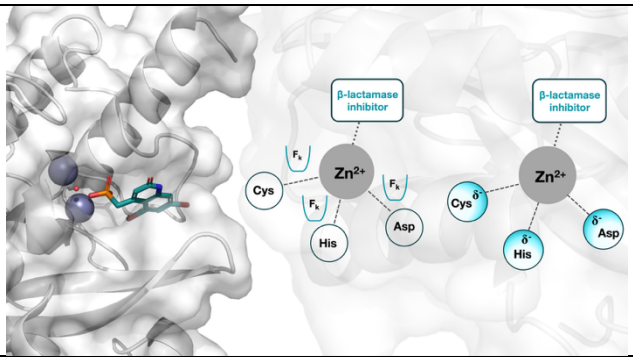

Figure 1: *Crystal structure image of VIM-2, a metallo-β-lactamase, and the two metal models.*

References

[1] Güven, J. Jasmin, Hanževački, Marko, Kalita, Papu, Mulholland Adrian J. and Mey, Antonia S.J.S. (2024). *Journal of Physical Chemistry B*, **128**(50), 12416–12424.
[2] Macchiagodena, Marina, Pagliai, Marco, Andreini, Claudia, Rosato, Antonio and Procacci, Piero, (2019), *Journal of Chemical Information and Modeling,* **59**(9), 3803-3816.
[3] Macchiagodena, Marina, Pagliai, Marco, Andreini, Claudia, Rosato, Antonio and Procacci, Piero, (2020), *ACS Omega*, **5**(25), 15301-15310.
[4] Pemberton, Orville A., Jaishankar, Priyadarshini, Akhtar, Afroza, Adams, Jessie L., Shaw, Lindsey N., Renslo, Adam R. and Chen, Yu, *Journal of Medicinal Chemistry*, (2019), **62**(18), 8480– 8496.

Enhanced Sampling Simulations of Biomolecular Systems

Edina Rosta [1]

[1]*University College London*

Corresponding author: e.rosta@ucl.ac.uk

Abstract

Important biological functions of biomolecular systems, such as catalysis, large-scale conformational changes, or ligand unbinding are often out of reach of atomistic molecular simulations. Enhanced sampling algorithms proved essential to mapping the pathway for these processes. However, the selection of collective variables is generally highly challenging.

Here, we present algorithms [1,2] that enable the identification of ligand unbinding pathways and pave the way for the calculation of unbinding free energy profiles. On one hand, we can determine **kinetically optimal bias** using **mean first passage times** (MFPTs) [2] to observe biomolecular pathways. On the other hand, without *a priori* defining CVs, we can identify useful atomic distances to map the unbinding pathway [1]. Subsequently, the finite-temperature string method is used to calculate the free energy barrier and obtain an estimate of the off rate. However, the key variables that determine the transition state (TS) are not necessarily obtained. To enable the identification of the most important CVs that play a major role at the TS, we developed a general machine learning (ML) approach, applicable to any molecular processes. We generate short downhill trajectories initiated near the TS and train ML networks to predict the trajectory outcomes. These trained networks are subsequently used to pinpoint the most relevant CVs at the TSs. Our calculations are applied to drugs targeting CDK2 [1] and the M3 muscarinic acetylcholine receptor [3]. We hope to provide key molecular features that help design inhibitors based on their allosteric and kinetic properties.

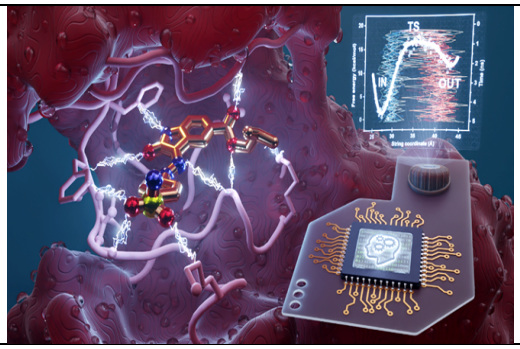

Figure 1: *Using machine learning to identify important collective variables for ligand unbinding.*

References

[1] M Badaoui, PJ Buigues, D Berta, G Mandana, H Gu, T Földes, C Dickson, V Hornak, M Kato, C Molteni, S Parsons, E Rosta, J. Chem. Theory Comput., **18**, 2543-2555 (2022)
[2] T Wei, B Dudas, E Rosta, "Kinetics-Optimized Enhanced Sampling Using Mean First Passage Times"; arXiv:2406.08964
[3] PJ Buigues, S Gehrke, M Badaoui, B Dudas, G Mandana, T Qi, G Bottegoni, E Rosta, "Investigating the Unbinding of Muscarinic Antagonists from the Muscarinic 3 Receptor", J. Chem. Theory Comput., **19**, 5260-5272, (2023)