

MGMS and RSC MMG Young Modellers' Forum 2020-21 PROGRAMME & ABSTRACTS

Programme of Oral Presentations

09:15-09:30	Opening Remarks
09:30-09:50	The Role of Hydration and Flexibility in Antibody Binding Mabel Wong, Southampton
09:50-10:10	QM/MM modelling of carbapenem hydrolysis by L1 gives new insight into the mechanistic pathway Rebecca Twidale, Bristol
10:10-10:30	Using imputation to predict in vitro toxicity Moritz Walter, Sheffield
10:30-11:00	Lightning Talks from Poster Presenters
11:00-11:30	BREAK and POSTER SESSION
11:30-11:50	Combining ToF-SIMS and Cheminformatics data to predict bacterial attachment to polymers Leonardo Contreas, Nottingham
11:50-12:10	Understanding Membrane Permeability with a Cyclic Decapeptide Series Shuzhe Wang, ETH Zurich
12:10-12:30	A Computational Workflow for the Discovery of Fluorescent Stapled-Peptides Binders of the SAR-cov2 Spike Protein Marie Bluntzer, Edinburgh
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	Computational Approach to Nanopore Design Zonghua Bo, Oxford
14:10-14:30	Predicting fluoroquinolone resistance: correlations or chemistry? Alice Brankin, Oxford
14:30-14:50	Protein-Ligand Free Energies of Binding from Full-Protein DFT Calculations: Convergence and Choice of Exchange-Correlation Functional Lennart Gundelach, Southampton
14:50-15:20	BREAK and POSTER SESSION
15:20-15:40	Combining Machine Learning and Enhanced Sampling Techniques for Efficient and Accurate Calculation of Absolute Binding Free Energies Rhys Evans, UCL
15:40-16:00	OctSurf: Efficient Hierarchical Voxel-based Molecular Surface Representation for Protein-Ligand Affinity Prediction Qinqing Liu, Connecticut
16:00-16:20	Visual Continuity in Protein Secondary Structure Rendering Alexander Jamieson-Binnie, Bristol
16:20-16:45	JUDGING, FINAL REMARKS, PRIZE ANNOUNCEMENTS

Talk 1

The Role of Hydration and Flexibility in Antibody Binding

Mabel T. Y. Wong^a, Sebastian Kelm^b, Xiaofeng Liu^b, Richard D. Taylor^b, Terry Baker^b, Jonathan W. Essex^a

^aSchool of Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom

^bUCB Celltech, 216 Bath Road, Slough, SL1 4EN, United Kingdom

Antibodies are increasingly attractive biological drugs, and at present are at the forefront of vaccine and therapeutic developments against COVID-19, a pandemic that has caused over 1.4 million deaths as of November 2020¹. The majority of an antibody's high affinity and specificity is dictated via six hypervariable loops at the antigen binding site, known as Complementarity Determining Regions (CDRs). These make antibody binding difficult to predict and understand at the atomic level. So far, computational antibody design attempts have developed binders of therapeutic affinity, but often require human intervention to filter out false positives. A possible reason for human involvement is a limited understanding on the nature of antibody-antigen interactions, and consequently the use of methods that may not account for all features affecting affinity. These computational methods include scoring designs with implicit solvent, and using static crystal structures or models despite the potential for CDRs to be flexible.

To further our understanding on the potential role of water and CDR flexibility at these interfaces, we have carried out a long-timescale simulation study on eighteen antibody crystal structures. Using Replica Exchange with Solute Scaling² and an explicit water model to explore interface conformation and hydration, we find significant numbers of bridging water molecules at the antibody-antigen interface. Additionally, a higher proportion of antibody-antigen interactions, including those mediated by water, are correlated with higher antibody affinity. While the difference in CDR flexibility between bound and unbound structures is not correlated with affinity, we find that fewer CDR conformers in the bound form is associated with enhanced binding. This observation is consistent with the fact that antibody-antigen binding is often enthalpically driven. We thus propose that interactions with waters and CDR flexibility are important in moderating antibody-antigen binding, and that explicit hydration and CDR sampling must be considered to improve antibody affinity prediction and computational design workflows.

References:

- [1] COVID-19 Map, <https://coronavirus.jhu.edu/map.html>, (accessed 30 November 2020).
- [2] L. Wang, R. A. Friesner and B. J. Berne, *J. Phys. Chem. B*, (2011), **115**, 9431–9438.

Talk 2

QM/MM modelling of carbapenem hydrolysis by L1 gives new insight into the mechanistic pathway

Rebecca M. Twidale^a, Philip Hinchcliffe^b, Jim Spencer^b, Adrian J. Mulholland^a

^aSchool of Chemistry, University of Bristol, Cantock's Cl, Bristol, BS8 1TS

^bSchool of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol BS8 1TD.

Widespread bacterial resistance to antibiotic therapies represents a serious threat to global medicine. In the fight against antibiotic resistance, understanding how enzymes produced by bacteria break down antibiotics is crucial in developing inhibitors to restore antibiotic activity. Among the most common of these bacterial enzymes are metallo- β -lactamases (M β L), a class of zinc-containing enzymes that can hydrolyse almost all clinically available β -lactam antibiotics. The hydrolysis mechanism of M β Ls is complex; all known M β Ls exhibit different modes of action.[1] This makes generalizing the mechanism difficult and requires each M β L to be studied individually. L1 is a clinically relevant class B3 M β L known to hydrolyse most classes of β -lactam antibiotics.[2] The aim of our work is to examine how L1 hydrolyses two β -lactam antibiotics, faropenem and ertapenem. To achieve this, we have used a hybrid quantum mechanics/molecular mechanics (QM/MM) approach to model the enzyme active site dynamics and hydrolysis mechanism.

Our simulations have revealed a new mechanistic pathway for the initial step in the hydrolysis mechanism of both faropenem and ertapenem by L1. Following the initial nucleophilic attack of the bridging hydroxide ion, a tetrahedral intermediate is observed that is stabilized by an oxyanion pair formed with an active site zinc ion. This intermediate has not been identified in previous mechanistic studies of L1. We also report on the stability of the double-bond tautomeric forms of each β -lactam. Faropenem and ertapenem display contrasting binding poses in the crystal geometry corresponding to a dihedral rotation around the C5-C6 bond. Ertapenem (non-rotated), shows no preference for either tautomeric form, whereas faropenem (rotated) is more stable in the C-protonated form. This suggests the rotation is important in understanding the final protonation step of the reaction. These findings have given both a greater understanding of how the hydrolysis mechanism of different antibiotics proceeds in L1 and highlight the importance of using appropriate levels of QM theory for mechanistic modelling of M β Ls.

References:

[1] T. Palzkill, *Ann N Y Acad Sci.*, (2013), **1277**, 91-104.

[2] J. H. Ullah, T. R. Walsh, I. A. Taylor, D. C. Emery, C. S. Verma, S. J. Gamblin and J. Spencer, *J Mol Biol.*, (1998) **284**(1), 125-36.

Talk 3

Using imputation to predict *in vitro* toxicity

Moritz Walter^a, Prof. Dr. Valerie J. Gillet^a, Dr. Antonio de la Vega de León^a, Dr. Samuel J. Webb^b

^aInformation School, University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP

^bLhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS

The amount of publicly available toxicity data has increased substantially in recent years, e.g., high-throughput screening efforts in the Tox21 and ToxCast projects. This data may be useful to better understand and predict human toxicity. However, multitarget toxicity datasets are typically sparse, i.e., not every compound was tested in every toxicity assay. Imputation describes the process of predicting missing values in a sparse dataset and hence may fill crucial gaps of knowledge. In contrast to classical QSAR models, which are based solely on relations between chemical descriptors and toxicities, imputation models leverage relations between different assays to make predictions. In comparison to multitask prediction methods, imputation models use known toxicity information to predict unknown toxicity endpoints for test compounds.

In the present study, two different imputation approaches, Feature Nets [1], which uses related assays as additional descriptors in supervised machine learning models, and Macau [2], a matrix factorisation technique, were compared to classical QSAR models using a dataset containing binary mutagenicity data for various bacteria strains (Ames test). Randomly selected cells of the original data matrix were removed when training the imputation models and held back to evaluate their performance. Both imputation approaches clearly outperformed classical QSAR models. Further analyses revealed that the effectiveness of imputation models increased with the amount of available training labels and that imputation models might overcome the limitations of classical QSAR models to make reliable predictions for compounds that are chemically dissimilar to compounds in the training set.

Our results show that imputation represents an attractive approach to predict toxicity in sparse datasets. The benefit over classical QSAR models is achieved by incorporating experimentally determined labels of related toxicity assays into the models.

References:

- [1] A. Varnek, C. Gaudin, G. Marcou., I. Baskin., A. K. Pandey and I.V. Tetko, *J. Chem. Inf. Model.* 2009. **49**, 133-144.
- [2] J. Simm, A. Arany, P. Zakeri, T. Haber, L. K. Wegner, V. Chupakhin, H. Ceulemans and Y. Moreau, 2015, arXiv: stat.ML/1509.04610v2.

Talk 4

COMBINING ToF-SIMS AND CHEMINFORMATICS DATA TO PREDICT BACTERIAL ATTACHMENT TO POLYMERS

Leonardo Contreas, Andrew Hook, Charles Laughton, Phil Williams and Morgan Alexander

School of Pharmacy, University of Nottingham, Nottingham, NG7 2RD

Bacterial infections represent a major problem due to the increasing drug resistance, and they are becoming ever more dangerous in medical environments. In fact, hospital-acquired-infections (HAIs) are more likely to occur if the host is weakened or immunocompromised. Moreover, bacterial colonization of biomedical devices, such as prostheses and catheters, is of growing concern. For this reason, the design of materials naturally resistant to bacterial attachment, and therefore to biofilm formation, is an important strategy in preventing post-implantation infections. Machine Learning (ML) has established itself as a powerful method to perform statistical modelling of complex data in a wide range of fields. Recently, several studies using a Bayesian Regularised Neural Network (BRANN) showed a strong correlation between bacterial adhesion on polymers and their physico-chemical or mass spectrometric descriptors^{1,2}.

However, those previous works were based on a complex, non-linear regression approach that generally focuses on predictive power rather than interpretability. In this study, Time-of-Flight Single Ion Mass Spectrometry (ToF-SIMS) ion peaks and chemoinformatic descriptors for two libraries, including more than five hundred polyacrylates, with experimentally determined bacterial attachment properties have been evaluated as features to train a binary classification model (Logistic Regression) to predict pro- or anti-attachment behaviour against the three most frequent bacterial species occurring in HAIs: *S.aureus*, *P.aeruginosa* and *E.coli*. Results suggest that with appropriate feature selection techniques, attachment behaviour of *P.aeruginosa* and *S.aureus* can be effectively modelled by solely relying upon chemoinformatic descriptors, whose role and importance can be analysed and explained. Subsequently, a virtual screening has been carried out and novel polymers with good predicted anti-attachment properties have been identified.

References:

- [1] Mikulskis, P. *et al. ACS Appl. Mater. Interfaces*, (2018), **10**, 139–149.
- [2] Epa, V. C. *et al. Adv. Funct. Mater.*, (2014), **24**, 2085–2093.

Talk 5

Understanding Membrane Permeability with a Cyclic Decapeptide Series

Shuzhe Wang^a, Gerhard König^a, Hans-Jörg Roth^b, Marianne Fouché^b, Stéphane Rodde^b, Sereina Riniker^a

^a Laboratory of Physical Chemistry, ETH Zürich, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland

^b Novartis Institutes for BioMedical Research, Novartis Pharma AG, Novartis Campus, 4056 Basel, Switzerland

Compared to traditional small-molecule drugs, macrocycles are interesting drug candidates capable of binding to larger but flat binding sites with high affinity. This opens doors for new therapeutic targets like protein-protein interfaces (PPIs) and GPCRs. Cyclic peptides are a subset of macrocycles with the advantage of being generally easier to synthesise and less toxic. However, an issue with these larger molecules is cellular uptake and bioavailability, as reaching the therapeutic targets may require them to first cross several membranes. The identification of peptides that are able to cross cell membranes by passive diffusion is a big challenge. Therefore, an improved understanding of the passive permeability of cyclic peptides using computational means will greatly extend the domain of promising therapeutic candidate and at the same time shorten the drug-development cycle.

In this work, we investigate a series of analogue cyclic decapeptides with side chain residues of different polarity as well as different levels of backbone rigidity. In a previous study [1] with six of these peptides, we observed a correlation between the descriptor extracted from extensive molecular dynamics (MD) simulations and the passive membrane permeability from experimental measurement. Here, our study is broadened to include a richer selection of 24 decapeptides. For each molecule, we conduct a large set of parallel MD simulations from different starting structures to achieve collective sampling of over 10 microseconds. From these simulations, we build kinetic models using Markov state modelling (MSM) to quantitatively determine different stable conformation states and transition time-scale of each decapeptide. This allows us to identify the “closed” (most apolar) conformational state needed for membrane permeation and compare its occupancy amongst the decapeptides. With the added insights from experimentally measured logD and PAMPA (parallel artificial membrane permeation assay) values, we formulate a more general hypothesis of how passive membrane permeability can be tuned by the balance of flexibility and polarity.

References:

[1] J. Witek, S. Wang, B. Schroeder, R. Lingwood, A. Dounas, H. Roth, M. Fouché, M. Blatter, O. Lemke, B. Keller, S. Riniker, *J. Chem. Inf. Model.* 2019, **59**, 294

Talk 6

A Computational Workflow for the Discovery of Fluorescent Stapled-Peptides Binders of the SAR-cov2 Spike Protein.

Marie T. J. BLUNTZER¹, Julien Michel¹, Alison N. Hulme¹

¹Joseph Black Building, David Brewster Road, University of Edinburgh, Edinburgh, EH9 3FJ, UK

Cheap, rapid, sensitive and portable SARS-CoV-2 diagnostics are crucial to monitor the ongoing pandemic and to contribute to the prevention of transmission. PCR tests for diagnosing COVID-19 are currently the most sensitive method to detect SARS-CoV-2 but are slow and expensive, while antibody tests lack sensitivity in the early stages of infection.¹ Here we report the computer-aided design of SARS-CoV-2 spike protein ligands for use as cell permeable optical probes in molecular imaging studies. Such molecules could form the basis of inexpensive point-of-care diagnostic kits, be used to study the life cycle of SARS-CoV-2 in real time, and potentially be developed further as antiviral agents.

We focused on designing SARS-CoV-2 S ligands with a stapled (cross-linked) peptide approach. Stapled peptides can be designed and prepared more rapidly than typical drug-like small molecules yet can achieve comparable bioavailability owing to their increased cell permeability and metabolic stability over traditional peptides.² They also offer a modular scaffold that can be further functionalised for diagnostic or imaging purposes via, for instance, incorporation of fluorescent dyes.

We will first report a study of an all-atom molecular dynamic simulation of the glycosylated spike protein. As sugars act as a protective shield against antigens, the analysis of these preliminary simulations has permitted us to focus on “weak spots” in this shield for the design of our peptides. We then extracted key structural features from published data on ACE2-RBD and antibody-spike protein interfaces and used those as input into our in-house ‘Stapline’ computational pipeline to screen thousands of putative sequences that incorporate staples and fluorophore tags against multiple regions of the spike protein. Structural and energetic descriptors from MD simulations of several hundreds of designs were generated using cloud computing and HPC resources kindly donated by Amazon Web Service and the DiRAC consortium through CCPBioSim and HECBioSim. The most promising peptides are being synthesised and assayed by CROs and academic collaborators in the School of Biological Sciences and the College of Medicine at Edinburgh. Preliminary biophysical characterisation of some of these designs will be presented.

References:

- [1] N. C. Grassly, M. Pons-Salort, E. P. K. Parker, P. J. White, N. M. Ferguson, *Lancet*, **2020**, doi : 10.1016/S1473-3099(20)30630-7
- [2] M. T. J. Bluntzer, J. O’Connell, T. S. Baker, J. Michel, A. N. Hulme, *Pep. Sci.*, **2020**, e24191.

Talk 7

Computational Approach to Nanopore Design

Zonghua Bo, Yujia Qing, Hagan Bayley, Fernanda Duarte

Department of Chemistry, University of Oxford, South Park Road, Oxford, OX1 3TA.

Site-selective modification of biomolecules is key to the development of new therapeutics. It has, however, proved challenging to pick out reactive sites based on chemical reactivity in the presence of chemically equivalent functional groups. Nanopores with a tubular channel present an alternative for achieving selectivity by spatial alignment. Among them, α -Hemolysin (α HL), a transmembrane pore, has been extensively studied as a DNA sequencer and single-molecule nanoreactor with tight spatial control over the pore-substrate interactions. The controlled molecular motion of a small molecule carrying a DNA cargo (hopper) was achieved within a modified α HL pore.¹ The molecule could move along an engineered cysteine track by thiol-disulfide interchange, with the overall directionality set by an applied electric potential.¹ Alignment between the substrate and the cysteine residues was found to be vital for a controlled loading. Recently, it was also shown that hopper variants were confined and elongated in a α HL nanotube to promote site-selective and regioselective chemistry not observed in bulk solution.² To redesign novel α HL-based nanoreactors for site-selective modification a deeper knowledge of its molecular basis is required. In this work, molecular dynamics simulations have been employed to evaluate the effect of confinement and electric field on the hopper-pore interactions and identify key interactions during substrate loading.

A six-cysteine protein track was built onto a β -barrel of α HL, which was inserted into a DPPC membrane. The hopper substrate was bounded to a traptavidin protein and pulled into the pore. Our results show a favorable alignment at foothold 113C and 115C, while unfavorable at 117C, in agreement with experimental results. Hydrogen bonding interactions were identified between the two amide moieties on either side of the disulfide linkage and the pore providing insights into possible mutations for future synthetic efforts.

References:

- [1] Y. Qing, S.A. Ionescu, G.S. Pulcu, H. Bayley. *Science*, (2018), **361**, 908-012.
- [2] Y. Qing, H. Tamagaki-Asahina, S. A. Ionescu, M. D. Liu, H. Bayley. *Nat. Nanotechnol*, (2019), **14**, 1135–1142

Talk 8

Predicting fluoroquinolone resistance: correlations or chemistry?

Alice E. Brankin, Philip W. Fowler, A. Sarah Walker

Nuffield Department of Medicine, University of Oxford, John Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU.

Moxifloxacin is recommended by the WHO for treating multi-drug resistant Tuberculosis. Resistance to moxifloxacin is increasing and is largely attributed to mutations in the *gyrA* and *gyrB* genes of the antibiotic target, DNA gyrase. It is important to know which mutations confer resistance in order to correctly treat patients and prevent the spread of resistance. Both alchemical free energy methods and machine learning using structure-based features have proven effective at predicting resistance associated with mutations in infectious disease¹. As studies to date have concentrated on small monomeric drug targets, we aimed to test whether these methods are applicable to a more complex target, the DNA gyrase cleavage complex, which comprises a 200kDa protein tetramer and covalently bound DNA.

The CRyPTIC consortium has collected over 20,000 *Mycobacterium tuberculosis* isolates, which have been sequenced and phenotyped². This provided our dataset of *gyrA* or *gyrB* mutations and their associated moxifloxacin minimum inhibitory concentration. These data were used to train and test a variety of machine learning algorithms, using structural features obtained from the crystal structure of moxifloxacin-bound DNA gyrase. Free energy calculations were performed using GROMACS to predict the effect of several known mutations on the binding free energy of moxifloxacin.

Machine learning models predicted moxifloxacin susceptibility with >90% specificity and >80% sensitivity and predicted the overall effect of multiple *gyrA* or *gyrB* mutations in individual isolates. Free energy calculations successfully predicted the effects of several known mutations but struggled when predicting the effect of mutations involving charged residues. Machine learning is fundamentally inferential in nature, and hence offers little understanding of the underlying chemistry associated with resistance. We propose that these approaches could be used together to augment genetics-based clinical microbiology diagnostic pipelines, by initially screening novel mutations using a machine learning model. With improvement in system set up, computational architecture and forcefields, alchemical free energy methods could then be applied to the mutations predicted resistant, however this requires more computational resource. Ultimately, alchemical approaches could facilitate the design of a better fluoroquinolone scaffold to increase potency and combat resistance.

References:

- [1] M. Aldeghi, V. Gapsys, B. L. de Groot, ACS Cent. Sci., (2019) 5(8), 1468-1474.
- [2] C. Allix-Béguec et al, N. Engl. J. Med., (2018) 379, 1403–1415.

Talk 9

Protein-Ligand Free Energies of Binding from Full-Protein DFT Calculations: Convergence and Choice of Exchange-Correlation Functional

Lennart Gundelach^a, *Thomas Fox*^b, *Christofer S. Tautermann*^b, *Chris-Kriton Skylaris*^a

^aChemistry, University of Southampton, University Road, SO17 1BJ, United Kingdom

^bDepartment of Lead Discovery, Boehringer Ingelheim Pharma GmbH & Co. KG D-88397, Biberach, Germany

The accurate prediction of protein-ligand binding free energies with tractable computational methods has the potential to revolutionize drug discovery. Modeling the protein-ligand interaction at a quantum mechanical level, instead of relying on empirical classical-mechanics methods, is an important step toward this goal. In this study, we explore the QM-PBSA method to calculate the free energies of binding of seven ligands to the T4-lysozyme L99A/M102Q mutant using linear-scaling density functional theory on the whole protein-ligand complex. By leveraging modern high-performance computing, we perform over 2900 full-protein (2600 atoms) DFT calculations providing new insights into the convergence, precision and reproducibility of the QM-PBSA method. We find that even at moderate sampling over 50 snapshots, the convergence of QM-PBSA is similar to traditional MM-PBSA and that the DFT-based energy evaluations are very reproducible. We show that in the QM-PBSA framework, the physically-motivated GGA exchange-correlation functional PBE outperforms the more modern, dispersion-including non-local and meta-GGA-nonlocal functionals VV10 and B97M-rV. Grimme's D3 is the most promising empirical-dispersion correction to PBE and the three-body dispersion term is significant and improves results. Inclusion of an entropy correction term sampled over less than 25 snapshots is detrimental while an entropy correction sampled over the same 50 or 100 snapshots as the enthalpies improves the accuracy of the QM-PBSA method. As full-protein DFT calculations can now be performed on modest computational resources our study demonstrates that they can be a useful addition to the toolbox of free energy calculations.

Talk 10

Combining Machine Learning and Enhanced Sampling Techniques for Efficient and Accurate Calculation of Absolute Binding Free Energies

Rhys Evans^a, Ladislav Hovan^a, Benjamin P. Cossins^c, Gareth A. Tribello^d, Carolina Estarellas^a, Francesco L. Gervasio^{a,b}

^a Department of Chemistry, University College London, London, WC1E 6BT, United Kingdom

^b Pharmaceutical Sciences, University of Geneva, 1211 Geneva, Switzerland

^c UCB Pharma, Slough, SL1 3WE, United Kingdom

^d Atomistic Simulation Centre, Queen's University, Belfast, BT7 1NN, United Kingdom

Fragment-Based Drug Design (FBDD) has emerged as a new standard for drug discovery. However, its deployment is hindered by several challenges, including the need to crystallise the ligand-bound targets and how to develop fragments into leads. Fast and accurate assessment of protein-fragment absolute binding free energies (ABFEs) is an essential step in the process, making simulations and modelling an attractive alternative to structural experiments, with the potential to enhance the success rate and the applicability of FBDD. While many efforts have been made to improve ABFE prediction, there continue to be problems associated with the computational cost and efficiency, especially for flexible ligands and complex targets. The emergence of machine learning, and its application to drug design, could address both the cost and accuracy problems still present in the field.

Consequently, calculating absolute binding free energies is both challenging and important. In this project, we test some recently developed metadynamics-based methods and developed a novel combination with a Hamiltonian replica-exchange approach: Funnel-SWISH. The methods were tested on 18 chemically diverse ligands, with a wide range of different binding affinities to a realistic and complex target; namely, human soluble epoxide hydrolase¹. The results suggest that metadynamics with a funnel-shaped restraint can be used to calculate, in a computationally affordable and relatively accurate way, the absolute binding free energy for small fragments. When used in combination with an optimal path-like variable obtained using machine learning (COMet-Path) or with the Hamiltonian replica-exchange algorithm (SWISH), this method can achieve accurate results for increasingly complex ligands, with a good balance of computational cost and speed. An additional benefit of using the combination of metadynamics and SWISH is that it also provides useful information about the role of water in the binding mechanism. These results were published in JCTC in 2020².

References:

- [1] L. Öster, S. Tapani, Y. Xue and H. Käck, *Drug Discovery Today*, 2015, **20**, 1104–1111.
- [2] R. Evans, L. Hovan, G. A. Tribello, B. P. Cossins, C. Estarellas and F. L. Gervasio, *J. Chem. Theory Comput.*, 2020, **16**, 4641–4654.

Talk 11

OctSurf: Efficient Hierarchical Voxel-based Molecular Surface Representation for Protein-Ligand Affinity Prediction

Qinqing Liu^a, Peng-Shuai Wang^b, Chunjiang Zhu^a, Blake Gaines^a, Tan Zhu^a, Jinbo Bi^{a,c}, *Minghu Song*^c

^a Computer Science and Engineering Department, University of Connecticut, 371 Fairfield Way, Unit 4155, Storrs, CT 06269-3247, USA

^b Microsoft Research Asia, Beijing, China

^c Department of Biomedical Engineering, University of Connecticut, A.B. Bronwell Building, Room 217, 260 Glenbrook Road, Unit 3247, Storrs, CT 06269-3247, USA

Voxel-based 3D convolutional neural networks (CNNs) have been applied to predict the protein-ligand binding affinity. However, the memory usage and computation cost of these voxel-based approaches increase cubically with respect to spatial resolution and sometimes make volumetric CNNs intractable at higher resolutions. Therefore, it is necessary to develop memory-efficient alternatives that can accelerate the convolutional operation on the 3D volumetric representation of the protein-ligand interaction. In this study, we implement a novel volumetric representation, OctSurf, to characterize the 3D molecular surface of protein binding pockets and bound ligands. The OctSurf surface representation is built based on the octree data structure, which has been widely used in computer graphics to efficiently represent and store 3D object data. Vanilla 3D-CNN approaches often divide the 3D space of objects into equal-sized voxels. In contrast, OctSurf recursively partitions the 3D space containing the protein-ligand pocket into eight subspaces or octants. Only those octants containing Van der Waals surface points of protein or ligand atoms undergo the recursive subdivision process until they reach the predefined octree depth, whereas unoccupied octants are kept intact to reduce the memory cost. Resulting non-empty leaf octants approximate molecular surfaces of the protein pocket and bound ligands. These surface octants, along with their chemical and geometric features, can be used as the input to 3D-CNNs. Two CNN architectures, VGG and ResNet, are applied to the OctSurf representation to predict binding affinity. The OctSurf representation consumes much less memory than the conventional voxel representation at the same resolution. By restricting the convolution operation to only octants of the smallest size, our method alleviates the overall computational overhead of CNN. A series of experiments are performed to demonstrate the memory and computational efficiency of the proposed learning method. Our codes are available at the following GitHub repository: <https://github.uconn.edu/mldrugdiscovery/OctSurf>.

Talk 12

Visual Continuity in Protein Secondary Structure Rendering

Alexander D. Jamieson-Binnie, *David R. Glowacki*

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

Ribbon diagrams are ubiquitous in protein visualisation, where they showcase the secondary structure motifs present. Most existing methods for generating these diagrams are aimed at static structures, and hence do not maintain visual continuity when viewing a molecular trajectory. This manifests itself as certain sections of ribbons flipping between clockwise and counterclockwise twists, as well as abrupt changes in assignment for each residue. We present an approach that yields smooth continuous ribbon diagrams, suitable for use in interactive molecular dynamics in virtual reality (iMD-VR).

There are two sources of discontinuity that require fixing. The issue with discrete assignment of secondary structure is resolved by linearly interpolating between assignments which are calculated once a second. This gives diagrams which transition smoothly between different assignments without any discontinuities. It also gives an indication of sections which are relatively unstable, where residues can be observed to fade between two or more assignments.

The second issue is in which direction to twist the ribbons to follow the shape of the curve, but also maintain continuity between frames. We address this issue by proposing a unique method of morphing the cross section of the ribbon, rather than rotating between two orientations. This yields ribbons which do not exhibit a sudden flip in the direction of their twist, instead squeezing the ribbon through a circular cross section when a large twist would occur. Combined, these two methods prevent visual artifacts that appear when viewing simulations in real time, aiding immersion in virtual reality.

Programme of Poster Presentations

1	Implementation and Validation of the QUBE force field in Sire MD Framework Sofia Bariami, Edinburgh
2	Reaction mechanism of Mevalonate diphosphate decarboxylases using QM/MM Sonia Santos, Northumbria
3	Computational Modelling of Allosteric Sites in the Chemokine Receptors Tianyi Ding, Belfast
4	Computational study of solvent effects in Lewis acid – Lewis base complexation: the BF3 scale Michele Assante, Liverpool JMU
5	Data-driven estimation of optimally-designed perturbation networks in relative alchemical free energy calculations Jenke Scheen, Edinburgh
6	In silico structural & in vitro functional analysis of GPR120 receptor to screen potential anticancer candidates Ajay Pal, TU Dublin
7	Quantum and classical effects in DNA point mutations: Watson-Crick tautomerism in AT and GC base pairs Louie Slocombe, Surrey
8	Comparison of Cellular Morphological Descriptors and Molecular Fingerprints for the Prediction of Cytotoxicity- and Proliferation-Related Assay Endpoints Srijit Seal, Cambridge
9	Improving phase-transfer catalysis by enhancing non-covalent interactions Inigo Iribarren, Trinity College Dublin
10	Identification of structure proposals for the interaction interface between h-FBP21 and the core-splicing protein SmB/B' Marius Wenz, FU Berlin
11	Translocation of flexible and tensioned ssDNA through in silico designed hydrophobic nanopores with two constrictions Punam Rattu, Southampton
12	Evaluating 3D Shape-Similarity Using the Kähler Potential Rachael Pirie, Newcastle

Posters will be available online on the YMF 2020-21 website from 10th- 17th 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

Implementation and Validation of the QUBE force field in Sire MD Framework

Sofia Bariami^a, Lauren Nelson^b, Ringrose Chris^b, Joshua T. Horton^b, Vadiraj Kurdekar^b, Antonia Mey^a, Daniel J. Cole^b, Julien Michel^a

^a Department of Chemistry, University of Edinburgh, Joseph Black Building, Edinburgh, EH9 3FJ

^b School of Natural and Environmental Sciences, Newcastle University, Bedson Building, Newcastle upon Tyne, NE1 7RU

Molecular mechanics (MM) force fields (FFs) are invaluable in the biomolecular modelling community for modelling protein interactions and dynamics, and particularly for computer-aided drug design where rigorous free energy methods may be employed to compute the binding affinity of a candidate drug molecule to a therapeutic target. QUBE is a bespoke molecular mechanics force field for small molecules and proteins developed by the Cole group in Newcastle.¹ Its uniqueness lies in the fact that all the parameters are derived by Quantum Mechanical (QM) calculations. We are interested in incorporating this FF into our alchemical free energy software framework (SOMD) in order to test and assess its performance in comparison to standard FFs such as GAFF. This is going to be achieved by predicting binding free energies between proteins and small molecules. My poster addresses how we implemented support in Sire for QUBE FF, with new features that include, but are not limited, to incorporation of geometric combining rules, position restraints and parsing of xml files, using a mixture of Python and C++ programming languages.

The implementation is tested by computing relative binding free energies for two congeneric series of non-nucleoside inhibitors of HIV-1 Reverse Transcriptase, for whom enhanced sampling has been previously conducted, using QUBE and AMBER/GAFF FFs. What we observed is relatively similar overall performance of AMBER and QUBE on these datasets. Modifications on the bespoke parameter derivation methodologies or updates of the FF functional form can improve its accuracy. The efficient use of GPUs acceleration by SOMD will ease future free energy calculation studies, with QM system-bespoke forcefields. Finally, a future release of SOMD will include these new features and it will result to the benchmarking of novel computational research capabilities.

References:

[1] Joshua T. Horton, Alice E. A. Allen, Leela S. Dodda, and Daniel J. Cole, *J. Chem. Inf. Model.*, (2019), **59**, 1366–1381

Poster 2

Reaction mechanism of Mevalonate diphosphate decarboxylases using QM/MM

Sónia F. G. Santos^{a,b,c}, Gary Black^{a,b}, Warispreet Singh^{a,b,c*}, Meilan Huang^{c*}

^a Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom;

^b Hub for Biotechnology in Build Environment, Newcastle upon Tyne, United Kingdom

^c Department of Chemistry & Chemical Engineering, Queen's University Belfast, BT9 5AG, United Kingdom

Isopentenyl diphosphate (IPP), a product of the mevalonate pathway, is a precursor of the profitable isoprenoids. Isoprenoids are of a valuable commercial interest, offering a novel biosynthetic route for biofuels, pharmaceutical applications and fine chemical production. IPP is synthesized from the mevalonate-5-diphosphate (MVAPP) by the Mevalonate diphosphate decarboxylases (MDD) enzyme. MDD belongs to the GHMP superfamily of metabolite kinases and is responsible for the final step of the mevalonate pathway. It is also a key enzyme in the engineered metabolic pathways. The successful implementation of the MDD enzyme in biotechnological process requires a detailed understanding of the reaction mechanism of IPP production. Despite the recent experimental and computational studies on MDD, the reaction mechanism remains elusive, particularly the magnesium coordination and location of the coordinated water.^{1,2} To elucidate the reaction mechanism of MDD enzyme, we have performed atomistic molecular dynamics simulations (MD) on the MDD enzyme in complex with Mevalonate 5-diphosphate, ATP and two Mg²⁺ ions. The MD simulations provided us with a conformational ensemble of the substrate and its interactions with the catalytic site residues, and also the starting points for the study of the MDD- catalysed phosphorylation and decarboxylation steps to be studied by using Quantum mechanics and Molecular mechanics (QM/MM) and QM/MM MD studies.

References:

1. Chen, C. L., Paul, L. N., Mermoud, J. C., Steussy, C. N. & Stauffacher, C. V. Visualizing the enzyme mechanism of mevalonate diphosphate decarboxylase. *Nat. Commun.* **11**, 3969 (2020).
2. McClory, J., Hui, C., Zhang, J. & Huang, M. The phosphorylation mechanism of mevalonate diphosphate decarboxylase: A QM/MM study. *Org. Biomol. Chem.* **18**, 518–529 (2020).

Poster 3

Alchemical Free Energy Methods Predict the Effect of DNA Gyrase Mutations on Computational Modelling of Allosteric Sites in the Chemokine Receptors

Tianyi Ding^a and *Irina G. Tikhonova*^a

^aSchool of Pharmacy, Medical Biology Centre, Queen's University Belfast, Belfast

Chemokine receptors are drug targets against immune and autoimmune diseases, as well as cancer (1). In recent years, several crystal X-ray structures of chemokine receptors in complex with allosteric ligands have been solved revealing different location of allosteric sites (2). These structures provide new opportunities to explore the properties and behavior of allosteric binding cavities in the chemokine receptor family and predict allosteric sites for the receptors for which structural knowledge is unknown. In this work, we conduct molecular dynamics simulations of the CCR7, CCR5 and CCR4 receptors in an empty form and the presence of allosteric ligands or probe molecules in a water-lipid bilayer. The results of simulations highlight key conserved and non-conserved amino acid residues holding a ligand in allosteric sites at extracellular and intracellular sides of the receptors. Analysis of druggability and physicochemical properties of allosteric cavities, and topology and functional groups of structures of allosteric ligands indicate unique features of allosteric binding. The location and amino acid residues of allosteric binding sites at the CCR4 receptor are predicted from experimental tests. Subtype selectivity of allosteric binding sites is discussed. The characterization of allosteric binding sites may provide new structure-based strategies in identifying a novel type of chemokine receptor modulators.

References:

1. Pharmacological modulation of chemokine receptor function. Scholten DJ, et al. *Br J Pharmacol.* 2012 Mar;**165(6)**:1617-1643.
2. Structural Basis for Allosteric Ligand Recognition in the Human CC Chemokine Receptor 7. Jaeger K, et al. *Cell.* 2019 Aug 22;**178(5)**:1222-1230.e10.

Poster 4

Computational study of solvent effects in Lewis acid – Lewis base complexation: the BF_3 scale

Michele Assante^a, Andrew G. Leach^b

^aLiverpool John Moores University, School of Pharmacy and Biomolecular Sciences, Byrom Street Liverpool L3 3AF.

^bUniversity of Manchester, Division of Pharmacy and Optometry, 46 Grafton Street Manchester M13 9NT.

Methods to account for solvent effects in quantum mechanical calculations are generally divided between efficient implicit (solvent as continuum) and more demanding explicit (solvent molecules themselves included) solvent models. When strong interactions between solute and solvent lead to solvent molecules having a significant residence time in particular locations around a solute, implicit solvent models cannot account for this effect. In such cases, adding a number of explicit solvent molecules to the representation of the solute, to create a 'supermolecule', can improve accuracy but anticipating the number and placement of such molecules remain challenges.

To address this, we computed the enthalpy of complexation of 72 diverse Lewis bases with the Lewis acid BF_3 in dichloromethane.¹ The corresponding experimental values had been reported previously. Several implicit solvation models and DFT approaches were tested. A good overall performance was achieved: RMSE of 1.42 kcal/mol for SMD/M06/6-31+G**. However, the enthalpy of complexation diverged markedly for some chemical groups: amines and esters were underestimated, whereas S=O and P=O containing compounds were overestimated. These are likely cases that will require explicit solvent molecules (although other sources of error may also be at play).

The solvent density and the electrostatic potential around the solute were analysed with 3D-RISM and quantum calculations.² These highlight likely sites and strengths of solvent-solute interactions and permit the design of supermolecular complexes that give the best that both implicit and explicit solvation have to offer. We will present graphical illustrations of the performance of implicit solvation and a decision tree that determines when and where explicit solvent molecules should be added. The insights provided by 3D-RISM and the electrostatic potential will be described.

References:

[1] P. C. Maria and J. F. Gal, *J. Phys. Chem.*, 1985, **89**, 1296-1304.

[2] Oliver, C. A. Hunter, R. Prohens and J. L. Rosselló, *Journal of Computational Chemistry*, 2018, **39**, 2371-2377

Poster 5

Data-driven estimation of optimally-designed perturbation networks in relative alchemical free energy calculations

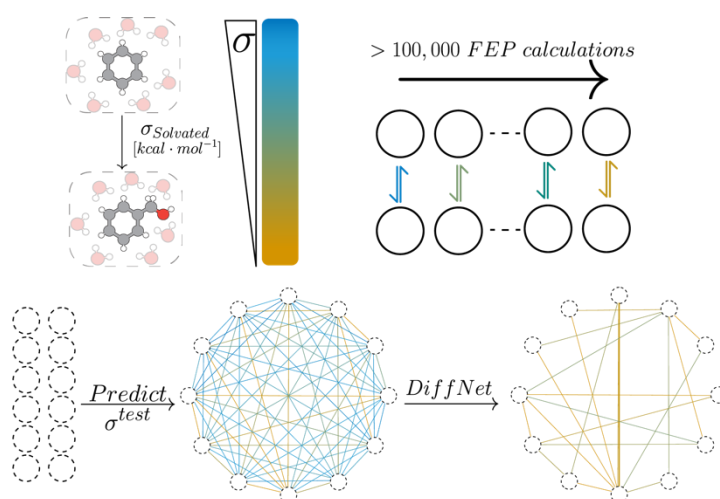
Jenke Scheen[†], Antonia Mey[†], Mark Mackey[‡], Julien Michel[†]

[†]University of Edinburgh, EastCHEM School of Chemistry, David Brewster Rd, Edinburgh EH9 3FJ, UK

[‡]Cresset Group, New Cambridge House, Bassingbourn Rd, Litlington, Cambridgeshire SG8 0SS, UK

Alchemical free energy (AFE) calculations are helpful in solving the ligand-optimisation problem in both academic and corporate drug discovery. A critical step in relative AFE workflows is the generation of effective perturbation networks to ensure transformations are being simulated with sufficient phase-space overlap; currently, state-of-the-art softwares use primarily molecular similarity metrics to estimate optimal edges in networks. Although fairly robust, this method often still requires the user to review and tweak the presented perturbation network, hindering development toward a fully-automated AFE workflow.

The current study attempts to use an alternative approach by using machine-learning algorithms trained on a large number of solvation AFE calculations to predict the difficulty of a given perturbation *a priori* and thereby whether the edge should be simulated or not. Additionally, we use optimal network design algorithms [1] to generate perturbation networks that do not require user intervention, allow per-edge sampling allocation given an input desired sampling quantity, provide statistically optimal free energy predictions and allow scaling of AFE calculations to congeneric series of >100 ligands.



Overview of presented workflow. Top row: a large number of solvation free energy perturbation calculations were performed spanning a chemically diverse space of molecular transformations. Subsequently, for each calculation the perturbation difficulty (σ ; colour range) was computed based on its molecular dynamics outputs resulting in a sizeable training set. **Bottom row:** machine-learning models trained to predict σ *a priori* explore all possible transformations for a given FEP congeneric series. Using a difference network approach [1], statistically-optimal perturbation networks are generated autonomously.

1. Xu, H. Optimal Measurement Network of Pairwise Differences. J. Chem. Inf. Model. 2019, 59, 4720– 4728, DOI: 10.1021/acs.jcim.9b00528

Poster 6

***In silico* structural & *in vitro* functional analysis of GPR120 receptor to screen potential anticancer candidates.**

Ajay Pal^{a,b}, James F. Curtin^a, Gemma K. Kinsella^a

^aSchool of Food Science and Environmental Health, College of Sciences and Health, Technological University Dublin, Dublin D07 ADY7, Ireland

^bEnvironmental Sustainability and Health Institute (ESHI), Grangegorman, Technological University Dublin, Dublin, D07 H6K8, Ireland.

The G-protein coupled receptor - GPR120 has recently been implicated as a novel target for colorectal cancer (CRC) and other cancer managements. In this study, a homology model of GPR120S (short isoform) was generated to identify potential anti-cancer compounds targeting the GPR120 receptor using a combined *in silico* docking-based virtual screening (DBVS), structure-activity relationship (SAR) study, molecular dynamics (MD) simulations and *in vitro* screening approach. ZINC and SPECS databases of synthetic chemical compounds were screened using the developed GPR120S model to identify molecules binding to the orthosteric pocket followed by an AutoDock SMINA rigid-flexible docking protocol.

DBVS hit molecules were then tested *in vitro* to evaluate their cytotoxic activity against SW480 – human CRC cell line expressing GPR120. The test compound **1** (3-(4-methylphenyl)-2-[(2-oxo-2-phenylethyl)sulfanyl]-5,6-dihydrospiro(benzo[h]quinazoline-5,1'-cyclopentane)-4(3H)-one) showed ~90% inhibitory effects on cell growth with micromolar affinities (IC₅₀ = 23.21-26.69 μM). Finally, SAR analysis of compound **1** led to the identification of two more active compounds from the SPECS database showing better efficacy during cell-based cytotoxicity assay –5 (IC₅₀ = 5.89 - 6.715 μM) and 7 (IC₅₀ = 6.789 - 7.502 μM). Simultaneously, MD simulations of DBVS hits interacting with a key residue Asn313 led to a predicted antagonism of GPR120S.

The GPR120S homology model generated, and SAR analysis conducted by this work discovered a potential chemical scaffold, dihydrospiro(benzo[h]quinazoline-5,1'-cyclopentane)-4(3H)-one, which will aid future research on anti-cancer drug development for CRC management.

Poster 7

Quantum and classical effects in DNA point mutations: Watson-Crick tautomerism in AT and GC base pairs

L. Slocombe¹, J. S. Al-Khalili² and M. Sacchi³

¹ *Leverhulme Quantum Biology Doctoral Training Centre,*

² *Department of Physics, and,*

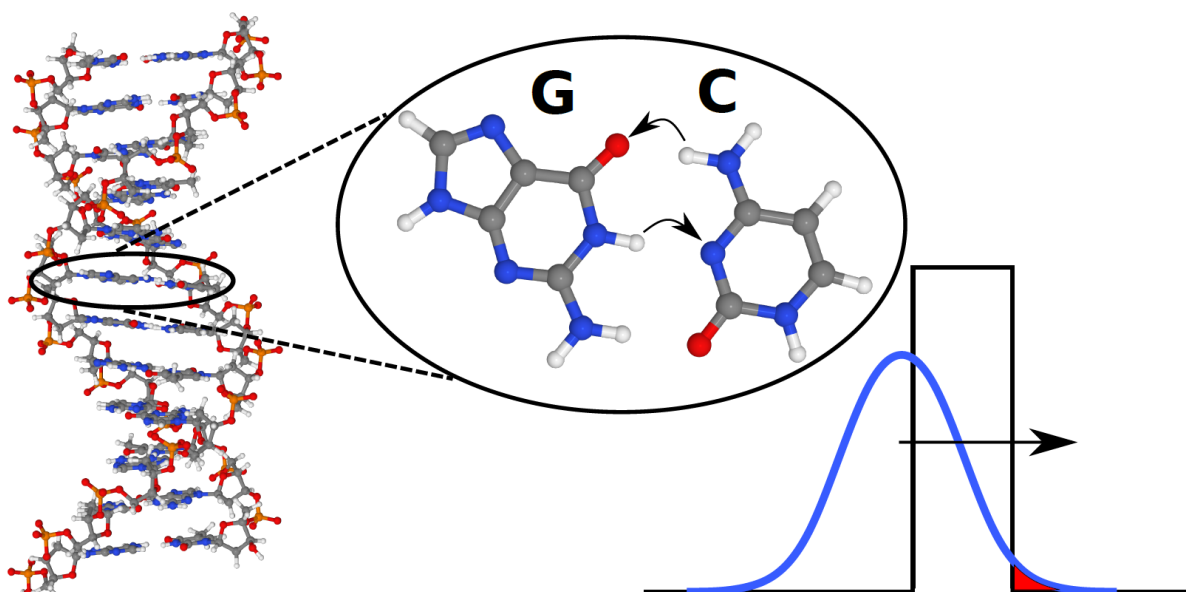
³ *Department of Chemistry, University of Surrey, Guildford GU2 7XH, UK*

Proton transfer along the hydrogen bonds of DNA can lead to the creation of short-lived, but biologically relevant point mutations [1] that can further lead to gene mutation and, potentially, cancer. In this work, the energy landscape of the canonical A-T and G-C base pairs (standard, amino-keto) to tautomeric A*-T* and G*-C* (non-standard, imino-enol) Watson-Crick DNA base pairs is modelled with density functional theory and machine-learning nudge-elastic band methods. We calculate the energy barriers and tunnelling rates of hydrogen transfer between and within each base monomer (A, T, G and C). Using a purely quantum mechanical approach, we show that the role of tunnelling in A-T tautomerisation is statistically unlikely due to the presence of a small reverse reaction barrier. On the contrary, the thermal populations of the G*-C* point mutation could be non-trivial and propagate through the replisome. For the direct intramolecular transfer, the reaction is hindered by a substantial energy barrier. However, our calculations indicate that tautomeric bases in their monomeric form have remarkably long lifetimes. On the other hand, the quantum motion of the protons could be treated separately within an open quantum systems approach [2], where the system is evolved while coupled to a thermal bath of quantum oscillators representing its surrounding cellular environment of water molecules [3]. This combined modelling of the potential energy surface and open quantum systems can accurately account for proton tunnelling in a biological environment which could play a relevant role in DNA replication.

[1] O. Brovarets', D. Hovorun, *J. Biomol. Struct. Dyn.*, (2018) 37(7), pp.1880-1907.

[2] H. Breuer, F. Petruccione, *The theory of open quantum systems*, (2010), Oxford: Oxford Univ. Press.

[3] A. Godbeer, J. Al-Khalili, and P. Stevenson, *Phys. Chem. Chem. Phys.*, (2015), 17(19), pp.13034-13044.



Poster 8

Comparison of Cellular Morphological Descriptors and Molecular Fingerprints for the Prediction of Cytotoxicity- and Proliferation-Related Assay Endpoints

Srijit Seal, *Andreas Bender*

Centre for Molecular Informatics, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom.

Cell morphology features, such as from the Cell Painting assay, can be generated at relatively low cost and represent versatile biological descriptors of a system, and thereby compound response. In this study, we explored cell morphology descriptors and molecular fingerprints, separately and in combination, for the prediction of cytotoxicity- and proliferation-related in vitro assay endpoints.

We selected 135 compounds from the MoleculeNet Toxcast benchmark dataset which were annotated with Cell Painting readouts, where the relatively small size of the dataset is due to the overlap of required annotations. We trained Random Forest classification models using nested cross-validation and Cell Painting descriptors, Morgan and ErG fingerprints, and their combinations. When using leave one-cluster-out cross validation (with clusters based on physicochemical descriptors), models using Cell Painting descriptors achieved higher average performance over all assays (Balanced Accuracy of 0.65, Matthews Correlation Coefficient of 0.28 and AUC-ROC of 0.71) compared to models using ErG fingerprints (BA 0.55, MCC 0.09 and AUC-ROC 0.60) and Morgan fingerprints alone (BA 0.54, MCC 0.06 and AUC-ROC: 0.56). When using random shuffle splits, the combination of Cell Painting descriptors with ErG and Morgan fingerprints further improved balanced accuracy on average by 8.9% (in 9 out of 12 assays) and 23.4% (in 8 out of 12 assays) compared to using only ErG and Morgan fingerprints, respectively.

Regarding feature importance, Cell Painting descriptors related to nuclei texture, granularity of cells and cytoplasm, as well as cell neighbours and radial distributions were identified to be most contributing, which is plausible given the endpoint considered. We conclude that cell morphological descriptors contain complementary information to molecular fingerprints which can be used to improve the performance of predictive cytotoxicity models, in particular in areas of novel structural space.

Poster 9

Improving phase-transfer catalysis by enhancing non-covalent interactions

Inigo Iribarren, *Cristina Trujillo*

School of chemistry, Trinity College Dublin, 152-160 Pearse St. Dublin 2, Dublin.

A wide variety of asymmetric transformations catalysed by chiral catalysts have been developed for the synthesis of valuable organic compounds in the past several decades. Within the asymmetric catalysis field, phase-transfer catalysis¹ has been recognized as a powerful method for establishing useful procedures for organic synthesis.

While the field of asymmetric organocatalysis is currently growing exponentially, an understanding of the mechanistic details involved in most of these reactions has often lagged far behind the pace of catalyst development, which in turn slows down the rational design of catalysts. Therefore, continuous efforts should be made toward the design and development of new catalyst classes, as well as understanding existing relationships between the structure of the catalyst and its ability to transfer stereochemical information². Herein, with the objective of providing a definitive binding mode for bifunctional PTCs, we present a theoretical study of different interactions that can be established between a well-known alkaloid quinine-derived PTC and four different anions of interest in organocatalysis: Cl⁻, Br⁻, MeCO₂⁻ and CN⁻.

Finally, a theoretical study of the free-energy profile corresponding to the enantioselective conjugate cyanation of a α,β -unsaturated ketone was performed. The reaction was studied under the presence of a widely used phase-transfer cinchona alkaloid based catalyst. The binding mode was explored as well as the type of interactions upon complexation characterised. Once that scenario was completely covered, an improved PTC derivative to the existing system was proposed in order to increase the enantioselectivity of a model reaction and therefore a theoretical study of the free-energy profile was performed.

References:

- [1] Graham, T., Renslow, R., Govind, N. and Saunders, S., (2016). *The Journal of Physical Chemistry C*, **120(35)**, pp.19837-19847.
- [2] Petrova GP, Li HB, Maruoka K, Morokuma K. , (2014) *J Phys Chem B.* ;**118(19)**, 5154-67.

Poster 10

Identification of structure proposals for the interaction interface between h-FBP21 and the core-splicing protein SmB/B'

Marius Wenz^a, Miriam Bertazon^b, Jana Sticht^{b,c}, Stevan Aleksić^a, Christian Freund^b and Bettina G. Keller^{*a}

^aDepartment for Biology, Chemistry, Pharmacy, Institute for Chemistry and Biochemistry, Molecular Dynamics Group, Arnimallee 22, Freie Universität Berlin, 14195 Berlin, Germany ^bDepartment for Biology, Chemistry, Pharmacy, Institute for Chemistry and Biochemistry, Protein Biochemistry Group, Thielallee 63, Freie Universität Berlin, 14195 Berlin, Germany ^cCore Facility BioSupraMol, Takustr. 3, Freie Universität Berlin, 14195 Berlin, Germany

The human formin-binding protein 21 (h-FBP21) was identified to interact with the core-splicing protein SmB/B' during the splicing process of the eukaryotic gene expression. This protein-protein interaction rests on the formation of a famous protein-protein interface that also occurs frequently in many other biological processes: tandem repeats of WW domains (tWWs) selectively recognising proline-rich sequences of the targeted protein. As this protein-protein interface is characterised by high reversibility despite high specificity, the identification of possible complex structures is, however, very demanding. In this study, we present the protein-protein interface formed between a h-FBP21 tWW and a specific proline-rich sequence of the SmB/B', called 'SmB2 ligand'.

We investigated the conformational ensemble of the h-FBP21 tWW using classical MD simulations and NMR spectroscopy. The h-FBP21 tWW consists of two successive WW domains, each comprising a rigid, triple-stranded β -sheet with a binding site, called 'XP groove'. Linked by a fully flexible interdomain region, these WW domains can adapt a huge variety of possible arrangements without changes in their respective folding. We therefore assume binding events of the h-FBP21 tWW to be driven by a conformational-selection mechanism, in which a potential ligand can 'select' any binding-competent conformation of the h-FBP21 tWW whenever it is formed.

To identify conformations that are possibly competent to bind the SmB2 ligand, we applied the density-based common-nearest neighbor algorithm on our MD data set and performed docking experiments with HADDOCK protocol. Characterising the resulting complexes in great detail, we finally determined two stable complex structures that we assume to represent the recognition event and therefore the formation of such a protein-protein interface best.

Poster 11

Translocation of flexible and tensioned ssDNA through *in silico* designed hydrophobic nanopores with two constrictions.

Punam Rattu, Bastien Belzunces, Taylor Haynes, Chris-Kriton Skylaris, Syma Khalid

School of Chemistry, University of Southampton, Highfield Campus, Southampton, SO17 1BJ.

DNA sequencing using nanopores is now a well-established 'next generation sequencing' method. Over the years, research has focused towards optimising the apertures within the membrane with the aim of controlling conformation and translocation rate of DNA to improve the resolution of the devices. Experiments involving protein engineering of *S. aureus* toxin α -hemolysin (a heptamer that forms a 14-stranded β -barrel in the membrane) revealed that basic residues introduced in the pore lumen slowed down the translocation of ssDNA. Subsequent simulation studies of simplified models of the α -hemolysin pore region showed that the strong electrostatic interactions between the basic residues and the acidic phosphate groups of the DNA led to transient tethering of the DNA to the pore, which resulted in undesirable conformations that may impact on the accuracy of the read.¹

Model pores with central hydrophobic constriction region maintained DNA in a largely linear conformation, resulting in bases translocating through the constriction in the correct order.² Taking this forward, and inspired by experimental work showing the potential advantages of nanopores with two constrictions for DNA sequencing, we have designed model pores with two hydrophobic constrictions (14- and 16-stranded β -barrels) to explore the impact of pore geometry and chemical nature of the constrictions on DNA translocation rate and conformations using MD simulations.

Our results show that DNA entry into these model pores is strongly influenced by the pore width under an applied electric field. However, no correlation was found between pore width and translocation rate when DNA is inside the pore. DNA translocation was slowed down by aromatic residues in the constrictions, which interacted with DNA and also caused the strand to deviate from its linear conformation. Overall, our pores with phenylalanine constrictions are the most promising as they slowed down DNA translocation whilst not causing too many conformational changes.

References:

- [1] P.J. Bond, A. T. Guy, A. J. Heron, H. Bayley and S. Khalid, *Biochemistry*, (2011), **50**, 3777-3783.
- [2] T. Haynes, I. P. S. Smith, E. J. Wallace, J. L. Trick, M. S. P. Sansom and S. Khalid, *ACS nano*, (2018), **12**, 8208-8213.

Poster 12

Evaluating 3D Shape-Similarity Using the Kähler Potential

Rachael M. E. Pirie,^a Daniel J. Cole,^a Matthew Forshaw,^b Stuart J. Hall.^c

^a School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.

^b School of Computing, Newcastle University, Newcastle upon Tyne, NE4 5TG.

^c School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.

Virtual screening has become popular in recent years to help reduce the cost of drug discovery in terms of human effort and resources used. One approach which has recently gained popularity is the use of shape-based similarity searching, derived from the principle that similar molecules are likely to display similar properties (including biological activity).¹ In this field mathematical descriptions are used to approximate the 3D shape of molecules. These are less expensive to compute than quantum-mechanical models, allowing for faster screening of large databases. The use of a shape-based descriptor to quantify similarity is advantageous compared to molecular fingerprints or physicochemical properties, as it allows for consideration of the conformers accessible to a molecule. These methods also enable scaffold hopping, which can aid rescue of drug candidates found to have issues such as poor solubility or toxicity.

We outline a new approach for approximating shape using the quantised Kähler potential (a concept taken from differential geometry) to represent the geometry of the molecular surface.² The surface is treated as topologically equivalent to a sphere embedded in complex projective space (CP^1). The Kähler potential is then calculated as a Taylor series to produce a Hermitian matrix which approximates the molecular shape. Representations can then be compared using a similarity metric to allow for shape-similarity between molecules to be determined. Our method offers the advantage of control over the level of approximation, allowing balance of accuracy and speed. In addition, these representations are translation invariant, and while not themselves rotation invariant, the behaviour of Hermitian matrices is well defined, allowing for straightforward identification of cases where the same molecule has simply been rotated. We present our results from initial development of the method, as well as some of the challenges faced, and an outline of our proposed benchmark study to compare our new method to existing open-source software.

References:

[1] A. Kumar and K. Y. J. Zhang, *Front. Chem.*, (2018), **6**, 1-21.

[2] S. K. Donaldson, *Pure Appl. Math. Q.*, (2009), **5**, 571-615.