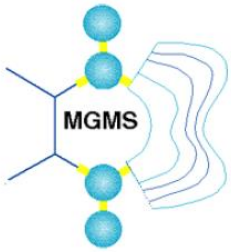

MGMS and RSC MMG
Young Modellers' Forum 2022
PROGRAMME & ABSTRACTS

Programme of Oral Presentations

09:15-09:50	Registration and Coffee
09:50-10:00	Introduction/Welcome
10:00-10:20	"Supramolecular organisation and dynamics of Mannosylated Phosphatidylinositol Lipids in the Mycobacterial Plasma Membrane" Chelsea Brown, Warwick University
10:20-10:40	"A World of Probabilities: An sMD/MSM Approach for Rational Design of Allosteric Modulators" Adele Hardie, Edinburgh University
10:40-11:00	"Searching membrane-facing allosteric sites of GPCRs: Neurotensin receptor 1" Abdul-Akim Guseinov, Queen's University Belfast
11:00-11:30	Flash Poster presentations
11:30-11:50	Break
11:50-12:10	"How sticky are our proteins? Quantifying hydrophobicity of the human proteome" Dea Gogishvili, VU Amsterdam
12:10-12:30	"Using WRITHE to produce realistic protein structure predictions from BioSAXS Data" Arron Bale, Durham University
12:30-12:50	"Sample efficient structure-based de novo drug design in the hunt for novel GPCR antagonists with Augmented Hill-Climb" Morgan Thomas, Cambridge University
12:50-14:10	Lunch and Posters
14:10-14:30	"QMMM study of the Aromatic hydroxylation by Cinnamate 4-hydroxylase (C4H) a lignin regulator" Sonia Santos, Queen's University Belfast & Northumbria University
14:30-14:50	"Molecular Dynamics of Cholesterol Membranes: A comparative Forcefield study" Jack Sawdon, Southampton University
14:50-15:10	"Molecular Dynamics Simulations of Engine Lubricant Additives" Angie Maťusová, Edinburgh University
15:10-15:30	Break
15:30-15:50	"Characterization of GLP-1R Ala316Thr variant. Does genetic variability in GLP-1R have an impact on the management of type 2 diabetes?" Roxana-Maria Rujan, Coventry University
15:50-16:10	"Complexity Matters: Polymyxin Antibiotics within the *E. coli* Cell Envelope" Iain Smith, Southampton University
16:10-16:30	"Can You Hear the Shape of a Drug?" Rachael Pirie, Newcastle University
16:30-16:45	Deliberation and Prizes
16:45	Close

Talk 1

Supramolecular organisation and dynamics of Mannosylated Phosphatidylinositol Lipids in the Mycobacterial Plasma Membrane

Chelsea M. Brown[a], Robin A. Corey[b], Ya Gao[c], Yeol Kyo Choi[d], Martine Gilleron[e], Nicolas Destainville[f], *Elizabeth Fullam*[a], Wonpil Im[c], *Phillip J. Stansfeld*[a,g], Matthieu Chavent[f]

[a]*School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK*
[b]*Department of Biochemistry, University of Oxford, Oxford, UK*
[c]*School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China*
[d]*Department of Biological Sciences, Lehigh University, Pennsylvania 18015, USA*
[e]*Institut de Pharmacologie et Biologie Structurale, CNRS, Université de Toulouse, 205 route de Narbonne, 31400, Toulouse, France*
[f]*Laboratoire de Physique Théorique, CNRS, Université Paul Sabatier, 31062 Toulouse,France*
[g]*Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK*

Mycobacterium tuberculosis (*Mtb*) is the causative agent of tuberculosis (TB), a disease that claims ~1.5 million lives annually. Drug susceptible infections are treated with a 6-moth regime which is riddled with serious side-effects. Drug resistance is unsurprisingly on the rise – an issue that is a global health concern. It is clear we need new methods to treat TB, but this pathogen is incredibly difficult to study experimentally and has physical barriers to prevent small molecule entry.

The cell envelope of *Mtb* contains glycans and *Mtb* specific lipids, separated into four discrete layers. The most well studied is the outer membrane which includes mycolic acids but less is known about the inner layers – especially the plasma membrane. This has been proposed to have a unique composition, with over 50% of the lipids being mannosylated phosphatidylinositol lipids (phosphatidyl-*myo*inositol mannosides, PIMs)[1]. How PIMs contribute to the structure and function of the *Mtb* plasma membrane is yet to be fully elucidated.

In this work we used multiscale molecular dynamics (MD) simulations to examine the structure-function relationship of the PIM lipids found in mycobacteria. The model was first validated using comparison of atomistic to coarse grained simulations. Then, a 50 x 50 nm bilayer was shown to be stable. Membrane proteins and antibiotics known to interact with them were also tested, and experimental results were replicated in these simulations.

Our results provide a pipeline to assemble a mycobacterial membrane *in silico* to study how biophysical properties change as lipid composition changes, and hence how that might influence the passage of antibiotics into the cell.

References:
[1] R. Bansal-Mutalik and H. Nikaido, *Proceedings of the National Academy of Sciences*, 2014, **111**, 4958-4963.

Talk 2

A World of Probabilities: An sMD/MSM Approach for Rational Design of Allosteric Modulators

Adele Hardie,[a] Silvia Lovera,[b] *Julien Michel*[a]

[a]*EaStChem School of Chemistry, Joseph Black Building, University of Edinburgh, Edinburgh, EH9 3FJ, UK*
[b]*UCB Pharma, Chemin du Foriest 1, 1420 Braine-l'Alleud, Belgium*

Multiple computational methods are available to predict ligand binding affinity and allosteric sites, but currently no robust protocol exists to assess the allosteric potential of a known binder. We propose a novel method that combines Markov State Models (MSMs) with steered molecular dynamics (sMD) to achieve this objective. Steered MD simulations are employed to sample protein conformational space inaccessible to routine equilibrium MD timescales. Protein conformations sampled by sMD provide starting points for "seeded" molecular dynamics simulations, which are used to build MSMs. State probabilities computed from these MSMs may be used to predict the effect of ligand binding on protein function.

The methodology is demonstrated on a dataset of protein tyrosine phosphatase 1B (PTP1B) ligands. Experimentally confirmed allosteric inhibitors with a wide range of activities[1,2] are put through the workflow to assess whether they allosterically regulate PTP1B function. The findings from our study confirm this methodology can be used for progressing hits towards lead molecules in drug design campaigns.

References:
[1] C. Wiesmann, *et al.*, *Nat. Struct. Mol. Biol.,* (2004), **11**, 730-737.
[2] D. A. Keedy, *et al.*, *eLife,* (2018).

Talk 3

Searching membrane-facing allosteric sites of GPCRs: Neurotensin receptor 1

Abdul-Akim Guseinov[1]; Dmitry S. Karlov[1]; Stephen Garland[2]; Bianca Plouffe[1]; Irina G Tikhonova[1]

1 Queen's University Belfast, 97 Lisburn Road, Belfast, BT9 7BL, UK
2 *NQuiX Ltd, Causeway House, Dane Street, Bishop's Stortford, CM23 3BT, UK*

The search for druggable allosteric sites is a hot topic in drug discovery. Molecular dynamics (MD)-based approaches are extensively used for this purpose. Among them, probe MD simulations demonstrate rapid development. In these methods, the target is simulated in the presence of probe fragments, which are a prototype of drug molecules. However, there are a number of challenges to use the probe simulations for membrane-embedded allosteric sites, including slow probe diffusion in membrane, probe aggregation and accumulation in the membrane.

Recently, our laboratory developed a probe-confined dynamic mapping protocol, which enhanced probe sampling within an area of interest. [1] This protocol was successfully applied to map allosteric sites of GPCRs at various locations. In this work, we apply it to neurotensin receptor 1 (NTSR1), which isa target for drug discovery against a wide range of cancers and several neurological diseases. To date no allosteric site has been reported for this receptor, therefore, NTSR1 could be a good blind system to test our protocol.

First, we identified 5 membrane-facing allosteric sites using MDpocket, which is a geometric algorithm for cavity detection in MD trajectories. We next conducted probe-confined dynamic mapping focusing on these sites. The probe simulations were further explored by MD simulations of a docked probe molecule. The probe simulations allowed us to estimate the accessibility of selected allosteric sites and outline the key probe-receptor interactions.

To study the importance of these sites for receptor signalling, we mutated several residues forming each site. We monitored the signalling of the resulting NTSR1 mutants using bioluminescence resonance energy transfer (BRET)-based biosensors in HEK293 cells. We identified three non-biased and one biased allosteric sites.

From the druggability assessment in the simulations and the NTSR1 signalling pattern in the functional assays, we selected 2 allosteric sites showing promise as targets for ligand design.

References

[1] Ciancetta A, Gill AK, Ding T, Karlov DS, Chalhoub G, McCormick PJ, Tikhonova IG (2021) *ACS Central Science*, **7**(11), 1847-1862.

Talk 4

How sticky are our proteins? Quantifying hydrophobicity of the human proteome

Dea Gogishvili, Juami van Gils, Jan van Eck, Robbin Bouwmeester, Erik van Dijk, *Sanne Abeln.*

Computer science, Vrije Universiteit Amsterdam, De Boelelaan 1111, 1081 HV Amsterdam, The Netherlands.

Proteins tend to bury hydrophobic residues inside their core during the folding process to provide stability to the protein structure and to prevent aggregation. Nevertheless, proteins do expose some 'sticky' hydrophobic residues to the solvent. These residues can play an important functional role, e.g. in protein–protein and membrane interactions. Here, we first investigate how hydrophobic protein surfaces are by providing three measures for surface hydrophobicity: the total hydrophobic surface area, the relative hydrophobic surface area and—using our MolPatch method—the largest hydrophobic patch. Secondly, we analyze how difficult it is to predict these measures from sequence: by adapting solvent accessibility predictions from NetSurfP2.0, we obtain well-performing prediction methods for the THSA and RHSA, while predicting LHP is more challenging. Finally, we analyze implications of exposed hydrophobic surfaces: we show that hydrophobic proteins typically have low expression, suggesting cells avoid an overabundance of sticky proteins.

Recently published in Bioinformatics Advances: https://doi.org/10.1093/bioadv/vbac002

Talk 5

USING WRITHE TO PRODUCE REALISTIC PROTEIN STRUCTURE PREDICTIONS FROM BIOSAXS DATA

Arron N Bale, Josh McKeown, *Chris Prior*

Department of Chemistry, MoSMed CDT (Centre for Doctoral Training), Durham University

The writhe is a measure of global self entanglement of a curve. The usual technique for studying entanglement of protein backbones requires artificial or probabilistic closures, thus altering the geometry of the curve. By using the open curve definition of the writhe we can study the geometry of protein curves in a more natural way. In particular, I will present bounds on the amount of global entanglement as measured by both the writhe and absolute writhe. With these bounds, we can produce realistically folded structures from BioSAXS data[1] with a much higher chance of success when passed into an MD simulation. These bounds, inspired by work in the magnetic helicity community[2], give rise to interesting classifications of proteins for further study. By studying the writhe of subsections of the protein's backbone, we are also able to highlight sites of interest such as highly flexible sections allowing large scale conformational changes or areas of maximal entanglement.

References:
[1] Prior, C., O. R. Davies, D. Bruce, and E. Pohl, 2020. *Obtaining tertiary protein structures by the ab initio interpretation of small angle X-ray scattering data*. Journal of chemical theory and computation 16:1985–2001.
[2] Cantarella, J.; DeTurck, D.; and Gluck, H., 2001. *Upper bounds for the writhing of knots and the helicity of vector fields*. AMS IP Studies in Advanced Mathematics, 24 (2001), 1–22.

Talk 6

Sample efficient structure-based de novo drug design in the hunt for novel GPCR antagonists with Augmented Hill-Climb

Morgan Thomas [a], Pierre Matricon [b], Noel M. O'Boyle [b], *Andreas Bender [a], Chris de Graaf [b]*

[a] *Centre for Molecular Informatics, University of Cambridge, Cambridge, UK*
[b] *Sosei Heptares, Cambridge, UK*

Artificial intelligence is now strongly embedded in computational drug discovery and design, showing particular promise *via* generative models for *de novo* molecule generation. Many goal-directed generative models use reinforcement learning (RL) to optimize towards a defined endpoint; however, this can be a sample inefficient process (sometimes requiring up to 10^5 molecule evaluations). This serves as a practical limitation when using more computationally expensive scoring functions, such as docking, for structure-based drug design. This is especially undesirable as structure-based methods can increase *de novo* molecule diversity and coverage of bioactive chemical space [1].

Augmented Hill-Climb (AHC) [2] is a hybrid RL strategy that improves sample efficiency by addressing the limitations of currently existing RL strategies. AHC improves both the optimization ability (by 1.5-fold) and sample efficiency (by 45-fold) compared to a baseline approach when conducting *de novo* structure-based design *via* Glide docking on four GPCR targets. AHC also outperforms other commonly used RL strategies on a benchmark of six tasks of varying difficulty with respect to sample efficiency and chemistry generated. The improvements outlined here enable the use of more computationally expensive structure-based scoring functions to be tractable on a reasonable timescale.

We now aim to share a comparison of AHC to a wider range of generative model approaches. In addition, we apply AHC in combination with Glide docking and multi-objective scoring strategies for *de novo* design of GPCR ligands. We will share approaches taken in this endeavour including, necessary docking and parameter constraints, novelty of proposed compounds, synthesizability and chemical space recovered with respect to crystal structure and parameters used.

References:

1. M. Thomas, R.T. Smith, N.M. O'Boyle, C. de Graaf and A. Bender, *J Cheminform*, (2021), **13**, 39.
2. M. Thomas, N.M. O'Boyle, A. Bender and C. de Graaf, *J Cheminform*, (2022) **14**, 68.

Talk 7

QMMM study of the Aromatic hydroxylation by Cinnamate 4-hydroxylase (C4H) a lignin regulator.

Sonia F. Santos[a,b,c], Gary W. Black[b,c], Meilan Huang[a] , *Warispreet Singh*[b,c]

[a]*Department of Chemistry & Chemical Engineering, Queen's University, Belfast, BT9 5AG, UK*
[b]*Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK*
[c]*Hub for Biotechnology in Build Environment, Newcastle upon Tyne, UK*

Being part of the plant metabolic system, Cinnamate 4-hydroxylase (C4H) enzyme belongs to the CYP73A family of the cytochrome P450 monooxygenase known for being a crucial biocatalyst. C4H is responsible for the hydroxylation of trans-cinnamic acid into p-coumaric acid.[1] This enzyme can have an impact on structure, development, and defence, which makes in-depth knowledge of the mechanism and interaction with the substrate during the catalytic reaction extremely important.

Playing different and impactful roles in plant's life, this enzyme must be properly studied in complex with their natural substrates. Trans-cinnamic acid and 2-Naphtoic acid were docked, into the active centre of C4H enzyme followed by comprehensive MD simulations and QM/MM calculations. However, unlike the hydroxylation of alkyl substrates in which the mechanism is initiated by the abstraction of the allylic hydrogen by the Iron-oxo complex, the case of aromatics is not so straightforward[2]. Based on our QM/MM calculations, we propose the aromatic hydroxylation to proceed through a tetrahedral sigma complex formation, to form a protonated porphyrin intermediate in comparison to the alternative epoxide and ketone pathways for aromatic hydroxylation.

References:

[1] B. Zhang, K. M. Lewis, A. Abril, D. R. Davydov, W. Vermerris, S. E. Sattler and C. Kang, *Plant Physiol.*, 2020, **183**, 957–973.
[2] C. M. Bathelt, A. J. Mulholland and J. N. Harvey, *J. Phys. Chem. A*, 2008, **112**, 13149–13156.

Talk 8

Molecular Dynamics of Cholesterol Membranes: A comparative Forcefield study

Jack Sawdon[a], Thomas J. Piggot[b], *Jonathan W. Essex*[a]

[a] *School of Chemistry, University of Southampton, Highfield Campus, Southampton, UK*
[b] *Chemical, Biological and Radiological Sciences Division, Defence Science and Technology Laboratory (DSTL), Porton Down, Salisbury, Wiltshire, UK*

Membrane properties are determined by lipid composition, and cholesterol plays a large role in affecting these properties. Cellular membranes show a diverse range of cholesterol compositions, the effects of which include alterations to: cellular biomechanics, lipid raft formation, membrane fusion, signalling pathways, metabolism, pharmaceutical therapeutic efficacy, and disease onset. Additionally, cholesterol plays an important role in non-cellular membranes, with its concentration in the skin lipid matrix being implicated with several skin diseases.

For phospholipid membranes, cholesterol increases tail ordering of neighbouring lipids, resulting in decreased membrane lateral area, and increased thickness. Decreasing lateral area as a function of increasing cholesterol concentration was first reported as the tendency of cholesterol lipid mixtures to negatively deviate from ideal mixing and has since been termed the cholesterol condensing effect. More recently, it has been shown that cholesterol DMPC membranes undergo much larger deviations from ideal mixing compared to DOPC membranes, resulting in a negative partial-specific area of cholesterol in DMPC membranes, but a positive value in DOPC membranes.

Here we present a comparative forcefield analysis of cholesterol models involving all-atom (CHARMM36, Slipids, Lipid17), united-atom (CKP, Poger, Berger) and coarse-grained (MARTINI, ELBA) forcefields. Simulations of cholesterol containing DMPC and DOPC membranes are performed, and average lipid areas are determined over a range of cholesterol concentrations to calculate mixing behaviour, partial-specific areas, and other condensing effect parameters.

While all the tested forcefields correctly predict small negative deviations from ideal mixing in cholesterol DOPC membranes, only all-atom forcefields capture the larger deviations expected in DMPC membranes. Compared to all-atom forcefields, united-atom and coarse-grained cholesterol models condense fewer neighbouring lipids by smaller magnitudes resulting in them under-predicting the negative deviation from ideal mixing in DMPC membranes. Results presented here suggest all-atom forcefields should be used if simulations of cholesterol-containing membranes are to properly capture the cholesterol condensing effect.

Talk 9

Molecular Dynamics Simulations of Engine Lubricant Additives

Angie Maťusová[a], *Philip J. Camp*[a], and *Peter J. Dowding*[b]

[a] *School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ*
[b] *Infineum UK Ltd, Milton Hill Business & Technology Centre, Milton Hill, Abingdon OX13 6BB*

Greenhouse-gas emissions caused by the transportation sector are a major contributor to climate change. Despite the development of hybrid and electric vehicles, the use of internal combustion engines in both road and marine transport will continue to rise for the foreseeable future.[1] Reducing friction, wear, fuel consumption, and emissions can be achieved with improved lubricant formulation, but it is a complex process involving a broad range of surfactant or polymer-like additives. Simulations help in providing molecular-level insights on the links between chemical composition of additives and tribological performance of lubricants.[2]

In this study, the LAMMPS package was used to carry out equilibrium and non-equilibrium molecular dynamics simulations on a brush-like oxazoline-based polymer in non-polar solvent. The aim was to investigate the effects that the polymer has on bulk properties of interest, such as viscosity and friction coefficients. The bulk systems were set up to complement small-angle neutron scattering experiments and offer clarification on clustering behavior and intermolecular interactions, justified through comparison in radii of gyration and form factors. Surface simulations provide better understanding of the tribological properties of the polymer-enriched lubricant in highly pressurized systems (up to 5 GPa). The simulation work shows excellent agreement with experiment, which lays the groundwork for further use of these methods in lubricant formulation and development.

References:
[1] R. Reitz et al., *International Journal of Engine Research*, 2019, 21, 3-10.
[2] K. Holmberg and A. Erdemir, *Friction*, 2017, 5, 263–284.

Talk 10

Characterization of GLP-1R Ala316Thr variant. Does genetic variability in GLP-1R have an impact on the management of type 2 diabetes?

Roxana-Maria Rujan[a]*, Liliane El Eid[b], Alejandra Tomas[b], Ben Jones[c], *Giuseppe Deganutti*[a] and *Christopher A. Reynolds*[a]

[a]*Centre for Sport, Exercise and Life Sciences, Faculty of Health and Life Sciences, Coventry University, United Kingdom*
**Medicines Discovery Catapult, Alderley Park, Macclesfield SK10 4ZF*
[b]*Section of Cell Biology and Functional Genomics, Imperial College London, London, United Kingdom*
[c]*Section of Endocrinology and Investigative Medicine, Imperial College London, London, United Kingdom*

Diabetes is a chronic metabolic disease characterised by high levels of glucose in the blood which can lead to serious health issues, including death. It is expected that the number of people affected world-wide by this disease will reach 642 million by 2040, with type 2 diabetes (T2D) accounting for around 90% of the cases. One major pharmacological target in the treatment of T2D is the glucagon-like peptide-1 receptor (GLP-1R), a Class B GPCR, due to its insulinotropic effects. Polymorphisms in this receptor are a clear challenge to treatment options in patients, with some variants leading to significantly poorer outcomes than others.

One missense variant, Ala316Thr, is particularly interesting as it is associated with lower T2D risk and generally beneficial treatment outcomes. In this work, we aim to characterise and explore the biological and atomic nature of this variant together with the biased signalling at this receptor to better understand what could be driving its physiological effects.

Using a combination of dry- (molecular modelling and molecular dynamics (MD) simulations) and wet-lab (confocal and high-content microscopy along with NanoBit assays) techniques we explored this variant in its apo and agonist bound states using the native GLP-1 agonist and its analogues: exendin-4, exendin-Phe1, and exendin-Asp3 in comparison to the wild-type receptor.

We show that this polymorphism leads to significant changes in receptor dynamics, signalling and trafficking. There is altered flexibility of the extracellular domain and transmembrane regions in both pre- and post-activation states as well as a loss of cell surface expression, decreased mutant receptor recycling, increased internalisation, and augmented degradation.

As we gain a better understanding of how genetic variation can drive treatment response, this work is timely and an essential step in the understanding and characterisation of polymorphisms in GLP-1R, paving the way for a personalised medicine approach for T2D.

References:

1. J. Klen and V. Dolžan, *Int. J. Mol. Sci.*, (2022) **23,** (7), 3451.
2. K. Ogurtsova, J.D. da Rocha Fernandes, Y. Huang, U. Linnenkamp, L. Guariguata, N.H. Cho, D. Cavan, J.E. Shaw and L.E. Makaroff, *Diabetes Res. Clin. Pract.*, (2017) **128**, 40–50.

Talk 11

Complexity Matters: Polymyxin Antibiotics within the E. coli Cell Envelope

Iain Smith[a], *Syma Khalid*[a,b]

[a] *School of Chemistry, University of Southampton, Southampton, SO17 1BJ*
[b] *Department of Biochemistry, University of Oxford, Oxford, OX1 3QU*

We present a computational study of the effects of biomolecular crowding on the behaviour of the antimicrobial peptides (AMPs) polymyxin B1 (PMB) and polymyxin E (PME) within the periplasm of *E. coli*. We have employed atomistic molecular dynamics simulations to explore the nature of the interaction between these AMPs and both Braun's lipoprotein (BLP) and the peptidoglycan (PGN) cell wall in models of the periplasm that are crowded to different extents. The most crowded model of the periplasm contained a range of osmolytes as well as crowding proteins (ubiquitin) in addition to the peptides, cell wall and Braun's lipoprotein. All systems also included a model of the *E. coli* outer membrane. Overall, three x 250 ns replicates of each system were performed (with both just neutralising ions and also 150 mMol concentration of NaCl); giving a total simulation time of 10.5 microseconds. Our results show the importance of D-phenylalanine to the primarily hydrophobic interaction between PMB and BLP; we characterise the change in this interaction with increased molecular crowding and highlight how the replacement of this moiety by D-leucine in PME altered the balance of interactions. A similar analysis of the interactions between the peptides and the cell wall highlight the prevalence of electrostatic interactions between the specific (DAP) residues of the cell wall and both polymyxins. Finally, the impacts of both polymyxins on the dimensions of the 'mesh' of the cell wall are discussed. Our results highlight the importance of considering the full biological complexity of a system when attempting to characterise details about the function and behaviour of antimicrobial peptides in vivo; as evidenced by an increase in the range of observed behavioural states of polymyxins as the complexity of the system increased.

Talk 12

Can You Hear the Shape of a Drug?

Rachael M. E. Pirie, *Stuart J. Hall, Daniel J. Cole*

School of Natural and Environmental Sciences, Newcastle University, Bedson Building, Kings Road, Newcastle Upon Tyne, NE1 8QB.

The principle that similar molecules will share biological activity has been used widely in hit identification, requiring only a small handful of known binders as templates to rapidly screen large databases. The need for protein-ligand complementarity to permit binding makes molecular shape useful in quantifying similarity.[1] Use of 3D shape, compared to other measures of similarity, allows consideration of different conformers and enables scaffold hopping, a useful property for rescue of candidates with undesirable properties. There is no absolute definition of molecular shape. Instead, we make use of mathematical approximations to produce vector descriptions of shape.

This talk presents two novel approaches to the approximation of molecular shape, derived from the theory of Riemannian geometry. Central to both methods is an object associated with the surface known as the *Riemannian metric*, that captures details of the geometry of the surface, but cannot be directly compared. In RGMolSA,[2] the metric is used to approximate the spectrum of the Laplacian, producing a simple 9-element vector descriptor of molecular shape. KQMolSA applies the theory of Kähler Quantisation, using the Riemannian metric and the associated Kähler potential to generate a matrix descriptor of shape. Both descriptors are alignment-free, quick to compute and easy to compare. The results of an initial case study comparing both approaches to the existing work in the field, as well as considering the ability to account for different conformers and potential performance in a real virtual screen will be discussed.

References:
[1] A. Kumar and K. Y. J. Zhang, *Front. Chem.*, (2018), **6**, 1-21.
[2] D. J. Cole, S. J. Hall and R. Pirie, *arXiv:2201.04230* (2022)

Programme of Poster Presentations

1	"Modelling type II polyketide synthase chain elongation reactions" Elise White, Bristol University
2	"Long-range dispersion-inclusive machine learning potentials for structure search and optimization of hybrid organic-inorganic interfaces" Shayantan Chaudhuri, Warwick University
3	"Molecular dynamic simulations reveal convergent structural evolution of distinct mutations within the membrane protein of SARS-CoV-2" Andrew Mack, Cardiff University
4	"Towards elucidating the molecular mechanisms of the skin penetration enhancer propylene glycol" Jade Mistry, Warwick University
5	"Investigation of the Effect of Cu(II) on the Structural Characteristics of α-Synuclein" Loizos Savva, Cardiff University
6	"Using BioSimSpace to create an interoperable Relative Binding Free Energy Pipeline" Anna Herz, Edinburgh University
7	"Enhancing Ligand Torsional Sampling using Fully Adaptive Simulated Tempering (FAST): A case study of relative binding free energies for P38 MAP Kinase" Anna Cavalleri, Southampton University
8	"Listening to Proteins with EARS (Extraordinary Acoustic Raman Spectroscopy)" Cameron McAllister, Durham University
9	"Transfer learning for heterocycle retrosynthesis prediction" Ewa Wieczorek, Oxford University
10	"Emerging variants of SARS-CoV-2 NSP10 have negligible effects on its binding to two non-structural proteins, nsp14 and nsp16" Huan Wang, University College London
11	"Conformational dynamics and putative substrate extrusion pathways of the *N*-glycosylated outer membrane factor CmeC from *Campylobacter jejuni*" Kahlan Newman, Southampton University
12	"Accurately and Rapidly Predicting Free Energies Using ML Potentials" Tristan Johnston-Wood, Oxford University

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

Modelling type II polyketide synthase chain elongation reactions

Elise White[a,b], *Matthew P. Crump[b], Marc W. van der Kamp[a]*

[a]School of Biochemistry, University of Bristol, University Walk, Bristol BS8 1TD
[b]School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS

Many polyketides are biomedically active; they are biosynthesised from malonate-based units by polyketide synthases (PKSs). In type-II PKSs, the growing polyketide is shuttled between individual PKS enzymes whilst tethered to an acyl carrier protein (ACP) by a phosphopantetheine (Ppant) arm. A ketosynthase chain-length-factor (KS-CLF) catalyses initial chain elongation reactions to synthesise the polyketide backbone. Due to the transient nature of the ACP-KS-CLF interaction and the reactive nature of the substrates, there are few structures of these enzymes in complex, with the polyketide present. Therefore, starting structures are not readily available for molecular dynamics (MD) simulations that would provide detailed insight into chain elongation reaction mechanisms.

For the anthraquinone (*ant*) and actinorhodin (*act*) PKSs, models were built of ACP-KS-CLF complexes for different states of the chain elongation reaction. *ant* structures were based on an *ant*ACP-Ppant-*ant*KS-CLF co-crystal structure with a captured polyketide intermediate. *act* structures were obtained following protein-protein docking of *act*ACP and *act*KS-CLF and manual placement of the polyketide and Ppant. These models were iteratively refined through analysis of key interactions and atomic distances during molecular mechanical (MM) MD simulations.

Analysis of polyketide behaviour in these final refined models, during MM MD simulations, has given insight into polyketide chain length control, ultimately informing possible mutations to alter polyketide product length. These final refined models were also used in combined quantum mechanical/molecular mechanical (QM/MM) reaction simulations to probe the carbon-carbon bond formation and carbon-sulphur bond breakage reactions that occur in the chain elongation reaction. The in-depth understanding obtained about this enzyme reaction could eventually inform engineering of PKSs to yield potentially bioactive novel polyketides.

References:

[1] Bräuer, A., Zhou, Q., Grammbitter, G. L. C., Schmalhofer, M., Rühl, M., Kaila, V. R. I., Bode, H. B., and Groll, M. (2020). Structural snapshots of the minimal PKS system responsible for octaketide biosynthesis. *Nat. Chem. 12*, 755–763.

Poster 2

Long-range dispersion-inclusive machine learning potentials for structure search and optimization of hybrid organic-inorganic interfaces

Shayantan Chaudhuri [a,b], Julia Westermayr [a], Andreas Jeindl [c], Oliver T. Hofmann [c], *Reinhard J. Maurer* [a]

[a] *Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom*
[b] *Centre for Doctoral Training in Diamond Science and Technology, University of Warwick, Coventry, CV4 7AL, United Kingdom*
[b] *Institute of Solid State Physics, Graz University of Technology, 8010 Graz, Austria*

The computational prediction of the structure and stability of hybrid organic-inorganic interfaces provides important insights into the measurable properties of electronic thin film devices, coatings, and catalyst surfaces and plays an important role in their rational design. However, the rich diversity of possible surface configurations and the important role of long-range interactions represent a challenge for structure exploration based on first-principles electronic structure calculations.[1] We present a machine learning (ML) approach that enables fast, yet accurate, structure optimisations by combining two different types of deep neural networks trained on first-principles electronic structure data. The first model is a short-ranged interatomic ML potential trained on local energies and forces, while the second is an ML model of effective atomic volumes derived from atoms-in-molecules partitioning. The latter can be used to connect short-range potentials to well-established density-dependent long-range dispersion correction methods. We investigate the formation of gold nanoclusters on diamond (110) surfaces and train models on sparse structure relaxation data from density functional theory to show the ability of the models to deliver highly efficient structure optimisations.

References:
[1] O. T. Hofmann, E. Zojer, L. Hörmann, A. Jeindl, R. J. Maurer, *Phys. Chem. Chem. Phys.,* (2021), **23**, 8132–8180

Poster 3

Molecular dynamic simulations reveal convergent structural evolution of distinct mutations within the membrane protein of SARS-CoV-2

Andrew Mack[a], Agnel Praveen Joseph[b], Georgina Menzies[a], Dafydd Jones[a], Thomas Connor[a]

[a]School of Biosciences, Cardiff University, Cardiff, United Kingdom
[b]Scientific Computing Department, Science and Technology Facilities Council, Research Complex at Harwell, Didcot, OX11 0FA, UK

Introduction

SARS-CoV-2 is an RNA virus and the causative agent of the COVID19 pandemic. With modern sequencing capabilities, the scientific community has been able to track its dissemination and evolution with unprecedented resolution. Here we deployed evolutionary methodologies to screen SARS-CoV-2 whole genome sequences and identified co-evolving signals within its homo-dimeric membrane protein. We conducted molecular dynamic simulations, with these mutant proteins embedded into membrane to determine their significance.

Methods

A representative cohort of SARS-CoV-2 whole genome sequences were obtained from GISAID repository. The whole genome sequences were translated to relevant protein encoding regions and a multiple sequence alignment was generated. The membrane protein encoding sequences were analysed for co-evolving signals using a Bayesian graphical network approach, which determines the probably of two mutations being co-dependent. Two clusters of distinct mutations were identified, one being I82T from the Delta lineage and the second cluster I63T & Q19E that is found in all three Beta, Alpha and Omicron lineages. These were modelled onto a pre-existing experimental structure using CHARMM-GUI. The two mutant and one wild-type structure were embedded into membrane structures also using CHARMM-GUI. Molecular dynamics simulations were ran in triplicate for 100ns using Gromacs.

Results

Analysis of the mutant membrane proteins revealed stabilisation and destabilisation at comparable locations across both mutant proteins when compared with the wild type. Although the mutations within the two mutants are at differing locations within the protein, the RMSF of the mutants were better correlated with each other (r= 0.92) than the wild type (r= 0.84 & r = 0.86).

Discussion

Here, we demonstrate an approach to identify independently evolved mutations with convergent structural consequence. This suggests that although different genotypes of the virus exist, they may in fact exhibit a similar phenotype and may contribute the same benefit to the overall fitness of the virus.

Poster 4

Towards elucidating the molecular mechanisms of the skin penetration enhancer propylene glycol

Jade Mistry, *Rebecca Notman*

Department of Chemistry, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK.

Transdermal drug delivery (TDD) is a non-invasive method for the controlled delivery of drugs through skin directly into the blood stream. However, a major limitation of TDD is that very few drugs have the correct physicochemical properties to cross the skin's main permeability barrier, the stratum corneum (SC). The SC is the outermost layer of the skin, and contains layers of corneocytes embedded in a lipid matrix composed of ceramides, fatty acids, and cholesterol. One of the most common strategies employed in TDD to temporarily overcome the barrier properties of the skin is the use of chemical penetration enhancers (PEs). PEs may facilitate the transport of drugs across the skin by altering the permeability of the SC.

Propylene glycol (PG) is a widely used skin PE that is known to diffuse into the SC[1] and interact with the SC lipids to increase the permeation of drugs.[2] Experiments have demonstrated that PG increases the mobility and disorder of SC lipids, and may extract cholesterol from the SC. However, little is known about the molecular mechanisms of drug permeation enhancement by PG.

In this work, all-atom and united atom MD simulations were performed to investigate the effect of PG on the structure and properties of model skin-lipid bilayers at the molecular level. First, the 2 forcefield models of PG were validated against experimental results, and then the interactions of various concentrations of PG with model skin-lipid bilayers of the 2 forcefields were investigated. PG was found to localise in the hydrophilic headgroup regions at the bilayer interface rather than permeate the bilayer, and was able to occupy lipid-water hydrogen-bonding sites, as well as increase disorder in the lipid tails.

References:

[1] A. C. Williams and B. W. Barry, *Adv. Drug Deliv. Rev.*, 2004, **56**, 603–618.
[2] P. D. A. Pudney, M. Mélot, P. J. Caspers, A. Van Der Pol and G. J. Puppels, *Appl. Spectrosc.*, 2007, **61**, 804–811.

Poster 5

Investigation of the Effect of Cu(II) on the Structural Characteristics of α-Synuclein

Loizos Savva[1]; *James A. Platts[1]*

[1]School of Chemistry, Cardiff University, United Kingdom

Over the last two decades, the binding domains of Cu(II) to α-Synuclein (αS) have been characterised, with experimental evidence on the aggregation capacities of the peptide upon metal ion binding, establishing the promotion of fibril formation and by extension accumulation of αS in the human brain – a possible trigger of Parkinson's disease. We employ implicit solvent molecular dynamics simulations, after validating the performance of the parameterisation technique against experimental data on the unbound peptide.[1] Stabilisation of the particular intrinsically disordered nature of the peptide, is observed upon metal coordination, with an increase in the compactness of the overall Cu(II)-bound system and decreased RMSF values in the regions involved in or neighbouring the coordination sites. We show increased lifetime of β-hairpin regions found in residue repeats, starting in the N-terminal and extending into the central, non-amyloid-β component (NAC) regions, upon Cu(II)-coordination.

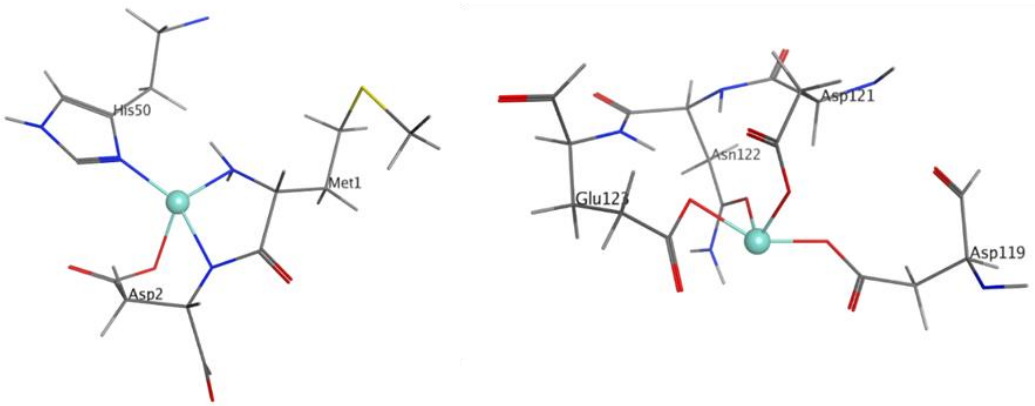

Figure 1 Coordination of ligating atoms in each of the Cu(II)-binding sites.

References:

[1] L. Savva and J. A. Platts, *J. Biomol. Struct. Dyn.*, (2022), 1–16.

Poster 6

Using BioSimSpace to create an interoperable Relative Binding Free Energy Pipeline

Anna M. Herz[a], Tahsin Kellici[b], Inaki Morao[b], *Julien Michel*[a]

[a] *School of Chemistry, The University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, United Kingdom*
[b] *EVOTEC, 114 Innovation Drive, Milton Park, Milton, Abingdon, Oxfordshire OX14 4RZ, United Kingdom*

FEP (Free Energy Perturbation) methods are used to computationally predict binding affinities, with RBFE (Relative Binding Free Energy) specifically being used to rank compounds during preclinical drug discovery stages. Previous work has shown that relative free energies of hydration are reproducible to within roughly 0.2 kcal/mol between the molecular dynamics engines AMBER, SOMD, GROMACS, and CHARMM, however no universally applicable protocol was established and considerable attention to detail and package specific simulation protocols needed to be observed.[1] For RBFE, recent work has shown a reproducibility of roughly 0.5 kcal/mol between OpenMM, NAMD2, and NAMD3.[2]

This work aims to establish an interoperable FEP pipeline to enable the rapid testing of RBFE calculations and evaluate the reproducibility between three different molecular dynamics engines - SOMD, GROMACS, and AMBER. Initial work has focused on expanding upon functionality within BioSimSpace, a python framework for biomolecular simulation. This has included the in-cooperation of AMBER for RBFE and Hydrogen Mass Repartitioning for all engines. Establishing the pipeline was done using tyrosine kinase 2 (TYK2), myeloid leukemia cell differentiation protein (MCL1) and mitogen-activated protein kinase 14 (p38α), along with some tractable ligand transformations as test systems. Initial results have shown a maximum mean unsigned error of 1 kcal/mol between the engines and highlighted important considerations regarding the need for appropriate equilibration protocols, consistent analysis methods, and engine specific automated identification of the transforming atoms for the engine specific atom masks.

References:
[1] H. H. Loeffler, S. Bosisio, G. Duarte Ramos Matos, D. Suh, B. Roux, D. L. Mobley and J. Michel, *J. Chem. Theory Comput,* 2018, **14**, 5567–5582.
[2] A. D. Wade, A. P. Bhati, S. Wan and P. V. Coveney, *J. Chem. Theory Comput,* 2022, **18**, 3972–3987.

Poster 7

Enhancing Ligand Torsional Sampling using Fully Adaptive Simulated Tempering (FAST): A case study of relative binding free energies for P38 MAP Kinase

Anna Cavalleri [1], Khaled Abdel-Maksoud [1], Kamaldeep K. Chohan [2], *Jonathan W. Essex* [1]

[1] *School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK*
[2] *Sygnature Discovery Limited, BioCity, Pennyfoot St, Nottingham NG1 1GR, UK*

Using classical MD methods to perform alchemical relative binding free energy calculations is well established and highly popular, despite the awareness that conventionally these methods are prone to poor sampling and forcefield errors. Often, slow modes in the protein secondary structure or within the ligand itself are missed or incorrectly captured in the typical timescales of a simulation, leading to imprecise and inaccurate free energy estimates. Recent work surrounding the choice of initial crystal structure has also been shown to influence the accuracy of free energy calculations [1].

Overcoming these challenges is of particular relevance for systems where the protein has a high degree of flexibility as well as when the ligand has multiple rotatable bonds. One such system is P38 MAP Kinase, which has flexible loops close to the binding site [2] and several rotatable torsions in the ligand structure.

In this work, we present a method based on Sequential Monte-Carlo (SMC) [3] and Simulated Tempering (ST) [4] to alchemically transform between two ligands with enhanced sampling directed towards user specified torsions within the ligand. SMC provides an initial lambda pathway for the alchemical transformation, and ST allows for long timescale sampling of these intermediates. As the simulation continues, the initial protocol is optimised further to provide a fully adaptive method regardless of the system used.

From our investigation, we sample more diverse binding modes and we gain free energy estimates which generally show greater variance than traditional alchemical methods. The implications of this balance between adequate sampling and variance in the calculated free energies will be discussed.

References:

[1] M. Suruzhon, M.S. Bodnarchuk, A. Ciancetta, R. Viner, I.D. Wall, J.W. Essex, *J Chem Theory Comput*, 2021, 17(**3**), 1806-1821
[2] L. Yang, X. Sun, Y. Ye, Y. Lu, J. Zuo, W. Liu, A. Elcock, S. Zhu, *Front. Oncol.,* 2019, **9**, 1294 [3] J.S. Liu, R. Chen, *J Am Stat Assoc*, 1998, 93(**443**), 1032-1044

Poster 8

Listening to Proteins with EARS (Extraordinary Acoustic Raman Spectroscopy)

Cameron McAllister, *Prof Elizabeth Bromley, Dr Matteo Degiacomi, Prof Kevin Weatherill*

Department of Physics, University of Durham, Lower Mountjoy, South Rd, Durham DH1 3LE

The low frequency global vibrational modes of proteins are theorised to be involved in a range of important biological processes, in particular dynamic allostery and enzymatic catalysis. However, there remains a paucity of experimental techniques to probe protein vibrations in this sub-THz range. In 2015, the technique of Extraordinary Acoustic Raman Spectroscopy (EARS) was demonstrated for the first time and appeared to demonstrate spectra of a range of proteins in the 3-300 GHz range[1]. To interpret these spectra, an accurate modelling procedure to characterise the dynamic of proteins in the experiment is required. However, currently no satisfactory modelling routine to simulate EARS spectra exists.

We have developed a method for simulating EARS spectra. To this end, we utilise Normal Modes Analysis (NMA) of an alpha carbon Anisotropic Network Model (ANM) to determine a protein's characteristic motion. Protein structures representing extremes of motion along each mode are solvated and energy minimised. This enables us to accurately calculate the changes in volume as a proxy for the change in electronic polarizability captured in Raman spectra.

With this modelling routine, we could simulate spectra for the Bovine Pancreatic Trypsin Inhibitor (BPTI) protein that more closely match the experimental results than previously published methods, particularly with regards to the frequencies of the dominant peaks.

Still, a limitation of an NMA-based approach is that it only captures dynamics around a single energy minimum. For this reason, we are currently improving our modelling routine by exploiting Molecular Dynamics (MD) simulations. Simulations analysis via Markov state modelling is enabling us to identify main protein states, generating characteristic EARS spectra for each of them.

References:

[1] S. Wheaton, R. M. Gelfand, and R. Gordon, *Nat. Photon.*, (2015), **9(1)**, 68-72.

Poster 9

Transfer learning for heterocycle retrosynthesis prediction

Ewa Wieczorek,[a,b] *Paul E. Brennan,*[b] *Fernanda Duarte*[a]

[a] Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford OX1 3TA, U.K.
[b] Alzheimer's Research UK Oxford Drug Discovery Institute, Centre for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, U.K.

Small heterocycles play a central role in synthetic and medicinal chemistry, with most pharmaceuticals and agrochemicals containing at least one of such rings. These motifs have been recognised as successful modulators of shape, pharmacokinetics, and binding properties of drugs.

The importance of heterocyclic scaffolds has recently sparked an interest in exploring them computationally, with several libraries being published enumerating new rings and their properties. However, the lack of synthetic protocols to make such rings obstructs medicinal chemists from incorporating them into drug-like molecules. A possible solution to aid their synthesis is to predict retrosynthetic routes computationally. While several computer-assisted synthesis planning tools are currently available, they tend to perform poorly for heterocycle-containing molecules, most likely due to under-representation of ring formation reactions in the training datasets.

In this contribution, we combine the use of transfer-learning strategies[1] for retrosynthesis prediction with a transformer model[2] to train a model that improves the prediction of ring-forming reactions. To do so, we use a general dataset of chemical reactions together with a carefully curated dataset of ring formations only. For this task, our heterocycle-focused model has higher accuracy and predicts more viable ring-breaking disconnections than the baseline model. Additionally, there is no decrease in performance for more common reaction types. This makes the model suitable for use in a multi-step retrosynthesis tool, being able to predict both the ring-opening disconnection as well as breaking down the linear intermediates into commercially available building-blocks.

References:
[1] G. Pesciullesi, P. Schwaller, T. Laino and J.-L. Reymond, *Nat Commun*, 2020, **11**, 4874.
[2] P. Schwaller, R. Petraglia, V. Zullo, V. H. Nair, R. A. Haeuselmann, R. Pisoni, C. Bekas, A. Iuliano and T. Laino, *Chem. Sci.*, 2020, **11**, 3316–3325.

Poster 10

Emerging variants of SARS-CoV-2 NSP10 have negligible effects on its binding to two non-structural proteins, nsp14 and nsp16

Huan Wang, Danni Dong, Jiaqi Lou, Qian Wang, Watanyoo Sopipong, Frank Kozielski and *Shozeb Haider*

School of Pharmacy, University College London, 29-39 Brunswick Square, London WC1N 1AX, United Kingdom

The coronavirus SARS-CoV-2 protects its viral RNA from being recognized by host immune responses by methylation of its 5' end, also known as capping. This process is carried out by two enzymes, NSP16 containing 2'-O-methyltransferase and NSP14 through its N7 methyltransferase activities, which are essential for the replication of the viral genome as well as evading the host's innate immunity. The non-structural protein 10 (NSP10) acts as a crucial cofactor and stimulator of NSP14 and NSP16. To further characterise the role of NSP10, we carried out a comprehensive analysis of >7 million globally collected whole-genome sequences (WGS) of SARS-CoV-2 obtained from the Global Initiative Sharing All Influenza Data (GISAID) and compared it with the reference genome Wuhan/WIV04/2019 to identify all currently known variants in NSP10. T12I, T102I and A104V in NSP10 have been identified as the three most frequent variants and characterised using biophysical and structural methods. In contrast to other non-structural proteins such as NSP1 and NSP12, the RNA-dependent RNA polymerase, a well-characterised and important drug target, NSP10 is significantly less prone to mutation due to its crucial role in replication. The functional effects of the variants were examined for their impact on the binding affinity and stability of both NSP14-NSP10 and NSP16-NSP10 complexes. These results highlight the limited changes induced by variant evolution in NSP10 and reflect on the crucial roles NSP10 plays during the SARS-CoV-2 life cycle. These results also indicate that there is limited capacity for the virus to overcome inhibitors targeting NSP10 via the generation of variants in inhibitor binding pockets.

Poster 11

Conformational dynamics and putative substrate extrusion pathways of the *N*-glycosylated outer membrane factor CmeC from *Campylobacter jejuni*

Kahlan E. Newman[a], *Syma Khalid*[a,b]

[a] *School of Chemistry, University of Southampton, Southampton, SO17 1BJ*
[b] *Department of Biochemistry, University of Oxford, Oxford, OX1 3QU*

The outer membrane factor CmeC of the efflux machinery CmeABC plays an important role in conferring antibiotic and bile resistance to *Campylobacter jejuni*. Curiously, the proteins in this machinery are *N*-glycosylated, with the glycans playing a key role in the effective function of the efflux system. Here we have employed atomistic equilibrium molecular dynamics simulations of CmeC in a representative model of the *C. jejuni* outer membrane to characterise the dynamics of the protein and its associated glycans. We show that the *N*-glycans are more conformationally labile than had previously been thought. The extracellular loops of CmeC visit the open and closed states freely suggesting the absence of a gating mechanism on this side, while the narrow periplasmic entrance remains tightly closed, regulated *via* coordination to solvated cations. We identify several cation binding sites on the interior surface of the protein. Additionally, we have used steered molecular dynamics simulations to elucidate translocation pathways for a bile acid, chenodeoxycholic acid, and a macrolide antibiotic, erythromycin. These, and additional equilibrium simulations suggest that the anionic bile acid utilises multivalent cations to climb a ladder of acidic residues that line the interior surface of the protein. In combination, these results further our understanding of the dynamics of outer membrane factors and allow insight into the translocation pathways of antibiotics and other known substrates in this region of the efflux machinery.

Poster 12

Accurately and Rapidly Predicting Free Energies Using ML Potentials

Tristan Johnston-Wood, Hanwen Zhang, Tom Young, *Fernanda Duarte*

Physical and Theoretical Chemistry Laboratory, University of Oxford, S Parks Rd, Oxford OX1 3QZ

Accurate prediction of free energy barriers is fundamental to elucidating kinetic and thermodynamic properties of a system. However, achieving accuracy such that you can make accurate predictions remains challenging. Traditionally, free energies have been computed using static electronic structure calculations, often at the DFT level, or classical molecular dynamics (MD) methods employing force fields. More accurate approaches include *ab initio* MD (AIMD); however, due to their computational cost, they are limited to small simulation times and small systems. Machine learning (ML) methods have the potential to transform molecular simulations, providing the best of both worlds.

We have previously used the Gaussian Approximation Potential (GAP)[1] framework to explore chemical reactivity of a range of systems, e.g., small organic reactions, metals in water and metallocages.[2] Here, we extend this methodology to comparatively study different ML potentials. The protocol is implemented in a Python package, *mlptrain*, which automates the training and free energy calculation. The performance of these methods is tested using a Diels-Alder reaction in gas phase as a model system. The most data efficient and accurate trained potential is then used to propagate both classical and quantum MD trajectories at an AIMD level of accuracy but at a reduced computational cost. Umbrella sampling and WHAM are then used to predict the free energy along the reaction coordinate.

We have applied these methods to gas phase systems, including methylchloride S_N2 reactions, ethene + butadiene and CPD + MVK Diels-Alder reactions. The free energy has computed within a few kcal mol^{-1} relative to experiment and under an hour after training. We then extend this methodology to more complex systems, such as migratory reactions involving hydrogen transfer, with the inclusion of quantum dynamics to account for tunnelling effects.

References:
[1] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Phys. Rev. Lett.*, 2010, **104**, 136403
[2] T. A. Young, T. Johnston-Wood, V. L. Deringer and F. Duarte, *Chem. Sci.*, 2021, **12**, 10944–10955

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