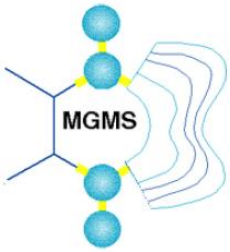


MGMS and RSC MMG
Young Modellers' Forum 2023
PROGRAMME & ABSTRACTS

Programme of Oral Presentations

09:15-09:50	Registration and Coffee
09:50-10:00	Introduction/Welcome
10:00-10:20	Modelling Solution Structures of Cyclic Peptides – How good are we? Daniel Crusius, University of Oxford
10:20-10:40	Understanding binding affinities of β-lactamase inhibitors: an alchemical study Jasmin Güven, University of Edinburgh
10:40-11:00	Accelerating gas-phase simulations of proteins with the fast multipole model Louise Persson, Uppsala University
11:00-11:30	Flash Poster presentations
11:30-11:50	Break
11:50-12:10	A Computational Investigation into the Cu-Catalysed Borylation of α,β-Unsaturated Compounds Jasmine Catlow, University of Sheffield
12:10-12:30	Towards in silico therapeutic T-cell receptor design: understanding and predicting cross-reactivity Annabelle Hartt, University of Bristol
12:30-12:50	Exploration of Intracellular Binding Pockets of CXC Chemokine Receptors for Allosteric Modulation Rhys Francis, University of Nottingham
12:50-14:10	Lunch and Posters
14:10-14:30	MDSubSampler: a Python library for a posteriori sampling of important protein conformations and data processing for machine learning prediction Namir Oues, Brunel University London
14:30-14:50	Enhancing Mixed Solvent Molecular Dynamics with Grand Canonical Nonequilibrium Candidate Monte Carlo William Poole, University of Southampton
14:50-15:10	A mesoscale model for the diffusion of nanoparticles in biological fluids: the challenge of the protein corona Beatrice Cipriani, Technological University of Dublin
15:10-15:30	Break
15:30-15:50	Engineering PET-degrading enzymes – targeting the energy barrier for PET binding Anna Jäckering, Heinrich-Heine University Düsseldorf
15:50-16:10	Accurate ranking (P-score) at flexible binding sites using SuMD and DA-QM-FMO Peter Ibrahim, University of Dundee
16:10-16:30	The Application of Machine Learning to Predict and Solve NMR Spectra in Structure Elucidation Ben Honoré, University of Bristol
16:30-16:45	Deliberation and Prizes
16:45	Close

Talk 1

Modelling Solution Structures of Cyclic Peptides – How good are we?

Daniel Crusius,[a] Jason R. Schnell,[a] Flaviu Cipcigan[b] and *Philip C. Biggin*[a]

[a] Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK
[b] IBM Research Europe, The Hartree Centre STFC Laboratory, Sci-Tech Daresbury, Warrington WA4 4AD, UK

Cyclic peptides (CPs) are macrocycles made from amino acids that can interact with classically 'undruggable' targets. Yet, their clinical utility remains hampered by the challenges in accurately predicting pharmacokinetic properties. A key issue lies in our limited understanding of CPs complex solution dynamics: CPs display a broad range of environment-dependent conformers, which can differ significantly from their bioactive structures. Although specialized simulation techniques have been developed to mimic available X-ray structures, the accuracy with which they capture the true ensemble of solution conformers remains largely unexamined.

To address this knowledge gap, we introduce "MacroConf," an expansive cyclic peptide solution NMR dataset curated from published studies. Employing molecular dynamics alongside specialized cheminformatics approaches, we model the solution structures of over 30 cyclic peptides. We conduct a thorough comparative analysis to: (a) assess the performance of different computational methodologies against experimental NMR structures, and (b) investigate whether less computationally intensive cheminformatics methods can approximate the accuracy achieved by state-of-the-art accelerated molecular dynamics simulations.

Through the implementation of automated computational workflows, our study serves not only as a rigorous assessment of current methodologies but also as a foundation for the development of future, more efficient conformer prediction tools for cyclic peptides. This work stands as a reference point, offering fully reproducible and extendable frameworks to guide development of future conformer generators for cyclic peptides.

Talk 2

Understanding binding affinities of β-lactamase inhibitors: an alchemical study

J. Jasmin Güven and *Antonia S.J.S. Mey*

EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Rd, Edinburgh

Alchemical free energy (AFE) calculations offer a computationally manageable way of obtaining free energies of binding between a protein and a small drug-like molecule. These calculations typically achieve accuracies of around 1 kcal/mol but ensuring reliability for some pharmaceutically relevant protein targets can still pose a challenge.[1] Examples of such targets are metallo-β-lactamases (MBLs), where active-site zinc ions directly interact with the ligands and thus influence their binding affinities. The production of β-lactamases by bacteria is the most common resistance mechanism against β-lactam antibiotics, such as penicillin.[2] Therefore, building reliable and easy-to-use AFE methods opens new opportunities for computational drug design for metalloprotein targets.

Here, we present an AFE workflow making use of the interoperable free energy tool BioSimSpace. The workflow is used to calculate relative binding free energies of a series of known inhibitors with cross-class affinity for serine- and metallo-β-lactamases. For KPC-2, a serine-β-lactamase, we obtain a mean unsigned error of $0.56^{0.57}_{0.55}$ kcal/mol and a Pearson's correlation coefficient of $0.836^{0.841}_{0.830}$. These results highlight the suitability and reliability of AFE calculations with serine-β-lactamases. In future, our workflow will be used to develop an AFE benchmark set of β-lactamases and their potential inhibitors which will aid in ensuring reliability of using these methods in computational drug design.

References:
[1] A. S. J. S. Mey, *et al*. LiveCoMS, 2020, **2**(1), 1-52.
[2] T. Palzkill, *Ann. N. Y. Acad. Sci.*, 2013, **65**(1), 91-104.

Talk 3

Accelerating gas-phase simulations of proteins with the fast multipole model

Louise J. Persson, *Erik G. Marklund*

Department of Chemistry – BMC, Uppsala university, Husargatan 3, 751 23 Uppsala, Sweden.

Molecular dynamics (MD) simulations find a powerful combination with mass spectrometry in the interrogation of proteins and protein complexes. While mass spectrometry is routinely used to probe multimeric proteins in the MDa range, MD simulations have been challenged by low computational performance due to a lack of efficient algorithms for computing electrostatic interactions in gas-phase systems. The particle mesh Ewald method, which is commonly used for condensed-phase simulations, is unsuitable for the gas phase, because it can only be used with periodic boundary conditions, and it is inaccurate for systems with a non-zero net charge. Thus, it has been necessary to calculate each pair-wise interaction explicitly, which is unfeasible for very large proteins due to the quadratic scaling of this approach.

A new algorithm for computing electrostatic interactions was recently adapted to the GROMACS MD package [1]. It is based on the fast multipole method (FMM), which partitions the system into octree boxes, and uses truncated multipole expansions to approximate the charge distribution of each box. Since it is compatible with open boundaries, treats systems with an overall charge accurately, and has been implemented with support for GPU acceleration already, it has the potential to accelerate simulations of large protein complexes in the gas phase.

In this work, we perform the first protein simulations using FMM, demonstrating significant performance enhancements for large proteins that even allow for MDa systems to be simulated over meaningful timescales. We provide guidelines for the method's use by evaluating how the performance and the accuracy of the FMM-computed Coulombic forces depend on the two FMM parameters *tree depth* and *multipole order*. Its use for both NVT and NVE ensembles is evaluated and discussed. Finally, we leverage the performance enhancement from FMM in applied simulations of a multimeric protein complex, showcasing the advantages of complementing ion mobility mass spectrometry with MD.

References:
[1] B. Kohnke, C. Kutzner and H. Grubmüller. *JCTC*, (2020) **16**, 6938-6949.

Talk 4

A Computational Investigation into the Cu-Catalysed Borylation of α,β-Unsaturated Compounds

Jasmine Catlow*, Ben Partridge, Grant Hill*

Department of Chemistry, The University of Sheffield, S3 7HF

Boronic esters are versatile building blocks for the synthesis of complex molecules, as they react under mild, functional group tolerant conditions. Boron reagents are also important as starting materials for one of the most common C-C bond-forming reactions in the pharmaceutical industry – the Suzuki reaction. Building a wider library of commercially available boronic ester building blocks would make adding functionality in organic synthesis more accessible. While there are currently many heteroaromatic boronic esters available, there are far fewer saturated heterocyclic boronic esters.[1]

The Partridge Group has developed a method for the synthesis of borylated lactams through the Cu-catalysed borylation of enoates, followed by their cyclisation.[1] These borylated lactams can be made in moderate-to-high yields, and the borylation step can be performed enantioselectively using a chiral catalyst. To aid in the design of future catalysts and new processes, and potentially tune the product selectivity, it would be advantageous to be able to understand how stereoselectivity is induced in these reactions.

This work develops a computational workflow to investigate the origins of stereoselectivity for the borylated lactam synthesis developed by the Partridge group. The main challenges of this work arise from modelling large organometallic complexes with two metal centres and chiral ligands. A simplified mechanism was modelled to determine the general mechanistic pathway, followed by benchmarking of different computational methods. The full reaction was then modelled, and semi-empirical methods such as xTB[2] and CREST were used to find low-energy conformers, before using DFT methods to obtain final geometries. This presentation will provide an overview of the computational benchmarking data, a detailed workflow, and a proposed mechanism with energy profile diagrams for the mechanisms modelled.

References:
[1] G. Rodgers, E. J. Wilson, C. C. Robertson, D. J. Cox, B. M. Partridge, *Advanced Synthesis and Catalysis,* 2021, **363**, 9, 2392–2395
[2] C. Bannwarth, S. Ehlert, and S. Grimme, *J. Chem. Theory Comput*., 2019, **15**, 3, 1652–1671

Talk 5

Towards in silico therapeutic T-cell receptor design: understanding and predicting cross-reactivity

Annabelle Hartt[a], *Stephen Harper[b]*, *Marc W. van der Kamp[a]*

[a] School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD
[b] Immunocore ltd., 92 Park Drive, Milton Park, Abingdon, Oxon, OX14 4RY

Corresponding author: ah16287@bristol.ac.uk

The potential of cancer immunotherapy using soluble T-cell receptors (TCRs) has increased drastically in the past few decades, with the recent approval of tebentafusp by the FDA demonstrating this[1]. Therapeutic TCRs require significant affinity enhancement for effective treatment, which often leads to unwanted cross-reactivity (binding to peptides that are not the intended target). Designing therapeutic TCRs is traditionally done through multiple rounds, combining experimental methods with human insights and structural data. Incorporating *in silico* design and evaluation could make this process faster and cheaper. Several studies have demonstrated the potential of *in silico* design workflows, but often come with the consequence of cross-reactivity[2]. Therefore, simulation-based evaluation of cross-reactivity should help reduce the failure rate of TCR drug candidates.

Here, we consider several TCRs that were engineered to bind the human leukocyte antigen (HLA) complexed with a cancer antigen peptide with high affinity. Surface plasmon resonance showed some TCRs to be cross-reactive towards off-target human peptides. Crystal structure data could not explain the increased cross-reactivity as all TCRs appeared highly similar. Molecular dynamics simulations of the TCR-pHLA complexes and the separate TCRs revealed that differences in specificity profiles were due to differences in 1) TCR-HLA interaction energetics, 2) flexibility of the CDR loops, or 3) both, depending on the given TCR. The fact that no one common explanation for cross-reactivity emerged demonstrates the complex nature of predicting it. We further show how molecular dynamics simulations (with off-target peptides modelled using Rosetta) can be used to predict whether a TCR is likely to be cross-reactive towards a given peptide. Overall, this research demonstrates how biomolecular simulation can contribute to therapeutic TCR development and improves our understanding of TCR specificity. In the future, this can result in more efficient generation of effective therapeutic TCRs, increasing the potential of drugs against cancer.

References:
[1] S. Dhillon, *Drugs*,(2022), **82**, 703–710.
[2] A. M. Rosenberg and B. M. Baker, *Current opinion in structural biology*, (2022), **74**, 102358.

Talk 6

Exploration of Intracellular Binding Pockets of CXC Chemokine Receptors for Allosteric Modulation

Rhys Francis[1], Shailesh Mistry[1], *Nicholas Holliday*[2], *Charles Laughton*[1]

[1]School of Pharmacy, University of Nottingham, Nottingham, NG7 2RD
[2]School of Life Sciences, University of Nottingham, Nottingham, NG7 2TQ
Corresponding author: rhys.francis@nottingham.ac.uk

The CXC chemokine receptor (CXCR)/CXC ligand (CXCL) axis plays a vital role in the cancer microenvironment and is attributed to proliferation, angiogenesis, invasion, and metastasis. Signalling of these receptors can be influenced by allosteric modulators binding to an intracellular site close to the G-protein binding region[1]. The most advanced class of compounds acting on this binding site are 3,4-diaminocyclo-3-ene-1,2-dione based antagonists[2]. The "best in class" of these negative allosteric modulators (NAM), navarixin, progressed to phase II trials, however undesirable side effects were observed. Navarixin exhibits 15-fold selectivity for CXCR2 over CXCR1 but inhibition of the CXCR2 pathway can lead to neutropenia suggesting that existing ligands should be retooled into selective CXCR1 NAMs. Currently, one barrier to further rational drug design is the lack of structural data.

The aim of this work is to develop novel CXCR1/2 intracellular, allosteric antagonists using a combination of in silico and in vitro methods. Homology modelling efforts yielded a CXCR1 model where docking of known ligands correlated with literature, allowing for structure-based design of CXCR1 NAMs. Through exploration of the intracellular binding site, key residues were identified to rationalise the relative antagonism between CXCR1 and CXCR2. A series of novel NAMs were synthesised and characterised in vitro to explore the antagonistic effect resulting from one of the key binding residues that differ between CXCR1 and CXCR2. Binding pocket and docking analysis of CXCR1 and CXCR2 revealed a potential extension to the binding region between TM3 and 5. Extension of the druggable binding pocket may play a key role in the design for a new class of CXCR1/2 antagonists. Fifteen 3,4-diaminocyclo-3-ene-1,2-dione based analogues were designed with the aim of extending the ligand and showed promising in silico characteristics. This series of compounds was subsequently synthesised and screened in-vitro against CXCR1 and CXCR2.

References:
[1] Salchow, K, Bond, M. (2010), *Br J Pharmacol,* **159**(7), 1429-1439
[2] Dwyer, Michael, Yu, Younong. (2014). *Expert Opin Ther Pat*, **24**(5), 519-534.

Talk 7

MDSubSampler: a Python library for a posteriori sampling of important protein conformations and data processing for machine learning prediction

Namir Oues[1], Sarath Chandra Dantu[1], Riktaben Jigarkumar Patel[1] and *Alessandro Pandini*[1]

[1]Department of Computer Science, Brunel University London, Uxbridge UB8 3PH

Introduction. Molecular dynamics (MD) simulations are routinely used to study protein dynamics and function. Advances in algorithms and hardware have made possible large-scale MD studies. However, there is a lack of tools for data-intensive tasks: extracting functionally relevant conformations from multiple trajectories and effectively using MD data to predict biological properties using machine learning.

Methods. We present MDSubSampler[1], a Python library for *a posteriori* subsampling of data from multiple trajectories and for the development of automated workflows for the streamlined processing of MD data for machine learning prediction. MDSubSampler is built on top of the MDAnalysis[2] and it includes three core classes: ProteinData to represent collections of trajectory subsamples, ProteinProperty to describe geometric properties calculated over trajectory frames and ProteinSampler to create automated processes to sample frames through different statistical methods. The toolkit implements automated workflows: to extract subsets of frames while preserving the original distribution of relevant properties and to subset and prepare trajectories to train machine learning models. MDSubSampler includes example scenarios as Python scripts and Jupyter Notebooks, a Unix-like command line interface and a set of reusable Python classes for ad-hoc customization and integration into existing workflows.

Results. Adenylate Kinase was selected as an example system. MDSubSampler was used to subsample a trajectory to <5% size while preserving the distribution of functional states. Additionally, a workflow was developed to train different supervised machine learning models to annotate frames according to two reference functional states.

References:
[1] N. Oues, S.C. Dantu, R.J. Patel and A. Pandini., *Bioinformatics* (2023), **39**, btad427
[2] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf and O.Beckstein, *J. Comput. Chem*., (2011), **32**, 2319-2327

Availability: https://github.com/alepandini/MDSubSampler

Talk 8

Enhancing Mixed Solvent Molecular Dynamics with Grand Canonical Nonequilibrium Candidate Monte Carlo

William G. Poole[1]; Victoria Nathan-Masier[1], Marley L. Samways[1], Davide Branduardi[2], Marcel L. Verdonk[2], *Jonathan W. Essex[1]*

[1] *School of Chemistry, University of Southampton, Southampton, SO17 1BJ*
[2] *Astex Pharmaceuticals, Cambridge, United Kingdom*
Corresponding author: wp1g16@soton.ac.uk

Mixed solvent molecular dynamics (MSMD) is a well-established simulation method for the identification of protein hotspots and ligand binding sites. There are many well established protocols for running MSMD simulations which often differ in the system setup or simulation details, but they are all united by the idea of simulating a protein target in a solution of water and a high concentration of organic probes. Designed to be an *in silico* alternative to the "Multiple Solvent Crystal Structures"[1] method which solves protein crystal structures in the presence of organic probes, MSMD aims to map regions on a protein surface with favourable interactions.

While many methods such as MixMD[2], MDMix[3] and SILCS[4] have had continued success at mapping protein surfaces and solvent exposed ligand binding sites, they often fail when it comes to more occluded/deeply buried binding sites as the diffusion time scales required for even the smallest probes is longer than what can be efficiently simulated. Furthermore, in simulation, some probes such as benzene aggregate at higher concentrations reducing the sampling at the protein interface.

For several years, Grand Canonical Monte Carlo (GCMC) based simulations have been used to efficiently sample water molecules within protein-ligand systems.[5,6] Combining GCMC insertions and deletions with conventional MD has been successful at accurately reproducing crystallographic water sites, sampling buried waters and predicting the effects of water on ligand binding.[7] In order to further improve sampling and increase acceptance rates we have combined the use of Nonequilibrium Candidate Monte Carlo with GCMC such that an insertion or deletion move happens over a short amount time with the molecule being slowly turned on or off. When applied to water, we have seen increased acceptance rates as well as a greater understanding of the binding process.[8] The incorporation of a "GCMC Sphere" means we focus our insertion and deletion moves to/from a region of interest, such as a protein, facilitating the sampling of buried pockets otherwise not accessible in traditional simulations.

More recently we have expanded the protocol to incorporate the insertion and deletions of small molecules [in preparation], with a focus on the MiniFrags library by Astex Pharmaceuticals.[9] Here, we apply this method to enhance typical MSMD simulations to predict small molecule binding sites for a range of protein systems with occluded binding pockets. The presentation will first give an overview of the current field, an introduction to Grand Canonical Nonequilibrium Candidate Monte Carlo and finally examples of the method being applied in the context of MSMD. Results will be compared to a traditional MD approach.

References:
[1] Mattos, C.; *et al* Journal of Molecular Biology 2006, 357 (5), 1471–1482. https://doi.org/10.1016/j.jmb.2006.01.039.
[2] Ghanakota, P.; *et al*. J. Med. Chem. 2016, 59 (23), 10383–10399. https://doi.org/10.1021/acs.jmedchem.6b00399.
[3] Alvarez-Garcia, D.; *et al.* https://doi.org/10.1021/ct500182z.
[4] Raman, E. P; *et al.* J. Chem. Inf. Model. 2011, 51 (4), 877–896. https://doi.org/10.1021/ci100462t.
[5] Samways, M. L.; *et al.* J. Chem. Inf. Model. 2020, 60 (10), 4436–4441. https://doi.org/10.1021/acs.jcim.0c00648.
[6] Woo, H.-J.; *et al.*. J. Chem. Phys. 2004, 121 (13), 6392–6400. https://doi.org/10.1063/1.1784436.
[7] Samways, M. L.; *et al.* J. Chem. Inf. Model. 2023, 63 (1), 387–396. https://doi.org/10.1021/acs.jcim.2c01225.
[8] Melling, O. J.; *et al.* J. Chem. Theory Comput. 2023, 19 (3), 1050–1062. https://doi.org/10.1021/acs.jctc.2c00823.
[9] O'Reilly, *et al.* Drug Discovery Today 2019, 24 (5), 1081–1086. https://doi.org/10.1016/j.drudis.2019.03.009.

Talk 9

A mesoscale model for the diffusion of nanoparticles in biological fluids: the challenge of the protein corona

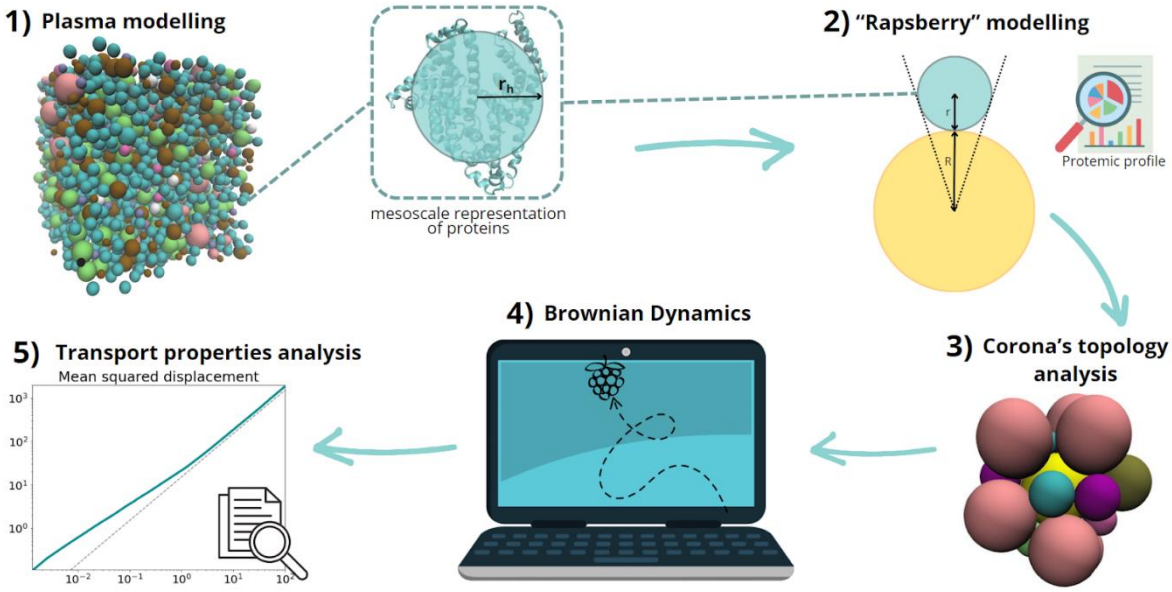
Beatrice Cipriani and *Hender Lopez Silva*

Technological University of Dublin, Physics & Clinical & Optometric Sciences
Beatrice.Cipriani@TUDublin.ie , Hender.LopezSilva@TUDublin.ie

The interaction of nanomaterials with biological media is an important factor that must be considered for every potential application in nanotechnologies: it is well known that in bio-fluids, Nanoparticles (NP) rapidly get coated by a thick protein layer called Protein Corona (PC) that highly affects the identity of the NP itself [1]. It has been observed the formation of a first, strongly bound layer (hard corona) whose protein can be irreversibly attached to the surface, and a second, weakly bound layer (soft corona), dynamically exchanging with the surrounding proteins. Although the dynamics of the soft corona's proteins is uncertain and its role on the nanoparticle's identity is still debated, it is well accepted that a major role on the NP's destiny *in vivo* is governed by the composition and thickness of the hard-corona.

In pharmaceutical applications, transport properties can have a dramatic effect on the permeability of therapeutic agents through biological barriers, and determining these parameters is urgently needed. Computational studies have already been performed in order to describe protein diffusion in crowded media, and valid meso-scale models have been developed: it has been demonstrated that in proteins' diffusion, volume exclusion (repulsive) interactions play the main role, followed by hydrodynamic interactions [2]. It is not clear if these two major factors would be reasonably describing the mobility of nanoparticles as well, or if alone they would result in a too broad approximation: an effective model for the analysis of diffusivity in nanomaterials has not been developed yet, and the studies performed so far focused mainly on the confinement and crowding effect of the dispersed medium, without considering in parallel the combined consequences of a protein aggregate such as the hard corona [3].

In this context, the aim of this work is to propose a computational model to simulate the dynamics of the NP in protein-rich media: to do that, both the polydispersity of the media and the interactions with the crowders (PC), must be considered. We propose a systematic workflow for the study of the role of the polydispersity of the medium and the selectivity of the NPs to certain crowders on the transport of proteins and NPs. For the analysis of such a large system at the relevant timescales, a full atomistic treatment would be computationally unaffordable: for this reason, a Coarse-Grained (CG) mesoscale representation of the proteins is carried out, and the hard-corona is treated as irreversibly absorbed to the NP's surface. With this study we present the effects of size, topology and heterogenity in the composition of the coronas for different NPs on their mobility in a crowded medium.



References:

[1] M. P. Monopoli, F. B. Bombelli, and K. A. Dawson, Nature Nanotechnology, vol. 6, no. 1, pp. 11–12, 2011.

[2] T. Ando and J. Skolnick, Proceedings of the National Academy of Sciences of the United States of America, vol. 107, no. 43, pp. 18 457–18 462, 2010..

[3] P. M. Blanco, M. Via, J. L. Garces, S. Madurga, and F. Mas, Entropy, vol. 19, no. 3, 2017

Talk 10

Engineering PET-degrading enzymes – targeting the energy barrier for PET binding

Anna Jäckering[a,b], *Birgit Strodel*[a,b]

[a]Department of Chemistry, Institute of Theoretical and Computational Chemistry, Heinrich-Heine University Düsseldorf, 40225 Düsseldorf, Germany

[b]Computational Biochemistry Group, Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Research Center Jülich, 52425 Jülich, Germany

In view of the worsening climate crisis and increasing plastic waste pollution, scientific interest in the development of an environmentally friendly enzymatic degradation mechanism for plastics is growing. However, the bottleneck in the industrial application of enzymes for plastic waste recycling is their insufficient activity and partial lack of stability under industrial conditions. [1]

To this end, we investigated the binding behaviour of highly active PET-degrading enzymes to polyethylene terephthalate (PET). Adsorption to the PET surface could be captured by classical molecular dynamics (MD) simulations. However, the entry of PET into the active site associated with the formation of productive binding poses was presumably hindered by an energy barrier limiting the activity of the enzyme. Using Hamiltonian Replica Exchange MD (HREMD) simulations, we were able to overcome this barrier and investigate entry pathways leading to productive conformations. In addition to hindering intramolecular PET interactions, we identified amino acids that potentially hinder entry into the binding site based on free energy surface profiles of amino acid-PET interaction. These residues serve as promising mutation sites to enhance PET degradation activity, which are currently investigated *in vitro*.

References:
[1] R. P. Magalhaes, J. M. Cunha and S. F. Sousa, *Int. J. Mol. Sci.*, 2021, **22**, 11257.

Talk 11

Accurate ranking (P-score) at flexible binding sites using SuMD and DA-QM-FMO

Peter E. G. F. Ibrahim, Fabio Zuccotto, Ulrich Zachariae, Ian Gilbert and *Mike Bodkin*

Drug Discovery Unit, Division of Biological Chemistry and Drug Discovery, University of Dundee, Dow St, Dundee DD1 5EH, UK

Introduction

Protein ligand binding prediction typically relies on docking methodologies and associated scoring functions to propose the binding mode of a ligand to a biological target. Significant challenges are associated with such approaches, including the flexibility of the protein-ligand system, solvent-mediated interactions, and associated entropy changes. In addition, scoring functions are weakly accurate due to the requirements of calculating enthalpic and entropic binding interactions quickly. The workflow described here attempts to address these limitations by combining Supervised Molecular Dynamics (SuMD) with Dynamical Averaging Quantum Mechanics Fragment Molecular Orbital (DA-QM-FMO).[1-2] This is illustrated using a set of five ligands targeting the SARS-CoV-2 Papain-like protease protein. We introduce a novel scoring function named the **P-score** that significantly enhances ranking accuracy.

Methods

The P-score ranks protein-ligand binding poses based on residence time (t) and interaction energy (ΔE) by evaluating the change in interaction energy over time. It provides a quantitative measure of complex stability.

The P-score prioritises a pose for each ligand in the different binding site conformations. The accuracy of the P-score predictions was assessed by comparing selected pose (highest P-score) for each ligand with its respective crystallographic reference.

Results and discussions

By ranking with the P-score we successfully identified the crystallographic structures of the five PL-pro inhibitors, independent from the starting position of the ligands or the receptor binding site conformation. This method represents a promising solution to overcome the challenges of exploring small molecule binding modes in flexible and dynamic solvated protein environments.

References:

[1] M. Bissaro, G. Bolcato, M. Pavan, D. Bassani, M. Sturlese, S. Moro, *ChemMedChem*., 2021, **16(13)**, 2075–2081.

[2] K. Takaba, C. Watanabe, A. Tokuhisa, R. Kanada, *Journal of Computational Chemistry.*, 2022, **43(20),**1362-1371.

Talk 12

The Application of Machine Learning to Predict and Solve NMR Spectra in Structure Elucidation

Ben Honoré, Calvin Yiu, *Craig Butts*

School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS

At Bristol we have developed IMPRESSION [1], a machine learning tool for predicting solution state chemical shifts and coupling constants of 3D molecular structures. The gold standard of IMPRESSION is a multitask graph transformer network that predicts ^{1}H and ^{13}C chemical shifts to within 0.1 and 1ppm, respectively, of quantum mechanical level (DFT) calculations. Predicted NMR parameters provide an important comparison when analysing NMR spectra and IMPRESSION takes a matter of seconds, rather than the hours/days that DFT can take to achieve the same result.

The inverse problem of generating the correct molecular structure from a set of NMR chemical shifts is much more challenging but would be ground-breaking in synthetic chemistry. Existing molecular generative models learn a distribution of chemical space, then sample new molecules from a region with favourable qualities, for example: ADMET properties. However, a set of NMR chemical shifts is likely too specific a property space for this method to be effective.

We have taken an approach that makes best use of IMPRESSION and our ability to rapidly predict chemical shifts. From a set of input fragments defined by the available spectral data, we generate and evolve a valency-valid structure, measuring the closeness of the IMPRESSION-predicted chemical shifts with the known true chemical shifts at each iteration. This error informs the next step in the evolution, creating a feedback loop. We then iterate until finding a structure with a suitably low IMPRESSION error to deem it a candidate.

This approach is effective for molecules with up to nine heavy atoms; however, as molecular size increases, the number of combinations to explore becomes exponentially larger and unfeasible to manage. Inclusion of more spectral data naturally narrows down this search space, but we are also looking into reinforcement learning as a technique to more intelligently and efficiently evolve the structure.

References:
[1] W. Gerrard, L. A. Bratholm, M. Packer, A. J. Mulholland, D. R. Glowacki and C. P. Butts, *Chem Sci*, (2020), **11**, 508-515.

Programme of Poster Presentations

1	**Deciphering the Mechanism of Avibactam Action Against Class A β-Lactamases through QM/MM Computational Assays** Jaida Begum, University of Bristol
2	MemPrO: Membrane Protein Orientation in Lipid Bilayers Matyas Parrag, University of Warwick
3	Characterisation of Persistence, Bioaccumulation and Toxicity of biologically active compounds with Machine Learning-based methods Dominga Evangelista, University of Bologna
4	Novel quantifiable metrics for the identification of non-systemic small drugs Miles Benardout, Kingston University London
5	Deciphering evolutionary improvements in designer enzymes by multiscale simulation to inform computational design Abbie Lear, University of Bristol
6	Combining Enhanced Sampling Methods – Investigating the Dual Binding Modes of a BRD4 Ligand Oliver J. Melling, University of Southampton
7	RNA translocation through an engineered nanopore Zephyr Ford, University of Oxford
8	Combining shape and electrostatics in a spectral geometry-based 3D molecular descriptor James Middleton, University of Sheffield
9	Electronic Polarization Tunes the Function of the Human Bestrophin 1 Cl^- Channel Linda Phan, University of Oxford
10	Dynamics based clustering of Class D β-lactamases using Molecular Simulations and Deep learning Fedaa Attana, UCL School of Pharmacy
11	Design and Synthesis of New Classes of Phosphatase Inhibitor Using Novel Computational Methods Emily Sampey, University of Manchester
12	Quantifying Ion Channel Pore Asymmetry: Exploring its influence on conductance and barrier to permeation David Seiferth, University of Oxford
13	Prediction of coiled coil oligomerisation using enhanced sampling Molecular Dynamics simulations Evangelia Notari, University of Edinburgh
14	Benchmarking structure-based 3D molecular generative models Benoit Baillif, University of Cambridge

Poster presenters will give their 2 minute "Lightning Talks" in the above order starting at 11:00 and will be available at the poster session to answer questions.

Poster 1

Deciphering the Mechanism of Avibactam Action Against Class A β-Lactamases through QM/MM Computational Assays

Jaida Begum[a], Adrian J. Mulholland[b], Jim Spencer[c], *Marc W. van der Kamp*[a]

[a]School of Biochemistry, University of Bristol, Bristol, BS8 1TD
[b]School of Chemistry, University of Bristol, Bristol, BS8 1TS
[c]School of Cellular and Molecular Medicine, University of Bristol, Bristol, BS8 1TD

Diazabicyclooctanes (DBOs) represent a crucial class of covalent inhibitors with clinical significance. Together with β-lactam antibiotics, they provide powerful combination therapies to treat complicated infections exemplified by avibactam-ceftazidime and relebactam-imipenem, and combat worldwide resistance driven by β-lactamase enzymes, which break down β-lactam antibiotics.

Unlike traditional covalent β-lactamase inhibitors which eventually get broken down, DBOs such as avibactam, remarkably regenerate upon leaving the β-lactamase. DBO inhibition efficiency is based on the balance between how quickly the drug binds (acylation) and is released (recyclization). Our aim is to generate an accurate and predictive computational protocol that can explain the molecular basis behind the experimentally observed differences between DBOs and different enzyme variants.

Here, we focused on the Class A β-lactamase, KPC-2, with avibactam. First, MM MD simulations were 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, confirmed 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 showed excellent agreement to the CI-NEB benchmarking studies and was used successfully to determine the minimum free energy pathway using two-dimensional umbrella sampling (2D US).

Now that we have established an accurate computational model of the rate-determining stage of avibactam acylation, 2D US will be used as a “computational assay” to predict the activity (i.e. free energy barriers) of different enzyme variants with various DBOs. Our simulations will help to dissect the reaction mechanism of DBOs on an atomistic level and rationalise experimental differences as well as suggestions for new and more efficient inhibitors or antibiotics to treat life-threating infections.

Poster 2

MemPrO: Membrane Protein Orientation in Lipid Bilayers

Matyas Parrag, *Phillip J Stansfeld*

Department of Life Sciences, University of Warwick, Coventry CV4 7AJ.

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, however these do not always work for some situations such as multi-bilayer systems or peripheral membrane proteins. This method enables the analysis of properties such as the width of intramembrane spaces, whilst also streamlining the setup of MD simulations for double membrane spanning proteins.

MemPrO can also be used to predict the approximate location of the peptidoglycan cell wall for periplasm spanning proteins in gram negative bacteria. The placement of the cell wall depends on many factors, some of which are external to the protein. This makes accurate predictions based only on the protein very difficult. It is however possible to provide information on potential locations.

The method consists of a minimisation in a mean field of potential constructed using Martini3[2] CG parameters. Combining this with an electrostatic potential results in a highly versatile approach for studying the bilayer insertion of membrane proteins. As minimisation methods can often get stuck in local minima, of which some may be non physical, many minimisations are performed and results assessed to give a selection of final orientation which are ranked by confidence.

References:
[1] Lomize, Andrei et al. (2009). *Protein Science,* **Volume 15** (Issue 6) , 1318-1333
[2] Souza, Paulo et al. (2021), *Nature Methods,***Volume 18**, 382-388

Poster 3

Characterisation of Persistence, Bioaccumulation and Toxicity of biologically active compounds with Machine Learning-based methods

Dominga Evangelista[a], Marinella Roberti[a], Bolognesi MariaLaura[a] and Giovanni Bottegoni.[b,c]

[a]Department of Pharmacy and Biotechnology, University of Bologna, Via Belmeloro 6, 40126, Bologna, Italy;
[b]Department of Biomolecular Sciences (DISB), University of Urbino Carlo Bo, Piazza Rinascimento 6, 61029, Urbino, Italy.
[cb]Institute of Clinical Sciences, University of Birmingham, Edgbaston, B15 2TT Birmingham (UK).

Assessing Persistence, Bioaccumulation, and Toxicity (PBT) is crucial for understanding the potential risk associated to the release of chemicals in the environment. Screening for PBT is still a challenging process due to the scarcity and poor quality of available data. This highlights the need for accurate and readily available methods to better predict the environmental impact of chemicals. PBT can be concisely expressed by the PBT index, defined thanks to a traditional QSPR model based on four structural molecular descriptors.[1] However, the accuracy of traditional QSP(A)R models tend to decrease when applied to large and heterogeneous datasets.

The aim of this project is to apply advanced machine learning (ML) algorithms to predict the PBT index of pharmaceutically relevant molecules, overcoming the limitations of more traditional approaches. Chemprop, a message passing neural network (MPNN) was employed for this purpose.[2] First, we assembled a comprehensive database of compounds already annotated by public agencies. After removing ambiguous and duplicated compounds, our final dataset encompassed 6386 molecules univocally labelled as PBT or non-PBT. Part of this collection (training set) was used to train a binary classification model that assigns each compound a PBT or non-PBT label based on features extracted from its structure. Once trained, the model was used to make predictions on the remaining molecules (test set). When applied to the test set, the figures of merit show how the model achieved an accuracy (g) of 0.947, and a recall rate (r) of 0.974. Optimizing hyperparameters, we only obtained a minor, if any, improvement in the results.

Overall, this study highlights the potential of ML algorithms for predicting PBT properties. We plan to further enhance the robustness and stability of our PBT prediction model by employing advanced bootstrap methodologies, thereby elucidating its capacity for generalization. Moreover, we are planning to apply the same framework to Drug-Induced Liver Injury (DILI) prediction, using multi-net classifiers to develop robust predictive models.

References:

[1] Papa, E. & Gramatica, P. QSPR as a support for the EU REACH regulation and rational design of environmentally safer chemicals: PBT identification from molecular structure. *Green Chem.* **12**, 836–84 (2010).
[2] Yang, K. *et al.* Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **59**, 3370–3388 (2019).

Poster 4

Novel quantifiable metrics for the identification of non-systemic small drugs

Miles Benardout[a], *Stephen P. Wren[a], Adam Le Greseley[a], Amr ELShaer[b]*

[a]Department of Chemical and Pharmaceutical Sciences, Faculty of Science, Engineering and Computing, Kingston University London, Penryhn Road, Surrey KT12EE, UK

[b]Department of Pharmacy, Faculty of Science, Engineering and Computing, Kingston University London, Penryhn Road, Surrey KT12EE, UK

Non-systemic drugs are compounds designed to have low oral bioavailability in order to target species located in the gut or small intestine, which standard drugs would struggle to target due to being absorbed into the small intestine. Typical drugs, designed with a high degree of oral bioavailability in mind, have relied upon "rule of thumb" druglikeness metrics, such as Lipinski's rule of five, to provide a framework in the design process of such drugs. Although some research has been undertaken in the quantification of druglikeness, such as Bickerton's QED [1], there has been very limited development of filters to identify small molecules with poor druglikeness, as Lipinski's rule of five fails to accurately identify small molecules (molecular weight ≤ 700) with poor oral biovailibity [2], let alone quantifiable metrics.

We have developed a metric for the identification of potentially non-systemic small molecules from publicly available datasets. A potential application for non-systemic small molecules is the development of non-systemic fructose scavengers for the treatment of fructose malabsorption, a potential cause of IBS and other gastrointestinal complaints. These molecules use a phenylboronic acid moiety to selectively bind to fructose. Performing a substructure search on PubChem for the phenylboronic acid moiety for compounds with a molecular weight of less than or equal to 700 produced a dataset of 5295 compounds for analysis.

The metric is designed to relatively compare compounds in a given dataset and quantify their normalised physiochemical (ADME) descriptors, evaluating them on the extent to which they deviate from hypothetical values as indicated by Lipinski's Rule of Five. From this relative analysis, we are able to trim down the dataset to find the sugar-binding compounds with the least druglike appearance. The trimming down of the dataset enables access to facile and fast semi-empirical analysis of sugar binding in order to evaluate a small library of compounds that we have identified and believe will both readily bind to fructose selectively and reside within the small intestine.

References:

[1] G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan and A. L. Hopkins, *Nat Chem*, 2012, **4**, 90.
[2] L. Z. Benet, C. M. Hosey, O. Ursu and T. I. Oprea, *Adv Drug Deliv Rev*, 2016, **101**, 89.

Poster 5

Deciphering evolutionary improvements in designer enzymes by multiscale simulation to inform computational design

Abbie Lear[1]; H. Adrian Bunzel[1,2] and *Adrian Mulholland*[1]

[1] Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK
[2] Department of Biosystems Science and Engineering, ETH Zurich, CH-4058 Basel, Switzerland
Corresponding author: adrian.mulholland@bristol.ac.uk

Reliable enzyme design would be a breakthrough for the chemical industry providing a wealth of specific, highly active biocatalysts for a huge range of syntheses. However, current enzyme design protocols only yield candidates with limited efficiency. Because rate-enhancing mutations are not usually predictable, random mutagenesis is used to improve activity through directed evolution. An example of a designed *de novo* enzyme is the Kemp eliminase 1A53-2, which catalyses proton abstraction from benzisoxazoles[1]. The limited activity of 1A53-2 was improved by 10^4-fold through directed evolution over five rounds of mutagenesis and screening. Unexpectedly, evolution also induced curved temperature dependence, signalling a negative activation heat capacity (ΔCp). Molecular dynamics simulations show that ΔCp in the evolved enzyme 1A53-2.5 stems from selective rigidification of the transition state ensemble,[2,3] but the relevance of ΔCp for efficient catalysis remains poorly understood.

Here, we aim to gain in-depth insights into ΔCp by combining quantum and molecular mechanics simulations to study the reaction barriers at a range of temperatures. Understanding how mutations introduced during evolution change the ligand-protein interactions in both the ground state and transition state, and the internal protein dynamics, will reveal the driving force behind the introduction of ΔCp alongside improved catalysis. A clearer picture of the origins of rate enhancement by directed evolution in 1A53-2.5 will inform algorithms for computational enzyme design.

References:

[1] H. A. Bunzel, H. Kries, L. Marchetti, C. Zeymer, P. R. E. Mittl, A. J. Mulholland and D. Hilvert, *J. Am. Chem. Soc.*, 2019, **141**, 11745–11748.
[2] H. A. Bunzel, J. L. R. Anderson, D. Hilvert, V. L. Arcus, M. W. van der Kamp and A. J. Mulholland, *Nat. Chem.*, 2021, **13**, 1017–1022.
[3] A. Lear, J. L. R. Anderson, D. Hilvert, V. L. Arcus, M. W. van der Kamp, H. A. Bunzel and A. J. Mulholland, *ACS Catal.*, 2023, **13**, 10527–10530.

Poster 6

Combining Enhanced Sampling Methods – Investigating the Dual Binding Modes of a BRD4 Ligand

Oliver J. Melling[a], Chun-wa Chung[b], Ian D. Wall[b], *Jonathan W. Essex[a]*

[a]School of Chemistry, University of Southampton, Southampton, SO17 1BJ.
[b]GSK, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK.

Despite their widespread application, free energy calculations can struggle to produce converged results, often due to the relatively short simulation length at each step along the relevant pathway. As a result, slow orthogonal degrees of freedom in the system, such as the rearrangement of a water network, or the sampling about a rotamer, cannot be sufficiently sampled and their effect on the resulting free energy is therefore incompletely resolved.

Many enhanced sampling methods have been developed to increase the sampling of states that are otherwise separated by high kinetic barriers. These methods have had success in reducing the computational time required to produce accurate and converged configurational ensembles. In addition, enhanced sampling methods have more recently been incorporated into free energy calculations and have shown promise in improving the accuracy of results when compared to experimental data. However, there are few cases where multiple enhanced sampling methods have been combined to tackle slow independent molecular rearrangements, in combination with free energy calculations.

Here, we present a bromodomain-containing protein (BRD4), bound to a ligand that has experimentally been shown to have two binding modes. Each binding mode is accompanied by a different bound water network and different side chain conformations of nearby protein residues. The rearrangement of the water network and sampling of the residue sidechains are both examples of slow orthogonal modes which are inadequately sampled in conventional MD simulations. The system is therefore an ideal test case for combined enhanced sampling methods.

We present results of standard MD on the two binding modes, enhanced water sampling (using grand canonical nonequilibrium candidate monte carlo[1]) on the bound water networks in each binding pose and compare the distributions of sidechain dihedral angles for a relevant binding site residue using both MD and Fully Adaptive Simulated Tempering (FAST)[2]. We finally propose a free energy cycle to calculate the relative stability of the two poses and discuss how this calculation could incorporate the enhanced sampling methods detailed above.

References:
[1] O.J. Melling, M.L. Samways, Y. Ge, D.L. Mobley and J.W. Essex, *JCTC,* (2023), **19**, 1050-1062
[2] M. Suruzhon, M.S. Bodnarchuk, A. Ciancetta, I.D. Wall and J.W. Essex, *JCTC*, (2022), **18**, 3894-3910

Poster 7

RNA translocation through an engineered nanopore

Zephyr Ford, *Syma Khalid*
Department of Biochemistry, University of Oxford, OX1 3QU

We propose the use of the protein FhuA, in a form previously engineered and experimentally tested[1], as a nanopore for the controlled translocation of double-stranded RNA. Using molecular dynamics, we characterise ion flux through the pore at a range of applied electric field strengths. We probe the translocation mechanism of the movement of the small dsRNA fragment.

We find that the RNA interacts for longer periods of time with specific charged residues in the pore lumen, and can be translocated in either direction through the pore based on the direction of the electric field.

In silico mutation of charged pore-lining residues allows the RNA to translocate more swiftly through the pore by removing these interaction or dwell points. Finally, we investigate the impact of lipid-RNA interactions on the speed of RNA translocation, including the role of the change in RNA conformation.

The results of these simulations can be compared to those of single-stranded DNA through smaller nanopores[2], showing more general principles of interactions that may influence the future design of nanopores for biotechnology, expanding beyond their use simply for DNA.

References:
[1] Niedzwiecki, D. J.; Mohammad, M. M.; Movileanu, L. Inspection of the Engineered FhuA ΔC/Δ4L Protein Nanopore by Polymer Exclusion. *Biophys J* (2012), **103 (10)**, 2115–2124. https://doi.org/10.1016/j.bpj.2012.10.008.
[2] Rattu, P.; Belzunces, B.; Haynes, T.; Skylaris, C.-K.; Khalid, S. Translocation of Flexible and Tensioned ssDNA through in Silico Designed Hydrophobic Nanopores with Two Constrictions. *Nanoscale* (2021), **13 (3)**, 1673–1679. https://doi.org/10.1039/D0NR04890A.

Poster 8

Combining shape and electrostatics in a spectral geometry-based 3D molecular descriptor

James Middleton[1], Gian Marco Ghiandoni[2], Martin J. Packer[3], Mengdie Zhuang[1], *Valerie J. Gillet[1]*

[1]Information School, University of Sheffield, The Wave, 2 Whitham Rd, Sheffield S10 2AH, United Kingdom
[2]AstraZeneca R&D IT, Academy House, 136 Hills Road, Cambridge CB2 8PA
[3]AstraZeneca Early Oncology R&D, Alderley Park, Macclesfield, SK10 4TG

It has been well established that shape complementarity plays an important role in molecular recognition. However, shape information alone does not suffice for describing accurately the binding process between a drug molecule and a therapeutic target, which also requires accounting for their electrostatic complementarity (Rathi et al., 2020).

In our work, the MolSG spectral geometry-based molecular descriptor workflow developed by Seddon et al. (2019) has been modified to include electrostatic information whilst retaining the beneficial alignment-invariance property of the original shape descriptor. Electrostatic potential (ESP) surfaces are computed using the Graph Convolutional Network developed by Rathi et al. (2020) which has been shown to produce potentials of comparable quality to computationally expensive density-functional theory (DFT) derived ESP surfaces whilst taking a fraction of the time (Rathi et al., 2020).

The electrostatic information is represented using alignment-invariant 1D histograms which take inspiration from the colour histograms popularized in content-based image retrieval. The ESP energy information is first projected onto a spherical surface which is subsequently separated into distance-based bins with respect to an intrinsic surface centroid. This centroid has been approximated using gradient descent with momentum alongside an implementation of the Riemannian center of mass as a cost function. The application benefit of using an intrinsic centroid as opposed to an extrinsic centroid based on the surface volume, means that the alignment-invariance property of the original MolSG descriptor is retained. 1D ESP energy histograms are then generated for each distance-bin to augment the MolSG 3D shape information with spatially enriched electrostatic information, in the program called Electro-MolSG.

Electro-MolSG has been benchmarked against a series of established descriptors in a ligand-based virtual screening setting using the DUD-E dataset. Initial findings suggest that the Electro-MolSG descriptor can outperform the shape-only implementation of MolSG, leading to enhanced enrichment in ligand-based virtual screening applications.

References:

[1] Prakash Chandra Rathi, R. Frederick Ludlow, and Marcel L. Verdonk, *Journal of Medicinal Chemistry* **2020** *63* (16), 8778-8790, DOI: 10.1021/acs.jmedchem.9b01129
[2] Matthew P. Seddon, David A. Cosgrove, Martin J. Packer, and Valerie J. Gillet, *Journal of Chemical Information and Modeling* **2019** *59* (1), 98-116, DOI: 10.1021/acs.jcim.8b00676

Poster 9

Electronic Polarization Tunes the Function of the Human Bestrophin 1 Cl^- Channel

Linda X. Phan [a,b], Aaron P. Owji [c,d,e], Tingting Yang [c], Jason Crain [f], *Mark S. P. Sansom [b], Stephen J. Tucker [a,g]*

[a] Clarendon Laboratory, Department of Physics, University of Oxford, Oxford, OX1 3PU, UK
[b] Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK
[c] Department of Ophthalmology, Columbia University, New York, NY, USA
[d] Department of Pharmacology, Columbia University, New York, NY, USA
[e] Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY, USA
[f] IBM Research Europe, Hartree Center, Daresbury, WA4 4AD, UK
[g] Kavli Institute for Nanoscience Discovery, University of Oxford, OX1 3QU, UK

Mechanisms of anion permeation within ion channels and nanopores remain poorly understood. Recent cryo-electron microscopy structures of the human bestrophin 1 chloride channel (hBest1) [1] provide an opportunity to evaluate ion interactions predicted by molecular dynamics (MD) simulations against experimental observations. We implement the fully polarizable AMOEBA forcefield in MD simulations of open and partially-open states of the hBest1. The AMOEBA forcefield models multipole moments up to the quadrupole; therefore, it captures induced dipole and anion-π interactions. By including polarization we demonstrate the key role that aromatic residues play in ion permeation and the functional advantages of pore asymmetry within the highly conserved hydrophobic neck of the pore. We establish that these only arise when electronic polarization is included in the molecular models. We also show that Cl^- permeation in this region can be achieved through hydrophobic solvation concomitant with partial ion dehydration, which is compensated for by the formation of contacts with the edge of the phenylalanine ring. Furthermore, we demonstrate how polarizable simulations can help determine the identity of ion-like densities within high-resolution cryo-EM structures. Crucially, neglecting polarization in simulation of these systems results in the localization of Cl^- at positions that do not correspond with their experimentally resolved location. Overall, our results demonstrate the importance of including electronic polarization in realistic and physically accurate models of biological systems.

References:
[1] A. Owji, J. Wang, A. Kitteridge, Z. Clark, Y. Zhang, W. Hendrickson, T. Yang. *Nat Comms*, (2022), **13**, 1-11.

Poster 10

Dynamics based clustering of Class D β-lactamases using Molecular Simulations and Deep learning

Fedaa Attana, *Prof. Shozeb Haider*

UCL School of Pharmacy, 29-39 Brunswick Square, London, WC1N 1AX

Antimicrobial resistance is currently identified as a global public health crisis with projections of 10 million deaths yearly by 2050 in addition to expected increased risks of invasive therapeutic interventions. This project focuses on class D β lactamases which have received much attention recently due to their rapid rate of mutations and their resistance to 'last resort' antibiotics such as carbapenems. Class D β-lactamases, also known as oxacillinases (OXAs), are known to have remarkably diverse sequences yet they exhibit structural similarity and previous work in the group showed conserved dynamics in other β-lactamase classes [1] thus we aim to harness the power of machine learning to build a model that can cluster the currently known OXA structures and eventually predict properties of emerging variants.

We defined the enzymes architecture and the conserved motifs via structural and sequence alignment of all non-redundant, experimentally obtained OXA crystal structures (32 in total). To explore the dynamics of the enzymes, adaptive bandit sampling molecular dynamics (MD) simulations are conducted for all OXAs for at least 20 μsec each. This enhanced sampling method uses machine learning algorithms to make the best exploration/exploitation decisions along the MD trajectory recording and thus use the simulation time in the most efficient way. It has been established that the flexibility in the loops framing the active site is the main determinant to the specificity of different OXA enzymes[2], especially the loop linking the B5 and B6 strands. Using Unsupervised deep learning approach, CVAE: Convolutional Variational AutoEncoder, a model that can capture the dynamics of these loops was able to produce 2D projections of millions of MD obtained conformations of the studied OXAs clustered in functionally relevant pattern.

This innovative approach provides valuable insights into the behaviour of OXAs and their dynamics. Furthermore, it presents a promising avenue to explore other protein families, notably in cases where experimental function evaluation for emerging mutants is challenging, highlighting the significant role that robust computational predictions can play.

References:

[1] Galdadas, I.; Lovera, S.; Pérez-Hernández, G.; Barnes, M. D.; Healy, J.; Afsharikho, H.; Woodford, N.; Bonomo, R. A.; Gervasio, F. L.; Haider, S. Defining the Architecture of KPC-2 Carbapenemase: Identifying Allosteric Networks to Fight Antibiotics Resistance. *Sci. Rep.* 2018, ***8*** (1), 12916. https://doi.org/10.1038/s41598-018-31176-0.

[2] Evans, B. A.; Amyes, S. G. B. OXA β-Lactamases. *Clin. Microbiol. Rev.* 2014, ***27*** (2), 241–263. https://doi.org/10.1128/CMR.00117-13.

Poster 11

Design and Synthesis of New Classes of Phosphatase Inhibitor Using Novel Computational Methods

Emily Sampey[a], *Andrew Leach[a], Sam Butterworth[a], Lydia Tabernero[b]*

[a]Division of Pharmacy and Optometry, University of Manchester, Oxford Road, Manchester, M13 9PT.
[b]Department of Evolution, Infection and Genomics, University of Manchester, Oxford Road, Manchester, M13 9PT.

The regulation of substrate phosphorylation levels within cells, by the opposing actions of phosphatase and kinase enzymes, is a process of huge therapeutic potential which is well established by the dozens of FDA-approved small molecule kinase inhibitors. In contrast to this, there is yet to be a single protein tyrosine phosphatase (PTP) inhibitor to enter the market owing to the "undruggable" nature of PTPs.[1] This work employs two computational approaches in an attempt to design potential inhibitor compounds of the PTPs. The first approach is the commonly used virtual screening, while the second is the more novel design of a "theoceptor model".[2] With regard to the latter approach, the methodology is discussed and will be employed to PTP allosteric sites following the validation of the method within the Leach group. For the virtual screening approach, the chosen therapeutic target is MptpA. MptpA is a low molecular weight PTP secreted by *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB). Targeting virulence factors such as MptpA may offer a new method of treatment for TB which overcomes the urgent issue of antibacterial resistance. Using Autodock Vina, a small commercially available fragment library was screening against the 21 available MptpA structures from the PDB. The resulting 'hits' were hybridised to create novel fragments, which were again docked into the 21 MptpA structures. The top fragments were identified, synthesised and subjected to biochemical assays, leading to the identification of a single, low-potency inhibitor (**25**, IC_{50} = 287 ⌈M). A small SAR study around this fragment led to the identification of a compound of improved potency (**39**, IC_{50} = 7.7 ⌈M), which is proposed to act through competitive inhibition. Future work on this series of compounds will include obtaining an x-ray co-crystal structure of the bound ligand, to aid further compound design, and investigation into the selectivity of the ligand.

References:
[1] M. Köhn, *ACS Cent. Sci.*, (2020) **6**, 476-477.
[2] I. Lukac, H. Abdelhakim, A. R. Ward, S. A. St-Gallay, J. C. Madden, A. G. Leach, *A. Chem. Sci*, (2019) **10**, 2218-2227.

Poster 12

Quantifying Ion Channel Pore Asymmetry: Exploring its influence on conductance and barrier to permeation

David Seiferth, *and Philip C. Biggin*

Structural Bioinformatics and Computational Biochemistry, Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK

Recent advances in structural biology have led to a growing number of ion channel structures featuring heteromeric subunit assembly, exemplified by synaptic Glycine receptors (GlyRs) and α4β2 nicotinic receptors. These structures exhibit inherent pore asymmetry, which has raised questions about the role of asymmetry in ion channel function. Furthermore, molecular dynamics simulations performed on symmetrical homomeric channels often lead to thermal distortion that means conformations of the resulting ensemble are also asymmetrical. When functionally annotating ion channels, researchers often rely on minimal constrictions determined through radius-profile calculations such as HOLE [1] or CHAP [2], coupled with an assessment of pore hydrophobicity. However, such tools typically employ spherical probe particles, limiting their ability to accurately capture pore asymmetry. In this study, we introduce an algorithm that employs ellipsoidal probe particles, enabling a more comprehensive characterization of pore asymmetries. A constriction is more asymmetric for a larger difference between the smaller and larger radius of the ellipsoidal probe particle. Our analysis reveals that the use of non-spherical ellipsoids for pore characterization, taking into account both a smaller and a larger radius, provides a more accurate and easily interpretable depiction of conductance, shedding light on the key structural properties governing ion conductance. To quantify the implications of pore asymmetry, we investigate Carbon Nanotubes (CNTs) with varying degrees of pore asymmetry as model systems. The conductance through these channels show surprising effects that would otherwise not be predicted with spherical probes. The results have broad implications not only for the functional annotation of biological ion channels, but also for the design of synthetic channel systems for use in areas such as water filtration.

References:
[1] Smart, O.S., Neduvelil, J.G., Wang, X., Wallace, B.A., Sansom, M.S.P., 1996. J. Mol. Graph. **14**, 354–360.
[2] Klesse, G., Rao, S., Sansom, M.S.P., Tucker, S.J., 2019. J. Mol. Bio. **431**, 3353–3365.

Poster 13

Prediction of coiled coil oligomerisation using enhanced sampling Molecular Dynamics simulations

Evangelia Notari[a], *Christopher W. Wood*[b], *Julien Michel*[a]

[a]EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, UK
[b]School of Biological Sciences, University of Edinburgh, C.H. Waddington Building, Max Born Crescent, Edinburgh EH9 3BF, UK

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 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 conformational preferences of coiled coils. To this end, we couple parametric design and metadynamics simulations with a novel 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), as well as rank designs with different sequences, in agreement with experimental observations. Preliminary results from the studies of pH-switchable systems will also be discussed.

References:
[1] G.G. Rhys, C.W. Wood, J.L. Beesley, N.R. Zaccai, A.J. Burton, R.L. Brady, A.R. Thomson and D.N. Woolfson, *JACS*, (2019) **141**(22), 8787-8797.

Poster 14

Benchmarking structure-based 3D molecular generative models

Benoit Baillif[a], *Jason Cole[b], Patrick McCabe[b], Andreas Bender[a]*

[a] Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Rd, CB2 1EW, Cambridge, United Kingdom

[b] Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom.

Deep learning models are widely used to generate diverse molecules or for goal-directed molecular optimisation. Recently, a large number of latent space and autoregressive models have been developed to produce three-dimensional (3D) representations of molecules [1]. The main advantage of 3D generative models is the generation of molecules optimised on 3D-dependent properties, such as potentially bioactive molecule generation within a binding pocket. The performance of generators producing SMILES or graphs are compared using benchmarks such as GuacaMol or MOSES, but these are not suited to 3D generators, where the quality of generated molecular conformations must also be assessed. In this work, we hence developed a new Python package which implements a new benchmark for models producing molecules within a binding pocket, that evaluates the validity of 3D conformations based on their bond lengths and valence angles compared to value ranges observed in the high-quality LigBoundConf dataset, and using MMFF94s strain energy; and compares the bioactivity relative to test ligands using the Vina score, PLP score and interaction fingerprints. The LiGAN, 3D-SBDD, Pocket2Mol and TargetDiff models were benchmarked. Between 86% and 100% of generated molecules show at least one invalid bond length or angle, and significantly higher MMFF94s strain energy than training ligands. These physically unrealistic conformations still resulted in poses with better Vina score than reference ligands ranging from 12% of cases for Pocket2Mol, up to 20% of cases for LiGAN despite being the model producing the most strained molecules, indicating that the Vina score does not fully account for the physical quality of ligand conformation. Using other scoring functions that give higher importance to ligand strain (e.g. PLP score) alleviates this caveat. Therefore, generated molecules in their raw poses should not be directly used for bioactivity scoring.

References:

[1] Baillif, B.; Cole, J.; McCabe, P.; Bender, A. Deep Generative Models for 3D Molecular Structure. *Curr. Opin. Struct. Biol.* **2023**, *80*, 102566. https://doi.org/10.1016/j.sbi.2023.102566.

Useful Information

Travel to Lady Margaret Hall

https://www.lmh.ox.ac.uk/sites/default/files/documents/2017-01/LMH%20Travel%20Information.pdf

YMF will be located in the Simpkins Lee Theatre and Monson Room