Molecular Graphics and Modelling Society
Young Modellers' Forum 2021-22
PROGRAMME & ABSTRACTS
11th February 2022

Programme of Oral Presentations

09:15-09:30	Opening Remarks
09:30-09:50	**The second extracellular loop of class A GPCRs: sixty structures, how many shapes?** Alessandro Nicoli, Leibniz Institute for Food Systems Biology at the Technical University of Munich
09:50-10:10	**A Computational Descriptor Analysis on Excited State Behaviours of a Series of TADF Emitters** Pelin Ulukan, Bogazici University
10:10-10:30	**Chiral knots display symmetry breaking in helical confinement** Renáta Rusková, Polymer Institute of Slovak Academy of Sciences
10:30-11:00	Lightning Talks from Poster Presenters
11:00-11:30	BREAK and POSTER SESSION
11:30-11:50	**Modelling Network Formation in Protein Hydrogels by Cluster Aggregation Kinetics** Kalila Cook, University of Leeds
11:50-12:10	**Benchmarking state of the art RNA Force Fields** Dimitris Stamatis, University of Southampton
12:10-12:30	**Active site sequences are enough for proteochemometric modelling of human kinases** Jannis Born, IBM Research Europe
12:30-13:50	LUNCH and POSTER SESSION 12:30- 13:10 Posters 1-6 available in Breakout Rooms P1-P6 13:10- 13:50 Posters 7-12 available in Breakout Rooms P7-P12
13:50-14:10	**Towards agreement between experiment and theory for barriers and kinetic isotope effects: addressing solvent and counterion interactions in solution** Michele Assante, Liverpool John Moores University
14:10-14:30	**A data-driven approach to relative Free Energy Perturbation reliability predictions for alchemical free energy calculations in drug design** Jenke Scheen, University of Edinburgh
14:30-14:50	**Solvent effects on the photochemical decarbonylation of cyclopropenones** Daniel Adrion, Northeastern University
14:50-15:20	BREAK and POSTER SESSION
15:20-15:40	**Investigating molecular mechanisms of antibiotic permeation through outer membrane porins in high dimensional conformational space** Nandan Haloi, University of Illinois
15:40-16:00	**Modelling Organic-Inorganic Interfaces of Urinary Calculi** Rhiannon Morris, University of Leeds
16:00-16:20	**Investigating the activity and inhibition mechanisms of SARS-CoV-2 Mpro using molecular simulations** Henry Chan, University of Oxford
16:20-16:45	JUDGING, FINAL REMARKS, PRIZE ANNOUNCEMENTS

Talk 1

The second extracellular loop of class A GPCRs: sixty structures, how many shapes?

Alessandro Nicoli[a], Andreas Dunkel[a], Toni Giorgino[b], Chris de Graaf[c], *Antonella Di Pizio[a]*

[a] Leibniz Institute for Food Systems Biology at the Technical University of Munich, 85354 Freising, Germany
[b] Biophysics Institute, National Research Council (CNR-IBF), 20133 Milan, Italy
[c] Sosei Heptares, Steinmetz Building, Granta Park, Abingdon, Cambridge CB21 6DG, UK

G-protein coupled receptors (GPCRs) are seven-transmembrane proteins involved in signal transduction across the cell membrane. Class A GPCRs constitute approximately 35% of drug targets.[1] The extracellular loop 2 (ECL2) is the longest and the most diverse loop among class A GPCRs, and contributes to the ligand binding and selectivity.

In this work, sixty experimental ECL2 structures were analyzed, based on their sequences, shapes and intramolecular contacts. To take into account the flexibility, and, consequently, to increase the ECL2 structural description, we incorporated into our analyses information from the Molecular Dynamics trajectories available on the GPCRmd website (https://submission.gpcrmd.org/home/, 84 MD trajectories for 28 unique GPCRs). To reduce the number of frames, but to keep the structural information, we clustered and extracted 10 frames from each MD trajectory.

All the structures were aligned to the κ-opioid receptor, the closest receptor to the consensus Class A GPCR sequence, and were cut at anchor positions 4.57 and 5.37, according to the GPCRdb structural numbering (https://gpcrdb.org/).

Despite the high sequence variability, the analyzed structures can be clustered into seven main groups based on volume overlaps:[2] two singletons, the long unstructured ECL2 of P2Y12 and the short ECL2 of MC4R; structures with a β-sheets fold oriented over the binding site (i.e. peptide GPCRs), and those with a β-sheets fold oriented over the binding site; the ECL2 of the adenosine A2a and A1 receptors; the ECL2 of the aminergic receptors and the longest ECL2 of the FFA1; and the ECL2 of lipid GPCRs, which group structures that form an intra-loop disulfide bridge instead of the conserved ECL2-TM3 disulfide bridge.

To compare the loops, we then investigated the contribution of side chains through a contact analysis (http://ecl2.giorginolab.it/). Interestingly, the number of contacts is a factor that discriminates between different conformations within the same GPCR subtype.

References:
[1] A. Hauser, M. Attwood, M. Rask-Andersen, et al., *Nat Rev Drug Discov*, (2017), **16**, 829–842
[2] Phase, Schrödinger, LLC, New York, NY, 2021

Talk 2

A Computational Descriptor Analysis on Excited State Behaviours of a Series of TADF Emitters

Pelin Ulukan,[1] Ekin Esme Bas,[1] Rengin Busra Ozek,[1] Cansul Dal Kaynak,[1] Antonio Monari,[2] Viktorya Aviyente,[1,]* and *Saron Catak*[1,]*

[1]Department of Chemistry, Bogazici University, Bebek, 34342 Istanbul, Turkey
[2]Université de Paris and CNRS, Itodys, F-75006 Paris, France.

Over the last decade, Organic Light Emitting Diodes (OLEDs) have attracted a widespread attention especially for display technology and a significant progress has been made for their development and applications. Along with many advantages such as mechanical flexibility of the devices and reduced power consumption, some important drawbacks subsist, especially related to the insufficient device efficiency, which limit their applications. Following spin-statistics rule, the recombination of electrons and holes leads to formation of singlets (25 %) and triplets (75 %), hence fluorescent devices inherently loos two third of the harvested energy.[1] To address this problem, Phosphorescent OLEDs with heavy metal containing organic molecules were suggested.[2] The use of heavy metals in these systems leads to increase in spin – orbit coupling (SOC) between spin angular momentum and orbital angular momentum, thus triggering phosphorescence.[2] However, due to high cost and environmental contamination, application of phosphorescent materials is limited. Accordingly, Thermally Activated Delayed Fluorescence (TADF) have been proposed as a high efficiency and environmentally friendly alternative.[2]

The key step leading to efficiency increase in TADF materials is the reverse intersystem crossing (RISC) process which is observed as the conversion from T1 to S1 state. RISC can be achieved in organic molecules with small energy gap between lowest lying singlet and triplet states (ΔE_{S1-T1}) and in molecules having strong SOC between the involved states.[2] Though several studies have been reported, a systematic exploration of the relation between structural and electronic properties is missing. In this study, our main aim is to propose a new computational strategy, investigating the relationship between molecular structures and photophysical properties of a set of TADF emitters. Our analysis of different excited state descriptors, such as ΔE_{ST}, SOC, and the nature of excited states, provides a useful computational protocol to rationalize TADF efficiency. Thus, we have succeeded in clarifying the reasons behind the performance of the different descriptors which will be used to predict RISC efficiency and we have observed that SOC is, indeed, the most reliable descriptor.

References:
[1] K. Traskovskis, A. Sebris, I. Novosjolova, M. Turks, M. Guzauskas, D. Volyniuk, O. Bezvikonnyi, J. V. Grazulevicius, A. Mishnev, R. Grzibovskis and A. Vembris, *J. Mater. Chem. C*, (2021), **9**, 4532–4543.
[2] I. E. Serdiuk, M. Monka, K. Kozakiewicz, B. Liberek, P. Bojarski and S. Y. Park, *J. Phys. Chem. B*, (2021), **125**, 2696–2706.

Talk 3

Chiral knots display symmetry breaking in helical confinement

Renáta Rusková, _Dušan Račko_

Department of Molecular Simulations of Polymers, Polymer Institute, Slovak Academy of Sciences, Dúbravská cesta 9, Bratislava, Slovakia.

Topological knots naturally occur on DNA due to conformational moves or activity of molecular machines. Around half of knots are chiral including the simplest and most abundant trefoil knot. In some bacteria, occurrence of chiral knots is preferred over the achiral ones [1] and in some one form of chiral knots is more abundant [2]. Experimentally, it was successful to separate chiral knots through 2D-gel electrophoresis, where their structure was altered by negative supercoiling [2]. In general, we know that chirality plays an important role in biological systems, yet there are still mysteries to uncover, especially regarding interaction of chiral molecules towards various environments. In this work, we observe differences between right- and left-handed forms of trefoil knots and their interplay with various confinements (cylinder, toroid, helix) through molecular dynamics simulations in ESPResSo. Simulations in cylinder provided general information about behaviour of knots in confined system, their size and shape properties, which we divided into two categories – geometrical properties independent of chirality and topological parameters influenced by handedness of the knot. Toroid confinement represents a system curved in space in one dimension. Such curving, however, was not enough to enhance distinguishability of chiral knots and they had similar topological properties as in cylinder. Right-handed helix, on the other hand, strongly influenced knot topology and induced symmetry breaking for chiral knots. Right-handed knot was delocalized and compressed at its centre, while the left-handed knot expanded its main loop towards confinement wall, making it more localized, spherical and better fitting in the helix. Different topological behaviour of chiral knots as well as introduced asymmetry visible on radius of curvature indicates that helical confinement is suitable for distinguishing and direct separation of right- and left-handed forms of knots.

References:
[1] Arsuaga, J., et al., DNA knots reveal a chiral organization of DNA in phage capsids. Proceedings of the National Academy of Sciences of the United States of America, 2005. 102: p. 9165-9.
[2] Valdés, A., et al., Quantitative disclosure of DNA knot chirality by high-resolution 2D-gel electrophoresis. Nucleic Acids Research, 2019. 47(5): p. e29-e29.

Talk 4

Modelling Network Formation in Protein Hydrogels by Cluster Aggregation Kinetics

Kalila Cook[a], *David Head[b], Lorna Dougan[a]*

[a]School of Physics and Astronomy, University of Leeds, Leeds, U.K.
[b]School of Computing, University of Leeds, Leeds, U.K.

Folded protein-based hydrogels are ideal materials in medicine and biomedical engineering because of their inherent biocompatibility, but little is understood about how their bulk properties extend from the single-molecule level. We develop and employ modelling tools to examine how experimentally controllable factors, such as protein volume fraction and cross-linking rate/probability, affect network formation, structure, and ultimately bulk properties of folded protein hydrogels.

We are currently employing a combination of rheology and scattering to study model folded proteins as building blocks in chemically cross-linked hydrogels[1]. Computational modelling offers a powerful way to gain insight into the factors governing protein gelation, and using the lattice-based model we have developed, we can efficiently simulate protein clustering mechanisms that interpolate between the limiting regimes of diffusion- and reaction-limited cluster aggregation (DLCA and RLCA) and precisely identify *in situ* the point at which the network percolates. Furthermore, we have performed structural analyses on our model gel systems by considering critical percolation times, maximum gelation rates, fractal dimensions and gel pore size distributions.

By varying the system volume fraction and protein cross-linking probability, our simulations reproduce the expected transition between DLCA and RLCA regimes. In addition, we have isolated the specific region at which this transition occurs in the cross-linking probability/volume fraction parameter space. We observe that a central point, characteristic of this transition region, is linearly proportional to the volume fraction in the range applicable to protein hydrogel experiments. The fractal dimensions and pore size distributions related to these simulations also show a clear regime change. With respect to protein hydrogels, these experimentally measurable parameters[1,2] have implications for the design and synthesis of hydrogels for drug delivery and scaffolds in tissue engineering.

References:
[1] A. Aufderhorst-Roberts, M. D. G. Hughes, A. Hare, D. A. Head, N. Kapur, D. J. Brockwell, L. Dougan, *Biomacromolecules,* (2020), **21**, 4253–4260.
[2] M. D. G. Hughes, S. Cussons, N. Mahmoudi, D. J. Brockwell, L. Dougan, *Soft Matter*, (2020), **16**, 6389-6399.

Talk 5

Benchmarking state-of-the-art RNA Force Fields

Dimitrios Stamatis[a,b], *Chandra Verma*[b], *Jonathan W. Essex*[a]

[a] School of Chemistry, University of Southampton, Highfield Campus, Southampton, UK
[b] Bioinformatics Institute (BII), A*STAR, 30 Biopolis Street, Singapore

RNA is a multipotent polymer of great biological and engineering interest. Despite the progress made towards elucidation of its structure and dynamics to enable predictive modeling for industrial and other applications, many physicochemical aspects remain elusive. Modeling the structural dynamics of RNA is particularly challenging, owing to the flexible, polyanionic backbone and the many polarizable moieties. Several empirical all-atom RNA force fields (FFs) have been developed over the last two decades to address this challenge, albeit with limited success.[1] In addition, the sampling profile of such FFs has rarely been explored in detail for RNAs larger than tetranucleotides.[2] This study attempts to address this issue by benchmarking four state-of-the-art FFs against four full-sequence RNA hairpins of varying conformational stability that incorporate a wide range of structural characteristics.

Two tetraloops UUCG and AUCG (PDB IDs: 2KOC, 2Y95), and two pentaloops CUUGU and GUAAU (PDB IDs: 2L6I, 2M22) were subjected to 2 μsec Hamiltonian Replica Exchange molecular dynamics (REST2 MD) simulations in explicit solvent under four state-of-the-art FFs (Amber14SB, Amber14SB+VdW$_{phosph.}$, Amber99SB$_{Chen-Garcia}$, and Amber99SB$_{DESRES}$). Simulations starting from distinct and independent conformations generated well-converged ensembles of loop conformations, enabling the assessment of FF performance. Additionally, backbone dihedrals, sugar puckers, and contact profiles were thoroughly analyzed and compared to the available experimental data.

None of the studied FFs was found to quantitatively agree with the NMR ensembles. For the thermodynamically stable loops containing a L1-L4 base pair, the native conformation was prominently featured (largest cluster) by Amber99SB$_{DESRES}$ (UUCG) or all FFs (CUUGU). The traditionally used Amber14SB and Amber14SB+VdW$_{phosph.}$ favored the sampling of a near-native structure over all others for the less stable AUCG loop. Finally, none of the tested FFs was able to reproduce the experimental conformation of the flexible and relatively unstable GUAAU pentaloop.

This is one of the few studies that decouple the problem of FF performance from the issue of converged sampling, at least for the conformational spaces of interest. Additionally, the featured FFs are thoroughly tested in a challenging dataset of hairpins that goes beyond the commonly used hyperstable tetraloops. The favorable comparison of our results for UUCG to those of previous studies hints at the validity of our approach. Despite the general failure of the tested state-of-the-art FFs to reproduce the experimental ensembles, some might be more appropriate when non-canonical base-pairing (Amber99SB$_{DESRES}$) or base-backbone (Amber14SB) interactions are involved. Lastly, different FFs sample significantly different conformational spaces for most cases, making the choice of an appropriate RNA FF non-trivial.

References:
[1] J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter and M. Otyepka, *Chem. Rev.*, (2018) **118**, 4177–4338.
[2] C. Bergonzo, N. M. Henriksen, D. R. Roe and T. E. Cheatham, *RNA*, (2015) **21**, 1578–1590.

Talk 6

Active site sequences are enough for proteochemometric modelling of human kinases

Jannis Born[1,2], Tien Huynh[3], Astrid Stroobants[4], Wendy Cornell[3], *Matteo Manica[1]*

[1]IBM Research Europe, Zurich laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland
[2]Department of Biosystem Science & Engineering, ETH Zurich, 4058 Basel, Switzerland
[3]Department of Chemistry, Imperial College London, SW7 2AZ London, United Kingdom
[4]IBM Research, Yorktown Heights, 10598 New York, USA

Recent advances in deep learning have enabled the development of large-scale multimodal models for virtual screening and de novo molecular design. The human kinome with its abundant sequence and inhibitor data presents an attractive opportunity to develop proteochemometric models that exploit the size and internal diversity of this family of targets.

Here we challenge a standard practice in sequence-based affinity prediction models: instead of leveraging the full primary structure of proteins, each target is represented by a sequence of 29 discontiguous residues defining the ATP binding site. In kinase-ligand binding affinity prediction, our results show that the reduced active site sequence representation is not only computationally more efficient but consistently yields significantly higher performance than the full primary structure. This trend persists across different models, datasets, and performance metrics and holds true when predicting pIC50 for both unseen ligands and kinases. Our interpretability analysis reveals a potential explanation for the superiority of the active site models: Whereas only mild statistical effects about the extraction of 3D interaction sites take place in the full-sequence models, the active-site models are equipped with an implicit but strong inductive bias about 3D structure stemming from the discontiguity of the active sites. Importantly, in direct comparisons, our models perform similarly or better than previous state-of-the-art approaches in affinity prediction.

Finally, we also investigate a de novo molecular design task and find that active site sequences provide benefits in the computational efficiency, but otherwise, both kinase representations yield similar optimized affinities (for both SMILES and SELFIES-based molecular generators). In sum, our work challenges the assumption that full primary structure is indispensable for modelling human kinases.

Talk 7

Towards agreement between experiment and theory for barriers and kinetic isotope effects: addressing solvent and counterion interactions in solution.

Michele Assante,[a] Simone Tomasi,[b] *Andrew G. Leach*[c]

a. School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK
b. AstraZeneca, Silk Road Business Park, Macclesfield, SK10 2NA, UK
c. School of Health Sciences, Stopford Building, The University of Manchester, Oxford Road, Manchester M13 9PT, UK

Solvent effects have a determining role in the phenomena happening in solution. From quantum mechanics calculations to classical simulations, computational chemists share the challenge of including such effects in their models. The most popular approaches are implicit solvent models (solvent phase as continuum) and explicit solvent models (discrete solvent molecules included). Albeit a simpler description of the solvent phase, implicit solvent models are often preferred over more demanding explicit solvent models for their efficiency. Despite that, when strong interactions between solute and solvent take place, they can become rather inaccurate, particularly when partial electron sharing is involved. In such cases, a better agreement with experiment can be reached by adding several explicit solvent molecules to the solute.[1] However, anticipating the number and positions of such molecules remain challenging. The system gets even more complicated when counter-ions are present, as they can potentially interact with both solvent and solutes as well.

Here, we present a model to tackle the above-mentioned challenges based on the electrostatic potential surface of solutes and solvents. Electrostatic potential surfaces of solutes can be exploited to identify and quantify likely sites of solute-solvent interaction.[2] This should be used to implement a correction to implicit solvent model-derived free energies, as well as providing a guide for the rational inclusion of explicit solvent molecules and counter-ions.

The said model has been developed for the prediction of logP and pK_as. Its performance was tested on relevant case studies: the prediction of kinetic isotope effects for the hydrolysis of phosphate monoester in water and for the Morita-Baylis-Hillman reaction in methanol, as well as reactions of boron-containing molecules.

References:
[1] J. R. Pliego Jr and J. M. Riveros, WIREs Computational Molecular Science, 2020, **10**, e1440.
[2] C. S. Calero, J. Farwer, E. J. Gardiner, C. A. Hunter, M. Mackey, S. Scuderi, S. Thompson and J. G. Vinter, Phys. Chem. Chem. Phys., (2013), 15, 18262-18273.

Talk 8

A data-driven approach to relative Free Energy Perturbation reliability predictions for alchemical free energy calculations in drug design

Jenke Scheen[a], Mark Mackey[b], *Julien Michel[a]*

[a]EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, United Kingdom
[b] Cresset Group, New Cambridge House, Bassingbourn Road, Litlington, Cambridgeshire, SG8 0SS, United Kingdom

Early-stage drug discovery campaigns pose major challenges to the pharmaceutical industry, mostly owing to large costs and high rate of failure of high-throughput screening of candidate molecules to the therapeutic target in question. Alchemical Free Energy (AFE) calculations have increasingly gained traction in solving the ligand-optimisation problem in both academic and corporate drug discovery. Free Energy Perturbation (FEP) - where candidate molecules are transformed into each other virtually to compute the relative free energy of binding - has been one of the most tractable AFE techniques[1] and is being offered by major commercial software suppliers such as *Cresset* and *Schrödinger*.

A critical step in FEP workflows is the generation of effective perturbation networks to ensure transformations are being simulated with sufficient phase-space overlap to maximise prediction reliability. Currently, state-of-the-art softwares use primarily molecular similarity based metrics to estimate optimal edges in networks. This method often still requires the user to review and tweak the presented perturbation network, hindering development toward a fully automated FEP workflow.

The current study uses a machine-learning (ML) approach to train models on a large number of solvation FEP simulations to predict the reliability of a given perturbation before simulating it. Such models can replace similarity-based metrics to plan perturbation networks. Additionally, we generate FEP-space: given all possible perturbations in all available FEP benchmarking sets, we have grafted these transformations onto a common scaffold such that we can generate a training set that encompasses the complete domain of FEP. Using this training set (n~2800), machine-learning models were built that outperformed cutting-edge network generators.

This open-source project has resulted in several advancements that further progress the field of FEP. Using the data-driven method, practicioners can generate more reliable FEP networks, are able to transfer-learn to alternative FEP software reliabilities and fine-tune the model to the molecule series that is being investigated. FEP-Space has been made publicly available to allow researchers to create alternative novel models in the context of FEP - outcomes of this work are projected to drive the field forward in aid of advancing drug discovery.

References:
[1] A. Mey, J. Scheen et al, *LiveCoMS*, (2019), 2, [Article v1.0].

Solvent effects on the photochemical decarbonylation of cyclopropenones

Daniel M. Adrion, *Steven A. Lopez*

Northeastern University, Department of Chemistry and Chemical Biology, Boston, Massachusetts, 02115, United States

Photochemical reactions are a powerful way to access high-energy and strained compounds while allowing for mild conditions and greener reaction pathways. This benefit is especially relevant in photoclick chemistry, where light generates the reagents to promote a 1,3-dipolar cycloaddition reaction between an alkyne and azide. This reaction is known as a strain-promoted azide-alkyne cycloaddition (SPAAC). A highly strained alkyne can be generated photochemically by irradiating an eight-membered ring with a cyclopropenone photo protecting group. This reaction generates a strained cyclooctyne (COT) and carbon monoxide. The photochemical mechanism is known for unsubstituted cyclopropenone, but not for any derivatives. It is also not known how the presence of alkyl groups or strained rings change the photochemical mechanism. To date, there has been one computational study done on cyclopropenone to investigate the dissociation into carbon monoxide and acetylene. This study used ab initio molecular spawning (AIMS) calculations to investigate the dissociation mechanism along the excited state. We have built upon this study to investigate how the mechanism is affected by methyl groups and an eight-membered ring and by using an explicit solvent model to determine the mechanism in water. We have used multiconfigurational calculations (CASSCF) with a (10,9) active space and the ANO-S-VDZP basis set to understand the photophysics and photochemistry of the decarbonylation reaction of unsubstituted cyclopropenone, dimethyl cyclopropenone, and cyclopropenone bonded to a cyclooctene ring (the COT-precursor). We found an allowed transition to the S_1 state through an $n \rightarrow \pi^*$ excitation for all three molecules. The excitation energies for the $S_0 \rightarrow S_1$ transitions are 4.61 eV (cyclopropepone), 5.13 eV (COT-precursor), and 5.13 eV (dimethyl cyclopropepone). We used fewest switches surface hopping to study the photochemical dynamics along the excited state and found that decarbonylation can occur for all three molecules along the S_1 surface. The dimethyl cyclopropenone dissociates at the highest rate (68%), followed by the COT-precursor (53%) and the parent structure (28%). We will also discuss the effects of water solvation on decarbonylation reactions.

Talk 10

Investigating molecular mechanisms of antibiotic permeation through outer membrane porins in high dimensional conformational space

Nandan Haloi, Archit Kumar Vasan, Emily Geddes, Rebecca Joy Ulrich, Arjun Prasanna, Mary Elizabeth Metcalf, Po-Chao Wen, William W. Metcalf, Diwakar Shukla, Paul Hergenrother, and *Emad Tajkhorshid*

NIH Center for Macromolecular Modeling and Bioinformatics, Beckman Institute for Advanced Science and Technology, Center for Biophysics and Quantitative Biology, Department of Biochemistry, Department of Microbiology, Department of Chemistry, Carl R. Woese Institute for Genomic Biology, and Cancer Center at Illinois,
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Antibiotic resistance of Gram-negative bacteria is largely attributed to the low permeability of their outer membrane (OM). Effective antibiotics typically permeate OM porins; thus, understanding porin permeation mechanisms would aid antibiotic development. Although molecular dynamics (MD) simulations can provide key structural information on molecular processes, simulating OM porin permeation is challenging due to its high dimensionality. Since OM porins spontaneously transition between open and closed states, an understanding of (1) porin functional states regulating permeation[1] and (2) antibiotic permeation pathways through porins is necessary[2]. To investigate the first aspect, we sampled the conformational landscape of the apo porin using MD simulations and Markov state models[1]. We found large-scale motion of the internal loop to opposing sides of the porin regulates transition between energetically stable open and closed functional states. Furthermore, mutations of key residues involved in the transition alter the dynamic equilibrium of the porin, as found in our MD simulations, and regulate antibiotic permeation inside the cell, as observed in our whole-cell accumulation assays. The identified open state of the porin was then used to investigate the high OM porin permeability of primary amine-appended antibiotics[2]. We compared the permeation of aminated and amine-free antibiotic derivatives by incorporating MD simulation with our novel, Monte Carlo and graph theory-based algorithm designed to improve sampling of conformationally flexible drugs. We found that the primary amine facilitates permeation by enabling the antibiotic to align its dipole to the luminal electric field of the porin and form favorable electrostatic interactions with highly-conserved charged residues of the internal loop. The importance of these interactions was validated with experimental mutagenesis and whole cell accumulation assays. Overall, our study demonstrating the regulation of antibiotic permeation by porin dynamics and direct antibiotic-protein interactions could help design effective antibiotics against resistant bacteria.

References:
[1] A. K. Vasan∗, N. Haloi∗, P. C. Wen, R. J. Ulrich, M. E. Metcalf, W. W. Metcalf, P. Hergenrother, D. Shukla, and E. Tajkhorshid, *PNAS* (2021), Under revision.
bioRxiv: https://www.biorxiv.org/content/10.1101/2021.09.09.459668v1.
[2] N. Haloi∗, A. K. Vasan∗, E. J. Geddes, A. Prasanna, P. C. Wen, W. W. Metcalf, P. J. Hergenrother, E. Tajkhorshid, *Chem. Sci.* (2021), **12**, 15028–15044.

Talk 11

Modelling Organic-Inorganic Interfaces of Urinary Calculi

Rhiannon Morris[1], Helen Chappell[1], Andrew Scott[2], Antonia Borissova[2] and James Smith[1]

[1] *School of Food Science and Nutrition, University of Leeds, Leeds, LS2 9JT, UK*
[2] *School of Chemical and Process Engineering, University of Leeds, LS2 9JT, UK*

Urinary calculi (kidney stones) is a common ailment effecting around 10% of the world's population and resulting in more than 97,000 finished consultant episodes (FEC) each year in the UK alone [1]. However, the key chemical interactions relating to stone crystallisation and aggregation are not fully understood. Urinary calculi are solid clusters of small stones, composed of crystals that have precipitated from urine, and built up on the luminal surface of the endothelial cell surfaces of the microtubule in the kidney. There are three main types of kidney stone: calcium, struvite and uric acid. Moreover, kidney stones are often found complexed to organic matrices, such as proteins and amino acids. The fundamental chemistry underlying this observation is unknown. This research uses first principles modelling (CASTEP) to help elucidate the crystallisation phenomena, and unravel the prevalent chemistry behind stone composition. Urine analysis of kidney stone patients has previously revealed that their urine contains higher amounts of phospholipids in comparison to healthy patients [2]. In order to begin to understand the nucleation process, we have constructed surface models of calcium oxalate monohydrate and calcium oxalate dihydrate, and modelled stone growth, by simulating further calcium oxalate adsorption onto these surfaces. To investigate the interactions between urinary macromolecules and the growing crystal surfaces at an atomic level, we have performed *ab initio* molecular dynamics of phosphocholine adsorption on calcium oxalate surfaces. We have shown that the phosphocholine head groups become entrapped within the growing crystal and the lowest energy structures are the ones were the calcium oxalate dihydrate surfaces have attempted to redefine their crystallographic structure. This could be because it is the crystallographic waters that are driving the encapsulation of the phosphocholine head group.

References:
[1] C.Y. Pak, *The lancet,* 1998, **351**, 1797-1801
[2] S. R. Khan, P. A. Glenton and D. R. Talham, *Kidney international*, 2002, **62**, 2062–2072.

Talk 12

Investigating the activity and inhibition mechanisms of SARS-CoV-2 M^{pro} using molecular simulations

H. T. Henry Chan, *Christopher J. Schofield*, *Fernanda Duarte*

Chemistry Research Laboratory, Department of Chemistry, 12 Mansfield Road, Oxford, OX1 3TA, U.K.

The SARS-CoV-2 coronavirus main protease (M^{pro}) is an attractive target for anti-COVID-19 drug development, given its essential role in viral replication and the lack of similar human cysteine proteases. However, there remain questions on the exact mechanism of how M^{pro} recognises and catalyses amide hydrolysis specifically at its 11 designated cleavage sites along the viral polyprotein chain. Based on molecular dynamics (MD) simulations of M^{pro} complexed with each of its 11 native substrates, each modelled as an 11-mer peptide, we identified a series of conserved protein-peptide hydrogen bonds and hydrophobic interactions that are key to substrate binding, particularly on the N-terminal side of the scissile amide.[1] Using complementary equilibrium and nonequilibrium MD simulations in the Kubo-Onsager approach, we obtained insights into the dynamic behaviour of the protein upon substrate binding. The catalytic mechanism was studied by employing quantum mechanics/molecular mechanics (QM/MM) methods.

The enzyme-substrate interactions identified from MD simulations provide interesting comparison with the binding modes of experimentally validated M^{pro} inhibitors, including a range of β-lactam-containing small molecule ligands,[2] as well as peptide inhibitors that were designed through *in silico* saturation mutagenesis.[1] A combination of computational methods including docking, MD, and QM/MM modelling enable prediction of the complexed conformations of these inhibitors and thus the expansion of opportunities for further inhibitor design and optimisation.

References:
[1] H. T. H. Chan, M. A. Moesser, R. K. Walters, T. R. Malla, R. M. Twidale, T. John, H. M. Deeks, T. Johnston-Wood, V. Mikhailov, R. B. Sessions, W. Dawson, E. Salah, P. Lukacik, C. Strain-Damerell, C. D. Owen, T. Nakajima, K. Świderek, A. Lodola, V. Moliner, D. R. Glowacki, J. Spencer, M. A. Walsh, C. J. Schofield, L. Genovese, D. K. Shoemark, A. J. Mulholland, F. Duarte and G. M. Morris, *Chem. Sci.*, 2021, **12**, 13686-13703.
[2] T. R. Malla, A. Tumber, T. John, L. Brewitz, C. Strain-Damerell, C. D. Owen, P. Lukacik, H. T. H. Chan, P. Maheswaran, E. Salah, F. Duarte, H. Yang, Z. Rao, M. A. Walsh and C. J. Schofield, *Chem. Commun.*, 2021, **57**, 1430-1433.

Programme of Poster Presentations

1	**Predicting protein-membrane interfaces of peripheral membrane proteins using ensemble machine learning** Alexios Chatzigoulas, Biomedical Research Foundation Academy of Athens
2	**Studying TRIM PHD-bromodomain interactions with histone H3 peptides using molecular dynamics simulations** Bernadette Lee, University of Oxford
3	**Multiscale Theoretical Investigation of the Second Harmonic Generation of Di-8-ANEPPS in (Un)Saturated Lipid Bilayers: Effect of the Cholesterol Content** Charlotte Bouquiaux, Université de Namur
4	**MoleculeACE: a benchmark for machine learning with activity cliffs** Derek van Tilborg, Eindhoven University of Technology
5	***In silico* modelling of cardiac myosin modulators** Fariha Akter, Queen Mary University of London
6	**Molecular modelling of biochars** Rosie Wood, University of Edinburgh
7	**Investigating Tetrel-Based Neutral Frustrated Lewis Pairs for Hydrogen Activation and Catalytic Hydrogenation** Pallavi Sarkar, JNCASR, India
8	**Investigating T-cell Receptor Selectivity using Molecular Dynamics Simulations and MMPBSA Calculations** Annabelle Hartt, University of Bristol
9	**Creating coarse grain parameters for lipids from the inner membrane of Mycobacterium tuberculosis** Chelsea Brown, Warwick University
10	**Machine learning for predicting regiocontrol in CH functionalisation** Joe Heeley, University of Bristol
11	**Investigating the molecular mechanism of transport proteins using HDX-MS and MD simulations** Ruyu Jia, King's College London
12	**Mechanistic insights into the functional dynamics of L1 β-lactamase** Zhuoran Zhou, University College London

Posters will be available online on the YMF 2021-22 website from 9th- 16th February. The password and link will be sent to all attendees using their registered email address.

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

11:00-11:30	BREAK and POSTER SESSION 11:05-11:20 Posters 1-12 available in Breakout Rooms P1-P12
12:30-13:50	LUNCH and POSTER SESSION 12:30- 13:10 Posters 1-6 available in Breakout Rooms P1-P6 13:10- 13:50 Posters 7-12 available in Breakout Rooms P7-P12
14:50-15:20	BREAK and POSTER SESSION 14:55-15:10 Posters 1-12 available in Breakout Rooms P1-P12

Poster 1

Predicting protein-membrane interfaces of peripheral membrane proteins using ensemble machine learning

Alexios Chatzigoulas[a,b] and *Zoe Cournia*[a,b]

[a]Biomedical Research Foundation, Academy of Athens, 4 Soranou Ephessiou, 11527 Athens, Greece
[b]Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 15784 Athens, Greece

Abnormal protein-membrane attachment of peripheral membrane proteins is involved in deregulated cellular pathways and in disease. Therefore, the possibility to modulate protein-membrane interactions represents a new promising therapeutic strategy for peripheral membrane proteins that have been considered so far undruggable [1]. A major obstacle in a drug design strategy targeting the protein-membrane interface is that the membrane binding domains of peripheral membrane proteins are usually unknown. Thus, the development of fast and efficient algorithms predicting the protein-membrane interface would identify these regions and shed light into the accessibility of membrane-protein interfaces by drug-like molecules. In this talk, we will present a complete drug design methodology for drugging protein-membrane interfaces using DREAMM (Drugging pRotein mEmbrAne Machine learning Method), a fast and robust predictor of protein-membrane interfaces of peripheral membrane proteins using machine learning, which is implemented as a web server [2]. We have devised an ensemble machine learning algorithm for identifying protein-membrane interfaces utilizing experimental data available in the literature for training 21 machine learning classifiers and a voting classifier for predicting membrane-penetrating residues [3]. Evaluation of the voting classifier performance yields a macro-averaged F_1 score=0.93 and an MCC score=0.87 for predicting correctly membrane penetrating residues on unknown proteins of an independent test set. Additionally, DREAMM also identifies binding pockets in the vicinity of the predicted membrane-penetrating residues using protein conformational ensembles provided by the user or generated by DREAMM, if the user does not have access to different protein conformations. Depending on the size of the protein, the calculation may last from a few minutes up to a few hours. A unique URL is provided to the user allowing them to bookmark and access the results at a later time. Results are visualized and can be downloaded from the web-server. The web-server is accessible via https://dreamm.ni4os.eu.

References:
[1] Z. Cournia and A. Chatzigoulas, *Curr. Opin. Struct. Biol.,* (2020), **62**, 197-204.
[2] A. Chatzigoulas and Z. Cournia, *Bioinformatics,* (2021), submitted.
[3] A. Chatzigoulas and Z. Cournia, *Brief. Bioinform.,* (2021), in proofs.

Poster 2

Studying TRIM PHD-bromodomain interactions with histone H3 peptides using molecular dynamics simulations

Bernadette Lee, Jessica K. Reynolds, Michael A. Platt, *Stuart J. Conway, Fernanda Duarte*

Department of Chemistry, University of Oxford, 12 Mansfield Road, Oxford, OX1 3TA

The tripartite motif (TRIM) protein family comprises over 70 proteins that often possess E3 ubiquitin ligase activity. Dysregulation of TRIM proteins has been linked to diseases such as breast and liver cancer, thus studying their function is of interest.[1] To target subsets of TRIM proteins, differences in their C-terminal domains can be exploited. For example, TRIM24 and TRIM33 (α and β isoforms) contain a C-terminal PHD finger-bromodomain (PHD-BRD) that could be targeted by chemical probes.[2] To aid development of ligands for these domains, molecular dynamics (MD) simulations were used to characterise interactions between the TRIM PHD-BRDs and a histone H3-mimicking peptide, to identify native protein–peptide contacts that could be disrupted by ligands.

Conventional MD simulations were performed on models of TRIM24, TRIM33α and TRIM33β bound to a $H3_{1-20}K9Me_3K18Ac$ ($K9Me_3$: trimethylated lysine 9, K18Ac: acetylated lysine 18). At the PHD finger, we found that $K9Me_3$ forms a cation–π interaction with a solvent-exposed tryptophan. We hypothesise that targeting this interaction could help with designing specific ligands for TRIM PHD fingers.

Interactions between the TRIM BRDs and peptide were also examined. In TRIM33α, a non-canonical BRD, we observed that K18Ac is not recognised *via* hydrogen bonding to a conserved asparagine. Instead, K18Ac forms hydrophobic contacts with the BRD pocket. Conversely, with TRIM24 and TRIM33β, we observed hydrogen bonding between K18Ac and the conserved asparagine. This may have implications for the design of ligands to differentiate between these BRDs.

Subsequently, the peptide unbinding process was studied using peptide Gaussian accelerated MD simulations. This highlighted the importance of N-terminal histone H3 residues in peptide binding to the PHD finger. Taken together, this work has helped inform the design of peptidomimetic ligands for TRIM PHD-BRDs, which are currently being synthesised.

References:
1. S. Hatakeyama, *Nat. Rev. Cancer*, 2011, **11**, 792–804.
2. F. D'Amico, R. Mukhopadhyay, H. Ovaa and M. P. C. Mulder, *ChemBioChem*, 2021, **22**, 2011–2031.

Poster 3

Multiscale Theoretical Investigation of the Second Harmonic Generation of Di-8-ANEPPS in (Un)Saturated Lipid Bilayers: Effect of the Cholesterol Content

Charlotte Bouquiaux[a], Frédéric Castet[b] and *Benoît Champage[a]*

[a]University of Namur, Theoretical Chemistry Lab, Unit of Theoretical and Structural Physical Chemistry, Namur Institute of Structured Matter, rue de Bruxelles, 61, B-5000 Namur (Belgium)
[b]Institut des Sciences Moléculaires, University of Bordeaux, UMR 5255 CNRS, cours de la Libération 351, F-33405 Talence Cedex (France)

The cell membranes perform diverse and essential functions that are controlled by the enormous lipid structure diversity. Among the different lipid classes, cholesterol has a strong influence on maintaining the correct fluidity and rigidity of the animals cell membranes, and so their functions. Therefore, having a tool to distinguish between membranes of various cholesterol content is helpful in the deeper understanding of lipid bilayers. In this work [1], we investigate three glycerophospholipid phosphatidylcholine (PC) lipids, namely dipalmitoylphosphatidylcholine (DPPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), and dioleoylphosphatidylcholine (DOPC) with different levels of cholesterol by analysing the second-harmonic generation (SHG) nonlinear optical (NLO) response of a probe molecule, di-8-ANEPPS inserted into the membranes. This technique as the advantage to be specific to interfacial region, like lipid bilayer, and used in conjunction with an ANEPP-like molecule, allows us rapid acquisition at relatively low laser power. Six independent 400ns Molecular Dynamics simulations were performed on both DPPC and DOPC lipids by varying the cholesterol mole fraction (from 0 to 0.66), and one on the pure POPC system, giving thirteen systems total. The effect of the cholesterol on the properties of the bilayer are studied, namely the thickness, the area per lipid, the hydrocarbon parameter and also the orientation of the diverse molecules within the membranes. All the analysis of the structural parameters of the bilayers studied converge toward one conclusion: as the mole fraction of cholesterol increases, the systems are more and more rigid. This is known as the condensing effect of cholesterol. The structural analyses are then confronted to the molecular NLO response, β, computed at the TDDFT/M06-2X/6-311+G* level, and in particular the contribution to beta parallel to the bilayer normal, β_{ZZZ}. Using the same methodology, the effect of the presence of (un)saturations(s) was also studied and highlighted the increase of β_{ZZZ} going from saturated to unsaturated lipids. This computational approach provides insights onto the link between the cholesterol content/presence of (un)saturation(s) and the diagonal component β_{ZZZ} of the first hyperpolarizability and so a first approach towards the unravelling of the changes due to the cholesterol content/presence of (un)saturation(s) of lipid bilayers.

References:
[1] C. Bouquiaux, F. Castet, and B. Champagne *J. Phys. Chem. B,* (2021), **125**, 10195-10212.

Poster 4

MoleculeACE: a benchmark for machine learning with activity cliffs

Derek van Tilborg[1], Alisa Alenicheva[2], and *Francesca Grisoni*[1]

[1]Eindhoven University of Technology, Dept. Biomedical Engineering, Institute for Complex Molecular Systems, Eindhoven, Netherlands.
[2]JetBrains Research. Saint Petersburg, Russia

Machine learning for quantitative-structure activity relationship (QSAR) is a crucial tool in numerous fields of chemistry and drug discovery. QSAR has reached high levels of accuracy thanks to advances in machine learning. However, remarkable mispredictions of activity still arise among similar molecules.[1] This aspect is related to the presence of the so-called activity cliffs, which arise when two compounds that are structurally similar exhibit a large difference in bioactivity on the same target[2]. Activity cliffs are of high interest in medicinal chemistry, especially during hit-to-lead optimization[1]. However, machine learning models are often ill-equipped to predict large SAR discontinuities.

Still, the ability of a model to perform well in the presence of activity cliffs is a testament to the model's 'understanding' of the underlying structure-activity relationship. Despite the importance of activity cliffs in drug discovery, it is unclear which machine learning strategies are best suited for handling complex structure-activity landscapes. Moreover, there is currently no well-established method for measuring activity cliff predictability.

Aiming to advance the capabilities of machine learning in QSAR, we developed a benchmark method called **MoleculeACE** (Activity Cliff Estimation). MoleculeACE is a modular and open-access resource, which can be used to (1) analyze and compare the performance on activity cliffs of machine and deep learning methods typically employed in QSAR, (2) identify best practices to enhance a model's predictivity in the presence of activity cliffs, and (3) design guidelines to consider when developing novel QSAR approaches. To this end, we collected and curated bioactivity data on 30 macromolecular targets, which were used to evaluate the performance of several machine learning algorithms, combined with common molecular descriptors, as well as neural networks based on molecular graphs, and recurrent neural networks based on SMILES strings.

Our results underscore the importance of molecular representation or molecular description in predicting activity cliffs compared to the used machine learning algorithm. MoleculeACE provides useful insights in QSAR modelling and we expect it to steer the efforts of the chemoinformatics community towards increased attention to the relevant issue of activity cliffs.

References:
[1] Stumpfe D, Hu H, Bajorath J. *ACS Omega*, (2019) **4**, 14360–14368.
[2] Maggiora GM. *J Chem Inf Model*, (2006) **46**, 1535.

Poster 5

In silico modelling of cardiac myosin modulators

Fariha Akter[a], Julien Ochala[b,c] and Arianna Fornili [a]

a. School of Physical and Chemical Sciences, Queen Mary University of London, Mile End Rd, Bethnal Green, London E1 4NS.
b. Centre for Human & Applied Physiological Sciences, King's College London, Great Maze Pond, London SE1 9RT
c. Department of Biomedical Sciences, University of Copenhagen, Blegdamsvej 3B, København N, DK-2200

Heart disease is the leading cause of mortality worldwide, despite medical advancements. Recent research has indicated that new promising treatments can be developed by designing drugs that can directly bind the key cardiac motor protein myosin II and modulate its contractile activity in diseases such as hypertrophic and dilated cardiomyopathy[1]. Myosin can adopt two important conformations during its functional cycle: pre-powerstroke (PPS) and post rigor (PR). Compounds that can selectively stabilise these states are expected to have different effects on myocardial contractility[2]. This research aims to rationally design new conformation-selective compounds of human β-cardiac myosin.

Various PR and PPS-selective ligands were identified using virtual screening, molecular docking, and molecular dynamics (MD) simulations. Hydrogen bond and contact map analyses showed that selectivity towards a specific conformation was related to a higher number of protein-ligand interactions. MD simulations confirmed a stabler binding to the expected state for each ligand. Moreover, new possible target conformations were identified with enhanced-sampling MD simulations (steered MD and umbrella sampling) of the PR-PPS transition (recovery stroke). Pocket tracking analyses revealed that the binding site is very dynamic with significant changes in physicochemical properties such as size and polarity. These findings will be used to guide the design of drugs that can selectively stabilise intermediate states in the myosin recovery stroke.

The best candidates will be tested *in vitro* on single cardiac myosin and on myocardial fibers to measure their effect on myosin force, velocity and power and determine the type of activity (activation or inhibition) of each ligand.

References:
1 F. I. Malik et al, *Science*, 2011, **331**, 1439–1443.
2 S. Hashem, W. G. Davies and A. Fornili, *J. Chem. Inf. Model.*, 2020, **60**, 6438–6446.

Poster 6

Molecular modelling of biochars

Rosie Wood [a], *Ondrej Masek [b], Valentina Erastova [a]*

[a] School of Chemistry, University of Edinburgh, Edinburgh, UK
[b] School of Geosciences, University of Edinburgh, Edinburgh, UK

Biochars are black carbonaceous solids produced through the pyrolysis of biomass under conditions of little or no oxygen. For centuries, biochar-like materials have been used within agriculture as soil amendments; however, more recently, these materials have found use as economical and environmentally friendly adsorbents with applications in water filtration and purification. This has led to an interest in their adsorption properties and a volume of experimental studies focused on optimisation of these materials for such applications. To this end, theoretical approaches can be advantageous. Molecular modelling, for example, can bring atomic-level insights into the adsorption process and provides an excellent platform upon which structure-property relationships can be investigated in a systematic and reproducible manner. Nevertheless, very few realistic molecular models of biochars exist within literature and only a small number of molecular modelling studies of biochars have been conducted. In an effort to remediate this, within this work, we have used molecular dynamics simulations to develop a set of three molecular models representative of biochars produced at low (400ºC), medium (600ºC) and high (800ºC) highest treatment temperatures. Our models replicate a wide variety of both chemical and physical properties of such biochars leading us to believe they are descriptive of reality. We hope that in developing these models we will facilitate the uptake of molecular modelling in the study of biochars and accelerate research into these materials for use as environmentally-friendly adsorbents.

Poster 7

Investigating Tetrel-Based Neutral Frustrated Lewis Pairs for Hydrogen Activation and Catalytic Hydrogenation

Pallavi Sarkar[a], Shubhajit Das[a,b] and *Swapan K. Pati[a]*

[a]Theoretical Sciences Unit, School of Advanced Materials (SAMat), Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, 560064, India.
[b]Present Address: Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fedé rale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

Frustrated Lewis pair (FLP) chemistry has gained significant attention in the past decade as the main group hydrogenation catalysts. The majority of the reported FLP catalysts utilize either boron-containing Lewis acids (LAs) or alternatively isolobal and iso(valence)electronic cationic tetrylium Lewis acids for hydrogenation. Neutral tetrel Lewis acids allude as less trivial options due to the absence of a formally empty p orbital on the acceptor atom. Recently, a series of intramolecular geminal FLPs $(C_2F_5)_3E$-CH_2-$P(tBu)_2$ (E = Si, Ge, Sn) featuring neutral sp^3 tetrel atoms as acceptor sites has been reported for activation of small molecules including H_2. In this work, through density functional theory computations, we elucidate the general mechanistic picture of H_2 activation by this family of FLPs. Our findings reveal that the sp^3 acceptor centre derives the required Lewis acidity utilizing the antibonding orbitals of its adjacent bonds with the individual contributions depending on the identity of the acceptor and the donor atoms. By varying the identity of the Lewis acid and Lewis base sites and attached substituents, we unravel their interplay on the energetics of the H_2 activation. We find that switching the donor centre from P to N significantly affects the synchronous nature of the bond breaking/formations along the reaction pathway, and as a result, N-bearing FLPs have a more favourable H_2 activation profile than those with P. The reactivity trends are rationalized in detail within the framework of the activation–strain model of reactivity along with the energy-decomposition analysis method. In addition, we have also examined their ability to perform catalytic hydrogenation using CO_2, carbonyls, and imines as substrates. Our mechanistic analyses reveal that the transfer of the activated hydrogens from FLP to the unsaturated bond follows a concerted yet asynchronous pathway. The sequence of proton/hydride transfer is largely influenced by the donor/acceptor centres of the FLP as well as the corresponding unsaturated functional group. We find that by varying the donor centres in these FLPs, the dihydrogen release barrier can be tuned. Finally, two major side reactions pertinent to the performance of the hydrogenation cycle, such as dimer formation of the catalyst and decomposition through reductive elimination, are also investigated in detail.

References:
1. P. Sarkar, S. Das and S. K. Pati, *Inorg. Chem.*, (2021) **60**, 20, 15180–15189.
2. P. Sarkar, S. Das and S. K. Pati, J. *Phys. Chem. C*, (2021) **125**, 41, 22522–22530.

Poster 8

Investigating T-cell Receptor Selectivity using Molecular Dynamics Simulations and MMPBSA Calculations

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

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

Cell-mediated adaptive immunity enables the surveillance of human cells, detecting when foreign or damaged proteins are present inside. Human leukocyte antigen (HLA) proteins can present short peptides made from degraded intracellular proteins on the cell surface, so that T-cell receptors (TCRs) can bind to non-self peptide-HLA molecules. This TCR-pHLA binding can lead to T-cell signalling events that eventually induce death of the antigen presenting cell. Over the past few decades, TCRs have been successfully developed to treat diseases. However, to use soluble TCRs as therapeutics, high affinity and low cross-reactivity are required (i.e. prevent binding to peptides that are not the intended target)[1].

Immunocore have identified a high affinity TCR that selectively binds to a KRAS protein peptide carrying an oncogenic mutation (G12D) over the wild-type peptide. The G12D mutation results in a constitutively active KRAS, which can cause mis-regulation of cell proliferation and therefore cancer. TCRs selectively binding to the G12D KRAS peptide thus have promise for cancer therapeutics.

Crystal structures of the two TCR-pHLA complexes showed minimal differences and thus could not explain the mechanism for selectivity. We thus performed molecular dynamics simulations and MMPBSA calculations to identify key residues with altered contributions to the binding affinity. The main changes identified were in positively charged residues in the TCR, which suggested a change in electrostatics within the pHLA. Electrostatic potential calculations indicated that the main reason for selectivity was a shift in electrostatics within the mutant KRAS peptide complex, allowing improved interactions with the TCR compared to the wild-type KRAS peptide.

As high affinity is essential for the therapeutic applications of soluble TCRs, understanding this mechanism for selectivity may enable *in silico* design guided by specificity enhancement, particularly for cases where peptides contain charged residues. We are now in the process of testing *in silico* binding and selectivity assessment based on MMPBSA as well as higher-throughput methods on larger sets of TCR variants for specific pHLA targets.

References:
1   Robinson, R. A., McMurran, C., McCully, M. L. & Cole, D. K. Engineering soluble T-cell receptors for therapy. *FEBS J* **288**, 6159-6173, doi:10.1111/febs.15780 (2021).

Poster 9

Creating coarse grain parameters for lipids from the inner membrane of *Mycobacterium tuberculosis*

Chelsea M. Brown[a], Ya Gao[b], Yeol Kyo[b], Mattheiu Chavent[c], Wonpil Im[b], *Elizabeth Fullam[a], Phillip Stansfeld[a]*

[a] School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK
[b] Department of Bioengineering, Lehigh University, Bethlehem, USA
[c] Biological Integrative NMR team, IPBS, Toulouse, France

Mycobacterium tuberculosis (*Mtb*) is the causative agent of Tuberculosis (TB) infections, one of the world's leading infectious killers. In 2020, there were 9.9 million new infections and 1.5 million deaths[1]. This is despite the fact that both a vaccine and treatment have been available for decades. Global infection rates remain high for a variety of complex reasons; however it is obvious that the treatment course needs improvement. The current regime for drug susceptible cases is 6 months long using four antibiotics given in combination. Treatment for drug resistant TB infections can take up to two years to treat and is 30% less successful. Intensive research into new treatments have been prompted by the World Health Organisation's yearly reports on global TB incidence which have been published since 1997, but no breakthrough treatments have been implemented so far.

One of the reasons for the slow progression is that the *Mtb* cell wall is fiendishly complicated, making it difficult for compounds to pass into the cell. Targeting membrane proteins reduces the number of barriers that need to be crossed for successful treatment. A second reason for slow progress on this problem is that experimental work with *Mtb* is slow, expensive and highly specialised. Using modelling to complement experimental findings can reduce the time and resources required to investigate promising targets for the treatment of *Mtb* infections.

Lipids can play an important part in the mechanism of transport for membrane proteins. Therefore, modelling known lipids for the inner membrane of *Mtb* is vital for accurate simulations of these important targets. The composition of the inner membrane of *Mtb* is 70% mannosylated phosphatidylinositol lipids (PIMs), 10% cardiolipin (CL), 5% phosphatidylethanolamine (PE), 3% phosphatidylinositol (PI), 3% trehalose monomycolate (TMM) and 9% aploar lipids[2]. Both atomistic and coarse-grained (CG) parameters exist for CL, PE, and PI but PIMs and TMM are not yet defined. Herein we have defined CG parameters for PIM lipids, the main constituent of IM, in the Martini3 forcefield and compared with atomistic parameters developed by the Im lab. The results of this work permit a multi-scale computational approach to accurately study *Mtb* membrane proteins and reveal potential targets for future treatments. This will be crucial to aid the development of novel antimicrobials, whilst also characterising transport pathways into the cell.

References:
1. WHO, *Global Tuberculosis Report 2021*, World Health Organization, Geneva, 2021.
2. R. Bansal-Mutalik and H. Nikaido, *Proceedings of the National Academy of Sciences*, 2014, **111**, 4958-4963.

Machine learning for predicting regiocontrol in CH functionalisation

Joseph Heeley, *Robin Bedford, Craig Butts, Natalie Fey*

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

In recent years, direct transition metal-catalysed CH bond functionalisation has emerged as a powerful method for the diversification of medicinally relevant compounds. These reactions offer numerous advantages over their traditional cross-coupling counterparts, namely as greener alternatives with higher atom and step economy. Despite this, they are not currently widely used in industry; their inherent lack of selectivity (thanks in part to the ubiquity of the CH group in organic chemistry) leads to unreliable and often promiscuous reactivity.

However, it is possible to discriminate between CH bonds *via* exploitation of a heterocycle's innate properties. Bedford *et al.* have shown that the most acidic and nucleophilic sites of pyrazolo[1,5-a]pyrimidines can be independently reacted by switching the mechanism of CH activation, achieving complete regiocontrol.[1] An added benefit of this methodology is it's predictive capability: these sites can be identified using DFT, demonstrating an impressive synergy between computation and experiment.

Attention has now turned to validating this approach across the breadth of heterocyclic chemical space, generating a wealth of experimental and computational data. This context is ideal for the application of machine learning for the prediction of these heterocyclic properties, reducing computational time from hours to seconds. The Butts group's 'IMPRESSION' algorithm[2] has proven to be capable of making accurate predictions within well-defined chemical space (MAE = 2.32 and 1.98 for acidities and nucleophilicities respectively). Quantitative reactivity prediction is also possible by defining relative metrics, which show good agreement to experimental observations.

References:
[1]   R. B. Bedford, S. J. Durrant, M. Montgomery, Catalyst-Switchable Regiocontrol in the Direct Arylation of Remote C-H Groups in Pyrazolo[1,5-a]pyrimidines. *Angew. Chemie Int. Ed.* (2015) **54**, 8787–8790
[2]   W. Gerrard, L. A. Bratholm, M. J. Packer, A. J. Mulholland, D. R. Glowacki, C. P. Butts, IMPRESSION – prediction of NMR parameters for 3-dimensional chemical structures using machine learning with near quantum chemical accuracy. *Chem. Sci.* (2020), **11**, 508–515

Poster 11

Investigating the molecular mechanism of transport proteins using HDX-MS and MD simulations

Ruyu Jia, Richard T. Bradshaw, Valeria Calvaresi, Chloé Martens, *Argyris Politis*

School of Chemistry, King's College London, 7 Trinity Street, London, SE1 1DB, London, United Kingdom

Despite amazing advances in structure prediction and elucidation in recent years, insights into biomolecular mechanism and function in solution are often unable to be inferred from static structural snapshots alone, and instead require an understanding of conformational dynamics. Hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) is an attractive, rapid, label-free method for investigating dynamics of biomolecules in near-native environments. Furthermore, HDX-MS is combined with molecular simulation approaches, to structurally interpret HDX-MS measurements at the atomistic level by directly correlating experimental and predicted deuterium uptake. However, sampling limitations and uncertainty in the experimental measurements can lead to difficulties in meaningfully interpreting differences between the predicted and observed exchange rates.

Reliable, quantitative, interpretation of HDX-MS measurements instead requires techniques to rigorously select the conformational ensemble that best represents the experimental data. One such approach, HDXer[1], 'reweights' a candidate ensemble of structures, for example from MD simulations, using a minimal bias, such that the predicted and target (experimental) HDX-MS agree within a given level of uncertainty. Here we apply the HDXer approach to identify the protein conformational populations present in apo, substrate- (xylose), and inhibitor- (glucose, phloretin, phloridzin) bound states[2] of the bacterial xylose transporter XylE, a representative of the wider major facilitator superfamily (MFS) transporters.

Making use of molecular docking, binding pose filtering, and extensive molecular dynamics simulations, we first generate a wide-ranging candidate ensemble of potential XylE structures, encompassing outward-facing, inward-facing, and occluded transporter states. Subsequently, we apply computational HDX-MS predictions and ensemble reweighting with HDXer, which suggests that the non-native inhibitors phloretin, and phloridzin exhibit a substantially different mode of action to the native ligands xylose and glucose. The final reweighted structural ensembles, fitted to each experimental dataset independently, allow us to probe this mechanistic difference at the atomistic level, and add crucial high-resolution detail to our inferences of MFS transporter mechanism from HDX-MS experiments.

References:
[1] Bradshaw, R. T., Marinelli, F. et al. Biophys. J. 2020, 118, 1649–1664.
[2] Jia, R., Martens, C. et al. Nat. Commun. 2020, 11 (1), 6162.

Poster 12

Mechanistic insights into the functional dynamics of L1 β-lactamase

Zhuoran Zhao, *Prof Shozeb Haider*

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

β-lactam antibiotic is one of the most widely used antibiotic drugs in the world. The inappropriate and widely use of β-lactam antibiotic leads to the growing of antibacterial resistance and become a major problem for global health. L1 metallo-β-lactamase (MBL) is a class B3 metallo-β-lactamase which expressed by Stenotrophomonas maltophilia. It has a unique tetramer structure and can inhibit the antibiotic function through hydrolyzing the β-lactam. There are two elongated loops exist between α3-β7 and β12-α5 in each monomer of L1. These two loops have shown the ability to influence the substrate binding. To explore how these loops work in the dynamic structure in L1, adaptive sampling molecular dynamics simulations have been employed, and Markov State Model (MSM) have been built. The results indicate that a gating interaction exist between α3-β7 and β12-α5 loops, and the loops can work together with the other key residues (P210, H144, Y212) around the binding site play an important role in maintaining the conformation of the ligand-binding site.