MGMS and RSC MMG
Young Modellers' Forum 2014
PROGRAMME & ABSTRACTS

Programme of Oral Presentations

8.45 – 9.20	Coffee and Registration
9.20 – 9.30	**Welcome and Introduction**
9.30 – 9.50	**Modelling the Selectivity of an Anticancer Drug Through Molecular Simulations** Silvia Lovera, *University College London*
9.50 – 10.10	**Towards a Complete Computational Site-Directed Mutagenesis Protocol** Pietro Aronica, *Imperial College London*
10.10 – 10.30	**Simulating the Role of Conformational Stabilization in p53 and its Effect on Oncogenesis** Zohra Ouaray, *University of Southampton*
10.30 – 10.50	**Prediction of hydration free energy involved in protein-ligand binding with Grid Cell Theory** Georgios Gerogiokas, *EaStCHEM School of Chemistry, Edinburgh*
10.50 – 11.20	**Tea/Coffee break**
11.20 – 11.40	**Global Optimisation of Hydrated Sulfate Clusters** Lewis Smeeton, *University of Birmingham*
11.40 – 12.00	**A DFT study of CO_2, H_2O and CO adsorption on Ni/YSZ(111): Solid Oxide Fuel Cell application** Abdelaziz Essadek, *University College London*
12.00 – 12.30	**Lightning Talks** *All poster presenters, in poster number order.*
12.30 – 14.00	**Lunch and Poster Session**
14.00 – 14.20	**QM/MM analysis of Lysozyme: obtaining novel insights into Biochemistry's 'classic' enzyme** Michael Limb, *University of Bristol*
14.20 – 14.40	**MOARF: Multi Objective Automated Replacement of Fragments** Nicholas Firth, *The Institute of Cancer Research*
14.40 – 15.00	**Improving MFS modeling using sequence analysis** Joanna Lee, *University of Oxford*
15.00 – 15.30	**Tea/Coffee break**
15.30 – 15.50	**Kinetic Model of Peptides: Automatic Construction Using Transferable Basis Sets** Francesca Vitalini, *Freie Universität Berlin*
15.50 – 16.10	**Ligand-based drug target prediction in the absence of structural similarity** Daniel Reker, *ETH Zurich*
16.10 – 16.30	**Similarity in the Context of Orphan Drug Legislation** Pedro Franco, *University of Sheffield*
16.30 – 16.45	**Judging, Prizes and Close**

Talk 1

Modelling the Selectivity of an Anticancer Drug Through Molecular Simulations

Silvia Lovera, Giorgio Saladino, Maria Morando, *Francesco L. Gervasio*

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

We studied the role of TKs dynamics in the binding of Type II inhibitors such as imatinib using state-of-the-art simulations and experiments. On the whole, our results support a new theory, according to which the degree of local motions in kinases is predictive of their sensitivity towards type-II drugs.

A lower drug activity is associated with a higher rigidity of crucial structural features (A-loop and αC-helix), which is translated into a destabilization of the inactive (DFG-out) state. On the contrary, a higher plasticity appears to favour the adoption of the DFG-out conformation and, as a result, the binding of the drug.

Interestingly, the alteration of the dynamics profile also has a relevant impact on the kinetics of the DFG-flip (passage from active DFG-in to inactive DFG-out state), with rigid kinases showing much higher energy barrier for the transition[1].

On a wider perspective, it appears that local fluctuations can predict for type-II inhibitors activity as they are an indicative of the stability of the inactive state. As the degree of local motions in these states is strictly associated with the protein sequence and own interaction network, plasticity could be considered an intrinsic property of each individual kinase: a “Dynamic Fingerprint”.

The ``Dynamic Fingerprint'' also reconciles results obtained for oncogenic mutants, which appear to have a similar shift towards the active state, accompanied by some degree of structural rigidification. Beside opening new ways for the rational design of new drugs, our results suggests that screening for the kinase dynamics profile could become a routinely used step in the identification of type-II drugs targets.

References:

[1] Lovera S, Gervasio F.L and Sutto L., Boubeva R., Scapozza L., Dölker N., The different flexibility of c-Src and c-Abl kinases regulates the accessibility of a druggable inactive conformation, *JACS*, (2012), **134**, 2496-2499.

Talk 2

Towards a Complete Computational Site-Directed Mutagenesis Protocol

P.G.A. Aronica, *Prof. R.J. Leatherbarrow, Dr. I.R. Gould*

Imperial College London, Department of Chemistry, Institute of Chemical Biology, South Kensington Campus, London SW7 2AZ

The use of computational library screens in order to aid with the development of drugs and the analysis of compounds is now widespread[1]. The main advantage of such methods is the possibility to automate processes where the experimental cost of testing may be too high to be profitable. By employing a virtual screen, previously intractable volumes of data may be accessed at a smaller cost.

Based on this observation, we are developing a computational tool for the mutagenesis of proteins. This protocol can be used to mutate residues in a protein or peptide and analyse the effects of that mutation on the protein’s properties or interaction with another compound. This, in turn, can lead to the rational design of proteins and can be an additional tool for their analysis.

The protocol was developed with the AMBER MD package[2], and special emphasis was given to its compatibility with other tools in the same suite. It is able to mutate at will between almost any pair of amino acids, in a highly automated process. Testing is ongoing to compare with current similar protocols and to determine the reliability of these computationally generated values in relation to experimental data, with encouraging results. In particular we have been testing the computational growth of chemical staples on peptides, which has had, so far, relatively good agreement with previously published laboratory results.

The ultimate aim of this protocol is to be employed in antibody design, though its flexible and versatile nature makes it possible to be used in any system that might require amino acid mutations.

References:

[1] Clark, D. E. What has virtual screening ever done for drug discovery? *Expert Opin Drug Discov.*, (2008), **(8)**, 841–851.

[2] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012), *AMBER 12*, University of California, San Francisco.

Talk 3

Simulating the Role of Conformational Stabilization in p53 and its Effect on Oncogenesis

Zohra Ouaray[a], *Prof. Jonathan W. Essex*[a], *Prof. Chandra S. Verma*[b]

[a]*School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK*
[b]*Bioinformatics Institute, 30 Biopolis Street, #07-01 Matrix, Singapore, 138671, Singapore*

The p53 tumour suppressor protein functions as a transcription factor to prevent cancer by inducing cell cycle arrest, DNA repair and apoptosis. 50 % of cancers are associated with a mutation on this protein; indeed mutated p53 either loses its capacity to bind DNA or is destabilised which induces the proliferation of DNA-damaged cells. 95 % of these mutations occur on the p53 DNA-binding domain (DBD). In this presentation, the case of mutations inducing p53 destabilization - so-called "structural mutants" - is investigated using computer simulations. One of the first "structural mutants" studied is V143A which is a temperature sensitive mutation. This mutation localized in the β-sandwich induces p53 destabilization by creating a cavity (at physiological temperature) and influences the DNA-binding by a long range effect (at low temperature, 30-32^{0}C).

Structural models of the p53 DBD wild-type and mutant, chosen from among the 70 human p53 DBD structures (NMR and crystallographic) listed in the protein data bank, were used for Molecular Dynamics (MD) study. Different protonation methods were tested and compared as well as methods for zinc parameterization to propose a model as close as possible to the original protein. MD simulations using AMBER 12 and CHARMM 36 were performed on these models at different temperatures and ion concentrations to sample conformations of p53 DBD wild type and mutant in different conditions. Several analysis methods were used to study these conformational samples to investigate their stability.

Large scale motions (folding and unfolding) on biological molecules occur over 10^{-7} to 10^{4} s. Knowing this, to increase the sampling of our simulations, two new molecular simulation methods were carried out: Accelerated Molecular Dynamics (aMD) [1] which by adding a function to bias the simulation sampling will induce a better exploration of the free energy surface and allow us to gather more conformations and the Diffusion Map Directed Molecular Dynamics (DMDMD) [2], a

technique combining locally scaled diffusion map and MD. In this presentation, the results of these simulations of p53 will be described, indicating the extent and mechanism by which DBD destabilisation occurs.

References:

[1] Pierce LCT, Salomon-Ferrer R, Augusto F de Oliveira C, McCammon JA, Walker RC (2012) Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. J Chem Theory Comput 8: 2997–3002.

[2] Zheng W, Rohrdanz M a, Clementi C (2013) Rapid Exploration of Configuration Space with Diffusion-Map-Directed Molecular Dynamics. J Phys Chem B.

Talk 4

Prediction of hydration free energy involved in protein-ligand binding with Grid Cell Theory

Georgios Gerogiokas *†,* Richard H. Henchman ‡, Michelle W. Y. Southey §, Michael Bodkin §, Michael Mazanetz §, Alexander Heifetz §, and *Julien Michel †**

† EaStCHEM School of Chemistry, Joseph Black Building, The King's Buildings, Edinburgh, EH9 3JJ, United Kingdom ‡ Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom and School of Chemistry, The University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom § Evotec (UK) Ltd 114 Innovation Drive, Milton Park, Abingdon, Oxfordshire OX14 4RZ, UK

Accurate prediction of the hydration energy involved in the forming of the protein-ligand complexes is vital for effective and cost-efficient drug discovery. Currently most of the free binding energy predictors do not take into account this hydration term. We have been developing a novel methodology to quantify and to visualize the water energetics involved in protein-ligand binding.

We applied the Grid Cell Theory (GCT) simulating approach to calculate and average interaction energies, forces and torques acting on each water molecule to parameterise a six dimensional harmonic potential used to compute the free energy of waters. The harmonic potential yields analytical expressions for entropies and enthalpies and are derived through ratios of bulk parameters (forces, torques, and energies) and solute perturbed water parameters. In order to spatially resolve these thermodynamic quantities we used a grid representation. This allowed us not only to calculate the intrinsic energetics for each water site but also to visualize it on the surfaces of proteins and ligands [1]. We established a GCT protocol to calculate the solvation energy of the protein and ligand in their unbound and bound states. The energetics of these two states can be visualized as isosurfaces which enabled us to distinguish between energetically stable and unstable water sites. The water network reorganization and the energetic cost incurred in these changes are also explored. These visualization methods together with energy calculations are useful for rationalizing and speeding up the hit-to-lead and lead optimizations steps of drug-discovery process.

The GCT solvation terms as well as the interaction energies are being assessed to find predictive protein-ligand scores. These protocols are being investigated with combinations of these terms. We performed simulations of the unbound protein and congeneric ligand series. Preliminary results suggest that for particular protein-ligand complexes certain combinations of GCT terms are more predictive which reflect differences in system properties.

References:

[1] Georgios Gerogiokas, Gaetano Calabro, Richard H. Henchman, Michelle W. Y. Southey, Richard J. Law, and Julien Michel, Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory, *J. Chem. Theo. Comput.,* (2014) **10**, 35-48.

Talk 5

Global Optimisation of Hydrated Sulfate Clusters

LC. Smeeton, *RL. Johnston*

School of Chemistry, University of Birmingham, Edgbaston, Birmingham, B15 2TT

The Hofmeister series ranks the ability of ions to affect a number of phenomena in physical chemistry [1], including surface tension, protein solubility and stability and bubble formation to name a few. SO_4^{2-} is a Hofmeister ion which sits at the far "kosmotropic" end of the series, meaning that it increases surface tension and decreases protein solubility. Despite the relative importance of the series, its exact chemical origin remains unclear, with evidence supporting both direct ion interaction and long range effects of the ion on solvent structure. Infrared photodissociation (IRPD) spectroscopy of gas phase hydrated sulfate ions provide information as to the hydrogen bonding of waters at the surface *and* interior of the cluster[2]. IRPD spectra of size selected $SO_4^{2-}(H_2O)_n$ suggest that for $n \leq 43$, all the waters in the cluster are hydrogen bonded to either the sulfate or another water. This contrasts with both water clusters of comparable size and at the surface of the bulk, where a number of waters exhibit "free", non-bonded OH-groups.

The aim of our study: Can simulation detect the size-dependent appearance of free OH-groups in hydrated sulfate clusters by searching the potential energy surface?

In order to model the potentially large system sizes, the sulfate ion and waters were initially modelled as rigid-bodies interacting through pairwise potentials. A Basin-hopping monte-carlo algorithm[3] with rigid-body translational and rotational moves, and a novel hydrogen-bond cage optimisation move set was used to search for low-lying energy minima. The lowest energy minima were then re-optimised at the density functional theory level, and their vibrational normal modes were computed. We demonstrate that even at the rigid-body potential level, we are able to replicate some of the behaviour of the hydrated sulfate cluster, with simulation predicting that free OH-groups appear around n = 33 waters.

References:

[1] W. Kunz, P. Lo Nostro, and B. W. Ninham, *Curr. Opin. Colloid. In.*, 2004, **9,** 1-18

[2] J. T. O'Brien, J. S. Prell, M. F. Bush and E. R. Williams, *JACS*, 2010, **132,** 7811-7819

[3] D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111-5116

Talk 6

A DFT study of CO_2, H_2O and CO adsorption on Ni/YSZ(111): Solid Oxide Fuel Cell application.

Abdelaziz Essadek, Alberto Roldan, *Nora H. de Leeuw*

Department of Chemistry, University College London, 20 Gordon Street, London, WC1H 0AJ

One of the main causes of global warming is the accumulation of greenhouse gases in the atmosphere by the combustion of fossil fuels. Solid Oxide Fuel Cells (SOFC) could be an alternative to mitigate the use of fossil fuels. SOFC devices convert chemical energy into electrical energy. Particular materials are needed for SOFC since the working temperature is approximately equal to 800˚C. For example, a suitable anode is Ni/YSZ (Nickel – Yttria Stabilized Zirconia = Y_2O_3-ZrO_2) since it has a high catalytic activity, mechanical and chemical stability and compatibility with

electrolyte, [1]. It is well known that the performance of the Ni/YSZ cermet depends on the microstructure and the distribution of Ni and YSZ phases in the cermet, [2]. Additionally, this performance depends on the key reactions occurring at the Triple Phase Boundary where the gas phase, Ni particles and YSZ surface meet. Thus, it is relevant to understand at the atomic scale the structure of Ni/YSZ and its interaction with the gas phase (H_2O, CO_2 and CO).

To simulate the interaction of H_2O, CO_2 and CO with Ni-YSZ, we used Density Functional Theory (DFT) thanks to Vienna Ab-initio Simulation Package (VASP) software. We studied first the deposition of Ni on YSZ(111). Taking the optimised structure of YSZ(111) and Ni/YSZ(111) we adsorbed molecules on both surfaces and paid special attention to the geometric and electronic structure of the most stable adsorption site. We have also analysed the vibrational modes of the three molecules in the gas phase and compared them with the adsorbed molecules. This frequency analysis helped us to understand the influence of the Ni/YSZ(111) surface on the geometry and electronic structure of the three molecules.

Finally, Ni tends to adsorb far from the Yttrium atom and on top of the oxygen vacancy leading to a mixed electronic structure with enhanced charge transfer. Concerning the adsorption of the molecules on the surfaces, H_2O and CO_2 prefer to adsorb on YSZ(111) while CO adsorbs preferentially on Ni/YSZ(111). Thus, Ni seems to play an important role mainly for CO adsorption. A diminution of the strength of the intramolecular bonds was observed upon adsorption. A decrease of the wavenumbers, for the three adsorbed molecules, was observed confirming this influence of the surface on the molecules intramolecular bonds.

References:

[1] Kim, S. *et al.* Ni-YSZ cermet anode fabricated from NiO-YSZ composite powder for high-performance and durability of solid oxide fuel cells, *Solid State Ionics*, (2007), **178**, 1304–1309.

[2] Kim, S. *et al.* Performance and durability of Ni-coated YSZ anodes for intermediate temperature solid oxide fuel cells, *Solid State Ionics*, (2006), **177**, 931–938.

Talk 7

QM/MM analysis of Lysozyme: obtaining novel insights into Biochemistry's 'classic' enzyme

Michael A.L. Limb, *Adrian J. Mulholland*

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

Quantum Mechanics/ Molecular Mechanics (QM/MM) methods are increasingly important in analyzing and predicting enzyme activity. These hybrid methods allow a detailed atomic level investigation of reactions in enzymes by coupling quantum chemical calculations on the active site with a simpler, empirical 'molecular mechanics' treatment of the rest of the protein. This has the significant advantage of probing possible reaction mechanisms in enzymes with quantum methods of potentially high accuracy, while retaining the ability to produce results for large, solvated enzymes, on reasonable time scales and at relatively small computational expense [1].

An example of QM/MM analysis of the effects of mutations, and investigations of alternative substrates, is provided by Hen Egg White Lysozyme (HEWL). HEWL was the first enzyme to have its crystal structure solved, and consequently is one of the most widely studied enzymes in Biochemistry. Despite the wealth of research conducted on the system, key gaps remain in our understanding. Among these essential and still unanswered questions for HEWL are the nature of substrate distortion and the effects of substrate fluorination (which is used in many experiments to probe the mechanism). MD (Molecular Dynamics) simulations, using GPUs (Graphics Processing Units), were performed on the enzyme system, giving insights into the nature and degree of

puckering of sugar ligand when bound in the active site. MD also provided suitable 'reactive frames' for simulating the reaction. QM/MM calculations were then used to determine the nature of the intermediate formed during the enzyme-catalyzed reaction. Reactions of mutant enzymes and alternative (fluorinated) substrates were modeled for comparisons with experiments where such modifications were necessary for the experimental trapping of a reaction intermediate [2]. QM/MM calculations compared the reactions with the wild-type and native substrate, and analyzed the changes caused by these modifications, testing the conclusions drawn from mutant enzymes and non-natural substrates. The results reveal the effects of fluorination may be more complex as has been previously suggested.

References:

[1] van der Kamp M.W., Mulholland A.J., Combined Quantum Mechanics / Molecular Mechanics (QM/MM) Methods in Computational Enzymology, *Biochemistry*, (2013), **52**, 2708-2728

[2] Vocadlo D.J., Davies G.J., Laine R., Withers S.G., Catalysis by Hen Egg White Lysozyme proceeds via a Covalent Intermediate, *Nature*, (2001), **412**, 835- 838

Talk 8

MOARF: Multi Objective Automated Replacement of Fragments

Nicholas C. Firth, Butrus Atrash, *Nathan Brown, Julian Blagg*

Cancer Research UK Cancer Therapeutics Unit, Division of Cancer Therapeutics, The Institute of Cancer Research, London, SM2 5NG, U.K.

Designing compounds to satisfy multiple objectives simultaneously is challenging and liable to subjectivity. Automated design or *de novo* design is a technique used to design compounds that satisfy multiple parameters, which requires minimal guidance if candidate solutions are appropriately scored. *de novo* design has historically suffered from one of two drawbacks, synthetic viability of solutions or poor sampling of the vast chemical space.

The aim of this work is to balance the drawbacks of *de novo* design using fragment replacement approach. A fragmentation algorithm, Synthetic Disconnection Rules (SynDiR), has been developed to enrich the design of synthetically tractable solutions. To permit the greatest possible coverage of chemical space we have designed and implemented a fragment replacement protocol, including a novel pharmacophoric fingerprint for Rapid Alignment of Topological Scaffolds (RATS). SynDiR and RATS have been integrated into a *de novo* design workflow, Multi Objective Automated Replacement of Fragments (MOARF), to enable the design of synthetically tractable molecules for multiobjective optimisation [1].

Both computational validation and prospective experimental validation has been carried out on MOARF. For the prospective validation MOARF has been applied to a previously described medicinal chemistry program targeting potent inhibitors of CDK2 with improved metabolic stability over Seleciclib [2]. For both validation experiments automated design will be described, and for the prospective experiments the synthesis and testing of both CDK2 activity and metabolic stability of a set of novel chemical entities will be reported.

References:

[1] Firth N. C., Brown N., Blagg J., MOARF: an integrated workflow for the multiobjective optimization of drug-like molecules, J. Chem. Inf. Model. (2014) *submitted.*

[2] Aldoss, I. T.; Tashi, T.; Ganti, A. K., Seliciclib in malignancies. *Expert Opin. Invest. Drugs*, (2009) **18**, 1957-1965.

Talk 9

Improving MFS modeling using sequence analysis

Joanna Lee[1], *Zara A. Sands[2]*, Florence Lebon[2], Jiye Shi[2] and *Philip C. Biggin[1]*

[1]Structural Bioinformatics and Computational Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom.
[2]UCB Pharma S.A., Chemin du Foriest, B-1420 Braine-l'Alleud, Belgium.

The Major Facilitator Superfamily (MFS) of transporters is highly diverse; as exemplified by the fact that the highest sequence identity between a pair of solved X-ray crystal structures is 29 % (GLUT1 and XylE). This low level of sequence identity significantly hampers an accurate alignment of novel MFS proteins to existing templates that might be used to build models to rationalize drug design from a structural point of view. Conventional alignment techniques can lead to models where the packing within the transmembrane helices is not optimal.

This research describes a protocol that aims to accurately predict the position of helices in the sequence of novel MFS proteins. Taking inspiration from the numbering of Class A GPCRs as determined by Ballesteros and Weinstein [1], the conservation patterns of hydrophobicity, polarity, as well as glycine and proline residues, were determined in multiple sequence alignments (MSAs) of MFS proteins. These were then used to create a conservation fingerprint that predicts the position of helices in test MSAs.

Analysis of the average lengths of helices in X-ray crystal structures and the position of conserved kinks, which relate to glycine conservation, were incorporated into the fingerprint. The resulting map contains 25 points used to position the 12 MFS helices in the test case to greater accuracy than conventional helix prediction tools. The automation of the system will provide a powerful tool to aid the production of MFS models, supporting research into this large and important superfamily of proteins.

References:

[1] Ballesteros, JA. and Weinstein, H., Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors, *Methods in Neurosci.*, (1995), **25,** 366-428.

Talk 10

Kinetic Model of Peptides: Automatic Construction Using Transferable Basis Sets

Francesca Vitalini[a], *F. Noé[b]* and *B.G. Keller* [a]

[a]Institute of Chemistry and Biochemistry, Freie Universität Berlin, Takustr. 3, D-14195 Berlin, Germany
[b]Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany

Modelling the dynamics of molecular systems via Markov State Models (MSMs) [1] has been of widespread use in recent years, as it allows to identify the interesting dynamical processes and the associated timescales. In MSMs, the dynamics is approximated via a Markov jump process between discrete states in the conformational space. The transition probabilities between states constitute the entries of the MSM transition matrix. The eigenvectors and eigenvalues of the transition matrix contain respectively the dynamical processes of the system and the associated relaxation timescales.

However, the approximation quality of the MSMs strongly depends on how well the discretization of the conformational space describes the energy-landscape features [2]. Finding a good discretization is not trivial for a high-dimensional space.

In this work we exploit a variational approach to MSMs [3] to compute the dominant eigenvalues and eigenvectors of a transition matrix. In this approach the crisp states in the conformational space are replaced by basis functions. We construct basis functions for peptide dynamics as combination of the slow conformational changes of the composing residues. The method correctly identifies the slow kinetic processes and associated relaxation timescales and additionally yields a straightforward interpretation of the kinetic processes in terms of conformational transitions.

References

[1] Prinz, J.H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J.D., Schütte, C. and Noé, F. Markov models of molecular kinetics: generation and validation. *J. Chem. Phys*., **134** (2011), 174105. ISSN 1089-7690. doi: 10.1063/1.3565032.

[2] Sarich, M., Noé, F. and Schütte, C. On the Approximation Quality of Markov State Models. *Multiscale Modeling & Simulation*, **8** (2010), 1154–1177. ISSN 1540-3459. doi:10.1137/090764049.

[3] Nüske, F., Keller, B.G., Pérez-Hernández, G., Mey, A.S.J.S. and Noé, F. Variational Approach to Molecular Kinetics. *Journal of Chemical Theory and Computation*, **10** (2014), 1739–1752. ISSN 1549-9618. doi:10.1021/ct4009156.

Talk 11

Ligand-based drug target prediction in the absence of structural similarity

Daniel Reker, *Gisbert Schneider*

Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland

Computational models allow predicting biomacromolecular targets of drug-like molecules. While the range of applied algorithms is broad, the underlying molecular representations are usually restricted to (sub)structural fingerprints. While these descriptors give accurate predictions according to the chemical similarity principle, they may struggle to both identify bioisosteric substitutions and perform scaffold hops.

We have developed a novel state-of-the-art prediction model based on pharmacophoric patterns and physicochemical properties [1]. Both approaches are workhorses in ligand-based drug design and were combined within a statistical framework for informed target prioritization with sustained sensitivity and markedly reduced false-positive rates.

Prospective studies validated the approach by confirming the target inference made from structurally unrelated ligands. Specifically, surprising *off*-targets for the drugs fenofibrate and fluoxetine have been revealed [1]. The application to new chemical entities (NCEs) retrieved from *de novo* design studies showed the complementary character of such a prediction method in comparison to other, publicly available approaches. In a second prospective study, the method was applied in concert with receptor-based pocket comparisons to identify innovative targets for molecules with a strong antiplasmodial effect [2].

References:

[1] Reker, D., Rodrigues, T., Schneider, P., Schneider, G. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proc. Natl. Acad. Sci. USA,* (2014) **111,** 4067-4072.

[2] Reker, D., Seet, M., Pillong, M., Koch, C. P., Schneider, P., Witschel, M. C., Rottmann, M., Freymond, C., Brun, R., Schweizer, B., Illarionov, B., Bacher, A., Fischer, M., Diederich, F. and Schneider, G. Deorphaning pyrrolopyrazines as potent multi-target antimalarial agents. *Angew. Chem. Int. Ed.,* (2014) **53,** 7079-7084.

Talk 12

Similarity in the Context of Orphan Drug Legislation

P. Franco[1], *N. Porta[2], J. Holliday[1] and P. Willett[1]*

[1] University of Sheffield, Sheffield, UK [2] European Medicines Agency, London, UK,

An orphan drug is a medicinal product that is intended for the treatment of a rare disease that affects only a small number of patients, e.g., five in ten-thousand. According to the current European orphan drug legislation [1], the European Union shall not, for a period of 10 years, accept another orphan medicinal product for the same therapeutic indication, in respect of a similar medicinal product. Thus far, the European Medicines Agency has used human judgments of similarity when assessing new medicines for rare diseases. The study reported here seeks to develop quantitative methods for this purpose.

The project began with an analysis of the correlation between human and computed judgments of similarity for 100 pairs of molecules chosen from the Drug Bank 3.0 database.[2,3] The human similarity assessments for these pairs of molecules were obtained from a total of 143 experts from Europe, Asia and the US, with the experts being asked to state whether each pair was, or was not, similar. The fraction of the experts judging a pair to be similar was then compared with the Tanimoto coefficient computed using a range of different types of descriptors (1D, 2D and 3D), with the aim of identifying those descriptors that correlated most closely with the human judgments.

The following types of fingerprint were studied: ECFP4, ECFC4, Daylight, Unity, BCI, MDL as implemented in the Pipeline Pilot system; and Extended, Standard, Estate, PubChem, MACCS, Morgan, Feat Morgan, Atom Pair, Torsion, RDKit, Avelon, Layers, FP:TGD and FP:TGT as implemented in the KNIME system. The 3D fingerprints studied were the following: FP:TAD and FP:TAT as implemented in the KNIME protocol. 1D molecular property descriptors were also studied but these proved to be of only limited effectiveness for this application. Logistic regression models were developed for each type of descriptor, relating the Tanimoto similarity for a pair of molecules computed with the probability of the human experts would regard that pair as being similar. The resulting regression models were then validated using a separate test-set containing 100 pairs of molecules that had previously evaluated by the European experts in the context of the authorisation of medicines for rare diseases. The best models were able to reproduce over 95% of the human judgments. This success rate was increased to 98% using a simple data fusion approach in which a pair of molecules is classified as similar (or non-similar) when three or more of the individual fingerprints are in agreement.

The results obtained here suggest that computed Tanimoto values based on 2D descriptors could provide a useful source of information when assessing new active substances that are being proposed for the treatment of rare diseases.

References:

[1] Regulation (EC) No. 141/2000 of the European Parliament and Council of 16 December 1999.

[2] Drug Bank, Drug Bank 3.0, 2011, available online at http://www.drugbank.ca/ [Accessed 29 January 2012].

[3] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Branco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo, D. S. Wishart, Nucleic Acids Research (2011) **39**, 1035-1041.

Programme of Poster Presentations

Poster 1	**How many matched molecular pairs are enough?** Iva Lukac, *John Moores University, Liverpool*
Poster 2	**Absolute Binding Free Energy Calculations of Bromodomain Inhibitors** Matteo Aldeghi, *University of Oxford*
Poster 3	**Hydration thermodynamics of small organic molecules using 3D RISM** Maksim Mišin, *University of Strathclyde*
Poster 4	**The Nature of the Bonding in Symmetrical Pincer Palladacycles** Sarote Boonseng, *University of Sussex*
Poster 5	**Exploring P-glycoprotein substrate access** Laura Domicevica, *University of Oxford*
Poster 6	**Molecular dynamics simulations for the study of drug nanoparticles coated with mPEG-PCL diblock copolymers** Ioanna-Danai Styliari, *University of Nottingham*
Poster 7	**Metadynamics Analysis of Potential Host Molecules for Detection of NPS Mephedrone** Kathryn Kellett, *University of Hertfordshire*
Poster 8	**Identification of Preferential Sites for Glucosepane Cross-link Formation in Fibrillar Type I Collagen, Using a Fully Atomistic Molecular Dynamics Approach.** Thomas Collier, *University College London*
Poster 9	**Analysing enzyme specificity through comparison of reaction products: applications in drug metabolism.** Jose Bach Hardie, *European Bioinformatics Group, Cambridge*
Poster 10	**Molecular Dynamics Simulations of Ionic Liquids for CO_2 Capture** Qamreen Parker, *University College London*
Poster 11	**A Random Forest Model for Predicting Allosteric and Functional Sites on Proteins** Ava Sih-Yu Chen, *EaStCHEM School of Chemistry, University of St Andrews*
Poster 12	**The Prediction of Small Molecule Binding Hotspots on Proteins** Chris Radoux, *Cambridge Crystallographic Data Centre, Cambridge*

Poster 1

How many matched molecular pairs are enough?

Iva Lukac, *Andrew Leach*

School of Pharmacy and Biomolecular Sciences, John Moores University, Byrom Street, Liverpool

Since the beginning of rational drug design, compound designers have been facing the same question: "What to make next?". One of the things that has changed is the amount of data available; with an estimated 90% of compounds entering clinical trials failing to reach the market, drug discoverers have been seeking ways to improve their chances of success by exploiting the knowledge generated through decades of experimental work.

One of the ways to decide what to make next (or to predict the properties of a new compound) is by analysing molecules that differ only by a particular, well-defined structural transformation, the so called "matched molecular pairs" [1]. This is because the change of a property is more predictable than the absolute values of that property for either molecule alone. Matched Molecular Pairs Analysis (MMPA) is a technique which aims to identify such matched pairs from a set of compounds and determine the associated property change [2]. The output can then be coded to suggest new molecules together with the probability of each property changing in the desired direction - much

like picking the winning horse by studying its previous results and current race conditions; the decision can be logical and data driven but success cannot be guaranteed.

In this work we examine matched pairs first as a coin flip problem ("heads" meaning property is improved, "tails" property got worse) and aim to detect biased coins, i.e. pairs where the distribution is different from 50:50, indicating that the structural change has a real effect. We analyse how many times we have to see an improvement in a property to be confident the change is real. We then analyse large datasets to determine how many matched pairs we need to observe in order to achieve a selected level of confidence.

References:

[1] Kenny, P.W. and Sadowski, J. (2005) Chemoinformatics in Drug Discovery. pp. 271– 285, Wiley-VCH Verlag GmbH & Co. KGaA

[2] Dossetter, A. G., Griffen, E. J., & Leach, A. G. (2013). Matched molecular pair analysis in drug discovery. Drug Discovery Today, **18**(15-16), 724–31.

Poster 2

Absolute Binding Free Energy Calculations of Bromodomain Inhibitors

Matteo Aldeghi[a,b], *Stefan Knapp*[b], Alexander Heifetz[c], *John J. Barker*[c], *Michael J. Bodkin*[c], *Richard J. Law*[c] *and Philip C. Biggin*[a].

[a] *Department of Biochemistry, Structural Bioinformatics and Computational Biochemistry Unit, University of Oxford, South Parks Road, Oxford. OX1 3QU. UK.*
[b] *Nuffield Department of Medicine, Structural Genomics Consortium, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Headington, Oxford, OX3 7DQ. UK*
[c] *Evotec (U.K.) Ltd., 114 Innovation Drive, Milton Park, Abingdon, Oxfordshire OX14 4RZ, U.K.*

Computational approaches have been increasingly employed for drug discovery in both academia and industry. The ultimate goal of rational drug design is the prediction of the biological activity of a compound, and such activity is driven, at the molecular level, by the specific intermolecular interactions between the small organic molecule, the biological target and the solvent. Similar molecular recognition processes are crucial for many biological functions and the strength of the recognition is characterized by its binding affinity. Thanks to important advances in theory and computing in the last decades, predictions of binding affinities using physics-based simulations are gaining popularity within the computational chemistry community. In particular, binding free energy estimates based on alchemical pathways have been shown to be a rigorous approach for the affinity prediction problem and hold the promise to be able to guide lead optimisation. However, whilst relative calculations have made significant contributions in a drug-discovery context, absolute calculations have so far been mostly applied to model systems despite the advantages related to the ability to estimate affinities for largely diverse sets of molecules.[1]

We present the results of a study where we evaluated the performance of alchemical free energy estimates for a set of drug-like molecules that have been developed to target bromodomains, a family of epigenetic marks readers with established therapeutic potential.[2] The study evaluates the performance of the protocol employed with particular attention to the precision of the calculations and comparing the theoretical results with high-quality isothermal titration calorimetry data. We show that, for this epigenetic target, good agreement between calculations and experiments is achievable even for very challenging compounds, suggesting that alchemical free energy calculations might be approaching the degree of reliability required in order to have an impact in drug discovery campaigns.

References:

[1] Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* (2011), **21**, 150–160.

[2] Filippakopoulos, P.; Knapp, S. Targeting Bromodomains: Epigenetic Readers of Lysine Acetylation. *Nat. Rev. Drug Discov*. (2014), **13**, 337–356.

Poster 3

Hydration thermodynamics of small organic molecules using 3D RISM

Maksim Mišin[a], *David Palmer*[b], *Maxim Fedorov*[a]

[a] *Department of Physics, University of Strathclyde, John Anderson Building, 107 Rottenrow East, Glasgow.*
[b] *Department of Pure and Applied Chemistry, University of Strathclyde, Thomas Graham Building, 295 Cathedral Street, Glasgow.*

Standard free energy, enthalpy and entropy of hydration are important thermodynamic quantities describing the state of a solvated molecule. In our research we have tried to predict them using three new free energy functionals within the 3D Reference Interaction Site Model (3D RISM) theory: UC, NgB, PMVc [1, 2]. Similar studies have been conducted using molecular dynamics, but none have covered the broad spectrum of organic compounds and different temperatures investigated here - a total of 292 compounds with 3194 experimental data points. The method employed is significantly faster than conventional MD with single molecule calculations requiring less than an hour and a half on a single CPU.

We have found that while the accuracy of free energy predictions of empirically fitted models (UC and NgB) is relatively good (RMSE = 1.8 kcal/mol), this result is achieved by the favorable cancellation of errors in enthalpic and entropic terms (with both terms having RMSE > 7 kcal/mol). On the other hand, accuracy of predictions of theoretical model PMVc is similar for all three terms and has RMSE = 3 kcal/mol. All these errors vary greatly among different functional groups, suggesting that there are some systematic errors in the van der Waals parameters of the selected force field and partial charges scheme (GAFF and AM1-BCC).

References:

[1] Palmer, D.; Frolov, A.; Ratkova, E.; Fedorov, M. Towards A Universal Method For Calculating Hydration Free Energies: A 3D Reference Interaction Site Model With Partial Molar Volume Correction. *J. Phys.: Condens. Matter.* 2010, 22, 492101.

[2] Sergiievskyi, V.; Jeanmairet, G.; Levesque, M.; Borgis, D. Fast Computation Of Solvation Free Energies With Molecular Density Functional Theory: Thermodynamic-Ensemble Partial Molar Volume Corrections. *J. Phys. Chem. Lett.* 2014, *5*, 1935-1942.

Poster 4

The Nature of the Bonding in Symmetrical Pincer Palladacycles

Sarote Boonseng, Gavin W. Roffe, *John Spencer* and *Hazel Cox*

Department of Chemistry, School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9QJ

Pincer palladacycles are a type of organometallic compound consisting of a metallic centre with a tridentate ligand. These complexes have a great number of applications, particularly as catalysts (or

precatalysts) in organic synthesis such as the Heck reaction or the Suzuki coupling reaction. In general, the complex is stabilised by the metal-carbon bond avoiding decomposition of the complex while the donor atoms influence the reactivity, stability and performance as a catalyst. In this study we investigate the role of the donor atoms in three symmetrical pincer complexes, referred to as XCX-pincers where X = N, S, or P, i.e., [ClPd{2,6-$(Me_2NCH_2)_2C_6H_3$}], [ClPd{2,6-$(MeSCH_2)_2C_6H_3$}] and [ClPd{2,6-$(Me_2PCH_2)_2C_6H_3$}], in terms of their thermodynamic and kinetic stability. The Atoms-In-Molecules (AIM) theory is used to determine the nature of the bond interactions between the donor atoms of the ligand and Pd. All calculations were performed using DFT at the ωB97XD/6-311++G(2df,2p)[SDD]//PBE/6-31+G(d,p)[SDD] level of theory, where [SDD] is the ECP used for the metal. It is found that the Gibbs free energy of the reaction: pincer ligand + $PdCl_2 \rightarrow$ pincer palladacyle + HCl, is negative indicating that the formation is spontaneous. Along the reaction pathway the barrier for C-H activation is the rate-determining step in all cases. Furthermore, the PCP pincer palladacyle has both the lowest energy barrier to formation and is the most thermodynamically stable. This energetic data is supported by the charge density analysis ($\rho(\mathbf{r})$) at the bond critical points. It is found that the Pd-P bond is stronger than the Pd-S bond, which in turn is stronger than the Pd-N bond. The Laplacian of the charge density ($\nabla^2\rho(\mathbf{r})$) and the total electron density ($H(\mathbf{r})$) indicate that in all cases the bonding has partial covalent character, i.e., $\nabla^2\rho(\mathbf{r}) > 0$ and $H(\mathbf{r}) < 0$.

Poster 5

Exploring P-glycoprotein substrate access

Laura Domicevica, *Philip C. Biggin*

Structural Bioinformatics and Computational Biochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU.

P-glycoprotein (P-gp), an ATP Binding Cassette transporter family membrane protein, extrudes mainly hydrophobic substrates in an ATP hydrolysis-dependant manner. Structurally, P-gp consists of two transmembrane domains (TMD) each comprised of 6 α-helices that together form the drug binding site, and two nucleotide binding domains (NBD) that via ATP binding and hydrolysis provide the energy for the conformational changes necessary to drive drug translocation.

Due to the wide variety of substrates it extrudes, P-gp is one of the main causes of multidrug resistance in cancer and other diseases. As multidrug resistance becomes an ever increasingly important issue, drug development is dependent on knowing what constitutes the difference between a substrate and a non-substrate of P-gp. Although the chemical structure is likely to play a role in recognition within the binding site, it is equally important to understand how access to the binding site is controlled. Therefore more information about way in which substrates access the binding site is of key interest.

In the absence of a human P-gp crystal structure, several human homology models were made from both eukaryotic and bacterial homologue crystal structures. The models were assessed for structural stability and conformational dynamics using molecular dynamics (MD) simulations. Following analysis, the homology model from the *C. elegans* template was chosen as the model for steered MD calculations to investigate possible substrate access pathways.

The results, when combined with bilayer localisation NMR experiments and MD simulations, give novel insights into the factors that govern Pgp substrate specificity.

Poster 6

Molecular dynamics simulations for the study of drug nanoparticles coated with mPEG-PCL diblock copolymers

Ioanna-Danai Styliari [a], *Andrew Theophilus [b], Cameron Alexander [a], Martin Garnett [a] and Charles Laughton [a]*

[a] Centre for Doctoral Training in Targeted Therapeutics and Formulation Sciences, School of Pharmacy, University of Nottingham, University Park, NG7 2RD, Nottingham
[b] Product Development, R&D GlaxoSmithKline, Stevenage

The development of nanoparticles for drug delivery has been - and still continues to be - a very interesting and promising field of pharmaceutical sciences. Unfortunately, nanoparticles used in pulmonary drug delivery formulations frequently fail to reach the lower respiratory tract due to their aerodynamic size and behaviour. One approach to improving their performance is via the production of controlled nanoparticle agglomerates. In order to control the agglomeration, there is the need to understand the molecular structures and interactions that govern the behaviour of the nanoparticles.

In this study we focus on drug nanoparticles whose surface has been coated using monomethoxy poly(ethylene)glycol (mPEG) - polycaprolactone (PCL) diblock copolymers. mPEG-PCL copolymers are biodegradable and can be tailored according to the needs of the system [1]. In order to evaluate their coating efficiency and the effect the coating has over the the formation of the agglomerates, polymers with different hydrophilic (mPEG) /hydrophobic (PCL) molecular weight ratios are tested both experimentally and via molecular dynamics simulations. Using GROMACS and AMBER tools a nano-sized drug particle (indomethacin or fluticasone propionate) and mPEG-PCL in four Mw ratios polymers are tested for their coating efficiency. The simulated system consists of box of an acetone solution of the polymer which surrounds a water droplet containing the nano-sized crystal. The simulation depends on the mixing through diffusion of the acetone-polymer region and the water region. Due to this requirement an extended investigation of the different available acetone models was performed and the PAC model [2] was selected as the most accurate for simulating acetone/water mixed solvent systems.

The results of these simulations are a starting point for the following experimental work and will be compared with them when time comes. Furthermore they are the first step towards the study of the agglomerates formation which will require smaller time scales and larger systems (coarse graining).

References:

[1] Hou, J., Guo, S. *et al. Journal of Biomedical Nanotechnology* (2013**)**, **9**, 231-237

[2] Pinke, A., Jedlovszky P., *Journal of Physical Chemistry B* (2012**)**, **116**, 5977-5984

Poster 7

Metadynamics Analysis of Potential Host Molecules for Detection of NPS Mephedrone

Kathryn Kellett, *Stewart Kirton, Jacqueline L. Stair, Suzanne Fergus,* Mire Zloh

Department of Pharmacy, School of Life and Medical Science, University of Hertfordshire, College Lane, AL10 9AB, United Kingdom.

The aim of this study was to use molecular dynamics to rationally inform the design of small-molecule sensors for the detection of mephedrone, a Novel Psychoactive Substance (NPS). Due to the large number of NPS, the development of a selective sensory device is challenging [1]. A selective host molecule would interact with as much of the target molecules as possible, not just a single functional group. The host molecules fall into two structural categories: calixarenes and

anthraquinones. Metadynamics has been used to investigate the binding affinities between five "host" molecules and mephedrone. In order to demonstrate selectivity towards mephedrone over common contaminant molecules, interactions with amphetamine and caffeine were also studied.

Host molecule structures used for the metadynamics simulations were validated and minimum conformation energies achieved *via* conformational searching and simulated annealing using the Schrödinger software suit. Metadynamics simulations were carried out in an aqueous environment using Desmond [2].

Binding energies were calculated from the metadynamics analysis. These simulations provide evidence on the selectivity of the host molecule with mephedrone and the common contaminant molecules, as well as the type of interactions that are observed. From the simulations performed di-thiourea anthraquinone shows the greatest affinity with mephedrone with a Gibbs free energy of 7.04 kcal mol^{-1}. Caffeine and amphetamine produce minimum system energy 12.26 and 21.02 Å away from the host molecule respectively, suggesting the lowest energy of the system occurs without the host and guest molecules interacting. The anthraquinone class of host molecules show poor selectivity between the guest molecules, with similar binding distances and Gibbs free energies for all three adulterants. The calixarenes show selectivity towards mephedrone, however not at the desired binding site. Metadynamics shows potential to be used in further work for rational design of small molecule binding sites. This information can be used to guide the synthesis of host molecules.

References:

[1] EMCDDA. *European Drug Report Trends and Developments*. 1–88 (2014).

[2] Bowers, K. J. *et al.* Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. in *Proc. ACM/IEEE Conf. Supercomput. (SC06), Tampa, Florida, Novemb.* 11–17 (2006).

Poster 8

Identification of Preferential Sites for Glucosepane Cross-link Formation in Fibrillar Type I Collagen, Using a Fully Atomistic Molecular Dynamics Approach.

Thomas Collier[a], Anthony Nash[a], Andrea F. Lopez-Clavijo[b], Laurent Bozec[c], *Nora De Leeuw[a], Helen L. Birch[b]*

[a] *Department of Chemistry, UCL, London, United Kingdom*
[b] *Institute of Orthopaedics and Musculoskeletal Science, UCL, RNOH Stanmore Campus, London*
[c] *Eastman Dental Institute, UCL, London, United Kingdom*

The functionality of the musculoskeletal system is believed to be jeopardised by glycation and the accumulation of advanced glycation end products (AGEs). Some AGEs are generated by the non-enzymatic reaction of oligosaccharides with proteins in physiological systems. In collagen rich tissues, such as tendons and ligaments, AGEs are believed to form covalent cross-links within and between collagen molecules thereby changing the properties of the tissue. Glucosepane is by far the most abundant AGE crosslink in collagen with levels 100-1000 times higher than all currently known cross-links [1], however little is known about their site of formation. This study aims to identify specific sites involved in forming glucosepane cross-links within the tropocollagen molecule based on a relative energetics model using a proven fully atomistic molecular dynamics approach [2].

A tropocollagen molecule is constructed by combining the sequences of COL1A1 and COL1A2, with details of the supermolecular structure from entry 1Y0F in the Protein Data Bank. A distance based criterion search identifies lysine and arginine residues within 5Å of each other within the molecule. The tropocollagen molecule is placed in a unit cell with dimensions determined by low resolution X-ray diffraction, periodic boundary conditions are applied to the faces of the cell to replicate the dense fibrillar environment. Water molecules and chloride ions are added to the unit cell

before constant pressure molecular dynamic simulations are conducted at pseudo-physiological pH and 37°C using Sander part of Amber12 for 60ns of simulation [2]. A site is a likely candidate for glucosepane formation if the total energy of the tropocollagen molecule is lower in the presence of a bound glucosepane cross-link compared to an unbound glucose molecule.

Of the 24 positions identified based on the distance criteria, only six were found to be energetically favourable (exothermic binding enthalpies) compared to unbound glucose. Some sites potentially have huge implications on the biological function of collagen, as they are within sites where key collagen-biomolecule and collagen-cell interactions occur.

Our model representing a realistic 3-dimensional model of a collagen fibril, has identified six likely sites for glucosepane formation within tropocollagen. The positioning of these sites is likely to have a significant effect on tissue function and integrity.

References:

[1] D.R. Sell, K.M. Biemel, O. Reihl, M.O. Lederer, C.M. Strauch and V.M. Monnier, Glucosepane is a major protein cross-link of the senescent human extracellular matrix. Relationship with diabetes. *J. Biol. Chem. (2005)* **280**, 12310-12315

[2] I. Streeter and N. H. de Leeuw, Atomistic modelling of collagen proteins in their fibrillar environment. *J. Phys. Chem. B*, (2010) **114**, 13263–70

Poster 9

Analysing enzyme specificity through comparison of reaction products: applications in drug metabolism.

Jose Bach Hardie, *John P. Overington*

Overington Group, European Bioinformatics Group, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom

Hydrolisis is a common part of phase I drug metabolism, and involves a limited amount of reactions (generally de-esterification or de-amination). However, even though chemically these reactions are very similar, they are undertaken by a wide variety of enzymes from disparate families [1]. Although multiple enzymes may act upon a given compound, they each do so with different kinetics which reflect their unique specificity. Elucidating these varied kinetics requires careful assays which have been conducted only for a handful of drugs, but they are invaluable for predicting drug interactions and pharmacogenetic effects. This reveals a need for chemoinformatics methods to analyse and estimate the enzymes involved in the metabolism of hydrolised drugs.

For this purpose we have developed a method whereby the products of a given reaction, rather than the enzyme substrate, are used to characterise it. This allows us to inspect cases in which different enzymes act on different reaction sites on the same compound, and we also demonstrate it to have greater discriminating power to separate reactions by enzyme specificity, with consequent applications in machine learning.

References:

[1] Fukami, T., & Yokoi, T., The Emerging Role of Human Esterases. *Drug Metabolism and Pharmacokinetics*, (2012), **27**, 466–477.

Poster 10

Molecular Dynamics Simulations of Ionic Liquids for CO_2 Capture

Qamreen Parker, Dr. Maxime Mercy, *Dr. Rob Bell, Prof. Nora de Leeuw*

Department of Chemistry, University College London, 20 Gordon Street – London – WC1H 0AJ

The climate debate, albeit controversial, provides an actively growing research field. The global atmospheric temperature has increased over the years, [1], and the steady increase of CO_2 levels in the atmosphere due to man-made processes such as fossil fuel burning is no unrelated fact. Thus CO_2 capture and conversion with novel and efficient mediums is an important process in green chemistry. Ionic liquids (ILs), low-temperature melting organic salts, have for over a decade been reported to have high CO_2 capacities and are thus worthy of research as a medium for industrial CO_2 capture.

There are many physical and chemical properties of ILs that are difficult to measure or study through experimental methods. As macroscopic observables are related to microscopic behaviour, simulations can prove useful in bridging this gap, specifically Molecular Dynamics (MD). MD uses predefined force fields to compute the movement of a set of interacting particles, based on Newton's equations of motion. Testing of the force fields used to model ILs and their performance for replicating experimental and previous data is an important start to the larger aim of MD modelling of ILs. An IL specific force field and a general organic force field are compared to determine the differences in capability of modelling these complex systems. Properties such as density and viscosity, along with gas solubility and separation are important to determine for the realistic use of ILs in industrial processes.

The Generalised Amber force field (GAFF), for example, does not precisely describe the long range order seen in ILs (first reported by Maginn et al. in 2002, [2]). The IL specific Canongia Lopez & Padua force field (CL&PFF), [3], however showed a comparable long range order. In addition to bulk simulations of pure ILs, mixtures of ILs with water and CO_2 gas molecules are valuable in determining solubility information. This poster will present some initial results and conclusions on the use of MD for IL research with the ultimate aim of aiding the UK industrial and research communities in understanding the capture and conversion of CO_2 using ILs.

References:

[1] Hansen, J., et al., *Global temperature change.* Proceedings of the National Academy of Sciences, 2006. **103**(39): p. 14288-14293.

[2] Morrow, T.I. and Maginn, E.J., *Molecular Dynamics Study of the Ionic Liquid 1-n-Butyl-3-methylimidazolium Hexafluorophosphate.* The Journal of Physical Chemistry B, 2002. **106**(49): p. 12807-12813.

[3] Canongia Lopes, J. and A.H. Pádua, *CL&P: A generic and systematic force field for ionic liquids modeling.* Theoretical Chemistry Accounts, 2012. **131**(3): p. 1-11.

Poster 11

A Random Forest Model for Predicting Allosteric and Functional Sites on Proteins

Ava Sih-Yu Chen[a], Lazaros Mavridis[b], *Nicholas J. Westwood*[a] and *John B O Mitchell*[a]

(a) EaStCHEM School of Chemistry, University of St Andrews, North Haugh, St Andrews, Scotland KY16 9ST, UK
(b) School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, UK

Allostery is a universal mechanism of regulation to control a protein's activity by the binding of a small molecule to a cleft other than the protein's active site. In contrast to conventional orthostery as a simple on-off device, allosteric modulation as a dimmer switch could offer greater selectivity and fine modulatory control over the level of protein activity [1]. As an advanced therapeutic option, the discovery of new allosteric sites is of interest for discovering promising new leads for drug design.

We used a computational method to identify allosteric sites using a machine learning method trained and tested on protein structures containing bound ligand molecules. The Random Forest machine learning algorithm was adopted to build our three-way predictive model. Based on the descriptors collated for each ligand and binding site, the classification model allows us to assign protein cavities as allosteric, non-functional or orthosteric for allosteric site identification. 43 structural descriptors per complex were derived and were used to characterize individual protein-ligand binding sites as belonging to the three classes, allosteric, non-functional and orthosteric. We carried out a separate validation on a further unseen set of protein structures containing the ligand *2-(N-cyclohexylamino) ethane sulfonic acid* (CHES), and compared these with previous results that had emerged from manual inspection of crystal structures [2].

Our initial results showed that the Random Forest classifier is sensitive to class imbalance. This was resolved by stratified sampling selecting equal sized subsets from each class. For such a dataset, the Random Forest classifier achieves an average accuracy of 69.02% (an error rate of 30.98%). Two structural descriptors, the scoring function RF-Score and its combination with ligand burial scores computed with our novel software PocketFinder, appeared to be important determinant factors in our model's success. Our study suggests that the binding of buffer molecules like CHES could serve to identify locations of possible allosteric sites.

References:

[1] del Sol, A., Tsai, C.J., Ma, B. and Nussinov, R., The origin of allosteric functional modulation: multiple pre-existing pathways, Structure, (2009) **17**, 1042-1050.

[2] Hardy, J.A. and Wells, J.A., Searching for new allosteric sites in enzymes, Curr Opin Struct Biol, (2004) **14**, 706-715.

Poster 12

The Prediction of Small Molecule Binding Hotspots on Proteins

Chris Radoux[a], *Colin Groom[a], Tom Blundell[b], Will Pitt[c]*

[a] *Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge*
[b] *Department of Biochemistry, University of Cambridge, Old Addenbrooke's Site, 80 Tennis Court Road, Cambridge*
[c] *UCB, 216 Bath Rd, Slough*

Hotspots are locations in the protein-ligand or protein-protein binding interface that make a disproportionately large contribution to the binding free energy [1]. Knowledge of hotspot locations can be utilised during multiple stages of the drug discovery process.

Early druggability assessment has been demonstrated by the correlation between fragment hit rate and the ability to find potent binders [1]. Hit prediction may be improved by using the calculated hotspots as anchor points during molecular docking. Similarly, lead optimisation can be aided by highlighting secondary hotspots to target.

A manually curated dataset of 26 protein-fragment complex crystal structures has been built to provide a test set for hotspot prediction methods. Hotspots are typically the only sites able to interact

efficiently enough with fragments to overcome the ligand's limited ability to bind due to the reduced number of interactions available to them.

Full-protein SuperStar maps weighted by solvent accessibility have been generated for 20 crystal structures of thrombin. SuperStar creates interaction maps using IsoStar data and calculates the propensity for a given probe functional group. The aromatic CH probe, uncharged NH probe and carbonyl O probe are used to locate areas with a high propensity for hydrophobic, H-bond donor and acceptor groups respectively.

Averaging the propensity over the ensemble focuses the majority of the propensity at known fragment binding sites. For a given thrombin crystal structure, additional high propensity sites can be seen around the protein, but most are the result of creating the map from a single snapshot and are smoothed out in the ensemble.

Snapshots from Molecular dynamics simulations for each protein in the dataset will be used as input for the calculation of weighted SuperStar maps. The maps will be used to test whether highest regions of propensity are consistently located at the fragment binding site.

References:

[1] Hajduk, P. J., Huth, J. R., & Fesik, S. W.. Druggability indices for protein targets derived from NMR-based screening data, *J med chem*, (2005) , *48*(7), 2518–25.