MGMS and RSC MMG
Young Modellers' Forum 2011
POSTER PRESENTATIONS

Poster 1	**Revisiting the General Solubility Equation: *In Silico* Prediction of Aqueous Solubility incorporating the Effect of Topographical Polar Surface Area** Jogoth Ali, *University of Hertfordshire*
Poster 2	**The Development and Application of a simplified QM/MM method for Free Energy Simulations** Michael Carter, *University of Southampton*
Poster 3	**Visualization and Analysis of Helix Flexibility Using Bendix** Caroline Dahl, *University of Oxford*
Poster 4	**Ligand- and Structure-Based Computational Approaches for the Design of Potent and Selective Compounds towards Adenosine Receptors.** David Rodríguez, *Unilever Centre: University of Cambridge*
Poster 5	**Rapid Prediction and Scoring of Water Molecules in Protein Binding Sites** Gregory Ross, *University of Oxford*
Poster 6	**The Assessment of Computationally Derived Protein Ensembles in Protein-Ligand Docking** Barbara Sander, *University of Southampton*
Poster 7	**Actinyl reduction potentials in solution: An assessment of computational methods to achieve experimental accuracy** Krishnamoorthy Arumugam, *University of Manchester*
Poster 8	**The Relationship between Drug-Lipid Interaction Energy and Non-Specific Binding** Callum Dickson, *Imperial College London*
Poster 9	**Understanding the Competing Reactions of Polyurethane Foam Formation: A Computational Study** Jack Gibb, *Unilever Centre: University of Cambridge*
Poster 10	**Exploring the mycolic acid potential energy surface: the ups and downs** Wilma Groenewald, *Bangor University*
Poster 11	**In silico target predictions using machine-learning methods: Naïve Bayes & Parzen-Rosenblatt Window** Alexios Koutsoukas, *Unilever Centre: University of Cambridge*
Poster 12	**An *In Silico* Scaffold Hopping Protocol for Identifying Novel Kinase Inhibitors** Sarah Langdon, *Institute of Cancer Research*

Poster 1

Revisiting the General Solubility Equation: Evaluating Data Selection and Incorporating the Effect of Topographical Polar Surface Area in Prediction of Aqueous Solubility

Jogoth Ali[a], Patrick Camilleri[b], Marc B Brown[a, c], Andrew J. Hutt[a] and Stewart B. Kirton[a]

[a] *School of Pharmacy, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK*
[b] *Bio-Chemical Solutions, 5 Morgan Close, Stevenage, Hertfordshire, SG1 4TG, UK*
[c] *MedPharm Ltd, Unit 3, Chancellor Court, 50 Occam Road, Guildford, Surrey, GU2 7YN, UK*

The General Solubility Equation (GSE) [1] is a QSPR model based on the melting point and logP of a chemical substance. It is used to predict the aqueous solubility of chemical compounds. However, its reliance on experimentally derived descriptors, particularly melting point, limits its applicability to both novel and virtual compounds. The studies presented show the GSE is able to predict to within 1 log unit of the experimental aqueous solubility (logS) for 81% of compounds in a dataset of 1265 diverse chemical structures ($-8.48 < logS < 1.58$). However, the predictive ability of the GSE is reduced to 75% when applied to a subset of the data (1160 compounds - $6.00 < logS < 0.00$) which discounts those compounds occupying the sparsely populated regions of data space. This highlights how sparsely populated extremities of datasets can significantly skew results for linear regression-based models. Replacing the melting point descriptor of the GSE with a descriptor which accounts for topographical polar surface area (TPSA) produces a model of comparable quality to the GSE (the solubility of 81% of compounds in the full dataset is predicted accurately to within 1 log unit of the experimental aqueous solubility). As such, we propose an alternative simple model for predicting aqueous solubility which replaces the melting point descriptor of the GSE with TPSA and hence can be applied to novel and virtual compounds. Incorporating TPSA into the GSE in addition to logP and melting point gives a three descriptor model that improves accurate prediction of aqueous solubility over the GSE by 5.1% for the full, and 6.6% for the reduced dataset respectively.

References:
[1] Jain, N. and Yalkowsky, S.H., Estimation of the aqueous solubility I: Application to organic nonelectrolytes. *J Pharm Sci*, (2001). **90**(2), 234-252.

Poster 2

The Development and Application of a simplified QM/MM method for Free Energy Simulations

Michael Carter[a], Anny Odlie-Colson[b], Jonathan W. Essex[a]
[a]*School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom*
[b]*Novartis Horsham Research Center, Wimblehurst Road, Horsham, RH12 5AB, United Kingdom*

The estimation of ligand binding affinities using computer simulations has now reached the point where it is no longer limited by underlying free energy methodology, but instead constrained by reliability of the forcefield and sampling. In recent years hybrid Quantum Mechanical

(QM)/Molecular Mechanical (MM) modelling techniques have become a popular tool to combat this problem. Many studies have implemented QM/MM methods to study large biomolecular systems, but it is also often applied to study processes in explicit solvent and to investigate large inorganic/organometallic and solid-state systems. Methodological issues are commonplace due to the varying combinations of MM and QM that can be applied and the complexities of coupling these two different approaches [1].

The hybrid QM/MM approach used here encompasses Density Functional Theory (DFT) and Free Energy Perturbation (FEP) to calculate QM corrections to classical MM free energy calculations [2]. This method applies electrostatic embedding to polarize our QM system with the charges of our MM environment. It also involves no complex coupling between our MM and QM systems, which allows it to perform much faster than other QM/MM methods. Here, this method has been applied to hydration free energy studies of small molecules to establish the applicability of this method to drug design. It has been shown that the QM/MM free energies obtained from these studies show very good agreement with results obtained via MM simulations and with those obtained experimentally. Analysis of outliers has shown that subtle changes in the interactions between our QM ligand and the MM environment can lead to large changes in the QM correction energy obtained. To validate our results we have utilized charge perturbation pathways, where the charges of the ligand are scaled by an arbitrary factor. As free energy is a state function, the resultant MM and QM/MM free energies should cancel to produce a closed free energy cycle.

References:
[1] Senn H. M., Thiel W., QM/MM Methods for Biomolecular Systems, *Angew. Chem. Int.*, (2009), **48**, 1198-1229
[2] Beierlein F., Michel J., Essex J. W., A simple QM/MM Approach for Capturing Polarization Effects in Protein-Ligand Binding Free Energy Calculations, *J. Phys. Chem. B.*, (2011), **115**, 4911-4926

Poster 3

Visualization and Analysis of Helix Flexibility Using Bendix

Anna Caroline E. Dahl, Matthieu Chavent and Mark S.P. Sansom

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Visualization is crucial for understanding ion channel structure-function relationships. With increasing amounts of publicly available crystal structures of larger and more complex molecules, and longer simulation trajectories possible with more computer power, there is an increasing need for software that abstracts and quantifies protein structure dynamics [1].

α-Helices are key structural elements in most proteins, and their flexibility plays a key role in the mechanisms of gating of ion channels and solute transporters. Examples of this can be seen in the voltage-gated potassium channel Kv1.2 and the mechanosensitive channel of small conductance, MscS, where molecular hinges within channel-lining helices are important for gating. Kv1.2 and MscS are not exceptional - membrane proteins often feature curved or kinked helices, and dynamics of such “non-ideal” helices is critical to protein function [2]. However,

non-linear helices are compromised by current visualization techniques, and call for better characterization.

Bendix is a software that colours, quantifies and simplifies both static structures and dynamic (MD) simulation trajectories, without sacrificing details of local conformation. It also enables visualization of secondary structure in coarse-grained simulations of proteins. It is available as a plugin for the molecular graphics program Visual Molecular Dynamics (VMD, http://www.ks.uiuc.edu/Research/vmd) and works under most operating systems.
Bendix offers custom colour schemes that highlight relevant helix curvature and allow you to explore protein structures. Generated helix data is easily presented using the in-built 2D or 3D graphing module, alternatively exported to other graphics packages.
I will illustrate Bendix' features using MscS. Bendix is user-friendly, free and available for download at http://sbcb.bioch.ox.ac.uk/dahl.php

References:
[1] Goddard, T.D. and Ferrin, T.E., Visualization software for molecular assemblies, *Curr. Opin. Struc. Biol.* (2007) **17**, 587–595.
[2] Sansom, M.S.P. and Weinstein, H., Hinges, swivels and switches: the role of prolines in signalling via transmembrane α-helices, *Trends Pharmacol. Sci.* (2000) **21**, 445-451.

Poster 4

Ligand- and Structure-Based Computational Approaches for the Design of Potent and Selective Compounds towards Adenosine Receptors.

David Rodríguez[a,b], Vicente Yaziji[c], Eddy Sotelo[c], *Hugo Gutiérrez-de-Terán[a], Andreas Bender[b]*

[a]Fundación Pública Galega de Medicina Xenómica. Hospital Clínico, Ed. Consultas, planta (-2). Santiago de Compostela, 15706, Spain. [b]The Unilever Centre for Molecular Sciences Informatics, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, UK. [c]Combinatorial Chemistry Unit (COMBIOMED), Institute of Industrial Pharmacy (IFI), Faculty of Pharmacy, University of Santiago de Compostela, 15782, Spain.

Adenosine Receptors (ARs) are membrane proteins belonging to the group of G-Protein Coupled Receptors. They are divided in four subtypes (A_1, A_{2A}, A_{2B}, A_3), and are considered unexploited drug targets, with application in disorders such as inflammatory processes, asthma or Parkinson disease. The recent crystal structures of the A_{2A} adenosine receptor allow reliable Structure-Based drug design on his family of proteins. However, Ligand-Based methods are also appropriate to discover and develop novel compounds with desired characteristics towards these receptors.

First we applied a combination of homology modelling, molecular docking and 3D-QSAR techniques in order to rationalize and grow a series of diaryl-amidopyrimidines with high potency and selectivity towards A_3AR [1]. In parallel we are aiming to obtain novel, potent and selective compounds for each ARs subtype by applying multi-objective selection with the molecular similarity searching tool Molprint 2D [2]. Models of bioactivity are trained with filtered datasets from the ChEMBL database, and are employed to screen both public and proprietary databases.

We obtained a reliable binding mode for the diaryl-amidopyrimidine series in a homology model of A_3AR, and further 3D-QSAR models illustrated the relationships between the main substituents of the ligands for accomplishing the desired activities. Also, the potency and selectivity are rationalized based on differential key binding site residues of ARs [1]. The further synthesis and evaluation of different novel series targeting A_3AR, as suggested by the model, is ongoing with successful results. Here we will also present results on prospective hits (under pharmacological evaluation) towards ARs subtypes obtained by molecular similarity searching. The agreement of the Ligand-Based Virtual Screening towards ARs compared to Structure-Based studies existing in the literature will be discussed.

References:
[1] Yaziji, V., Rodríguez, D., Gutiérrez-de-Terán, H., Coelho, A., Caamaño, O., García-Mera, X., Brea, J., Loza, M.I., Cadavid, M.I., Sotelo, E., Pyrimidine derivatives as potent and selective A_3 adenosine receptor antagonists, *J. Med. Chem.*, (2011), **54**, 457-471.
[2] Bender, A., Mussa, H.Y., Glen, R.C., Reiling, S., Similarity searching of chemical databases using atom environment descriptors: evaluation of performance, *J. Chem. Inf. Comp. Sci.*, (2004), **44**, 1708-1718.

Poster 5

Rapid Prediction and Scoring of Water Molecules in Protein Binding Sites

Gregory A. Ross[a], *Garrett M. Morris[b] and Philip C. Biggin[a]*

[a]*Structural Bioinformatics and Computational Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom.*
[b]*InhibOx, Pembroke House, 36-37 Pembroke Street, Oxford, OX1 1BP, United Kingdom.*

Water is a key structural feature of protein-ligand complexes and plays a critical role in drug binding [1]. It is common practise to design new compounds that displace ordered water molecules to improve ligand affinity and target specificity [2]. Yet it still remains challenging to predict which water molecules are displaceable and what contribution these water molecules will have to the total binding free energy. Crucially, it is important to know the locations of water molecules before and after ligand binding. We introduce a new methodology called WaterDock that within a matter of seconds is able to accurately predict the location of water molecules in *apo* and *holo* binding sites. When validated against high resolution X-ray crystallography and neutron diffraction data, WaterDock predicted 85% of well characterised water molecules and had a false positive rate of 18%. Moreover, WaterDock utilises free available software. WaterDock was also applied to a high quality data set of protein-ligand complexes after the ligands had been removed from the structures. Hypothetically "displaced" water molecules were re-scored using a hydrogen bond function that was calibrated to fit water free energies as calculated by thermodynamic integration methods. We found statistically significant differences in the energy distributions of conserved and displaced water molecules as well with water molecules that were functional displaced and sterically displaced. It was found that water molecules that were displaced by ligands had significantly fewer and weaker hydrogen bonds compared to the water molecules that were conserved. These results should prove useful for

anyone wishing to undertake rational design of new compounds where the displacement of water molecules is being considered as a route to improved affinity.

References:
[1] Lu, Y.; Wang, R.; Yang, C.Y.; Wang, S., Analysis of Ligand-Bound Water Molecules in High-Resolution Crystal Structures of Protein−Ligand Complexes, *J. Chem. Inf. Mod.*, (2007), **47**, 668-675.
[2] Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Et al., Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as HIV Protease Inhibitors, *Science*, (1994), **263**, 380-384.

Poster 6

The Assessment of Computationally Derived Protein Ensembles in Protein-Ligand Docking

Barbara Sander[a], Oliver Korb[b], Jason Cole[b], Jonathan W. Essex[a]

[a]*School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom*
[b]*Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, United Kingdom*

Protein-ligand docking is one of the key steps in modern drug discovery. Accounting for the flexibility of the ligand is common practice; however, the inclusion of receptor dynamics is still at its beginnings and has become a major research interest [1]. One of the possible methods applied is the use of multiple discrete protein conformations, so called ensemble docking, which has been shown to improve docking results in a number of previous studies. With computational techniques like Molecular Dynamics (MD), the limitations imposed by the availability of crystal structures are overcome and a large number of different conformations can be generated. Not all of these can or should be included in the docking or virtual screening process - a matter of quantity vs. quality. The question arises if and how suitable protein conformations can be selected systematically *a priori* based on quantifiable conformational features.

The approach used for this work is based on the "relaxed complex scheme"[2], where snapshots of MD simulations are clustered based on root mean square deviations of binding site residues and only central cluster members as a representative of the particular cluster are used for docking. Here, this method is extended by clustering snapshots of large scale MD simulation trajectories of neuraminidase and cyclin-dependent kinase II (CDK2) based on different structural properties like pocket shape or the positions of binding site residues, using a variety of clustering methods. To establish a possible correlation between docking performance and target conformational attributes, all of the clustered snapshots have been subjected to extensive self- and cross-docking experiments as well as virtual screening using the GOLD docking programme. It is shown that conformationally similar snapshots do not necessarily result in a similar docking or virtual screening performance, i.e. that docking results of members of one cluster can be very different. The selection of the particular structural property on which to base the clustering appears to be the essential problem.

References:
[1] Carlson H. A., Protein Flexibility and drug design: How to Hit a Moving Target, *Curr. Opin. Chem. Biol.*, (2002), **6**, 447-452
[2] Lin, J.-H., Perryman, A. L., Schames, J. R. and McCammon, J. A., Computational drug design accommodating receptor flexibility: the relaxed complex scheme, *J. Am. Chem. Soc.*, (2002), **124**, 5632-5633

Poster 7

Actinyl reduction potentials in solution: An assessment of computational methods to achieve experimental accuracy

Krishnamoorthy Arumugam, *Dr Neil Burton*

School of Chemistry, University of Manchester, Brunswick Street, Oxford Road, Manchester M139PL.

The interest to study the chemistry of actinides by computational methods has been growing everyday; experimental studies are in need of more sophisticated instruments for handling actinides because of radioactive and as well as hazardous nature of actinide elements. The safe disposals of radioactive wastes from nuclear reactors require proper knowledge of their reactivity, redox chemistry, speciation and transport properties in solution. In actinide chemistry the electronic structure, and especially the role of the *f*-electrons, is often crucial to accurately predict reactivity and molecular structure. Understanding the chemistry of actinides at molecular level is very important to prevent any long term damage. Actinides (U, Np and Pu) form stable linear $[AnO_2]^{2+}$ actinyl cations and they bind with a variety of ligands at their equatorial position.

Density Functional Theory (DFT) methods used for the prediction of reduction potentials. Relativistic effects were included by employing relativistic small core effective core potentials (SC-ECP) combined with segmented basis set for the heavy metals and 6-31g(d) basis for other atoms. Solvation is another important factor which was included by using Conductor like Polarizable Continuum Model (CPCM). The predicted solvation energy is sensitive to the choice of solute cavity model. Spin orbit effect was included explicitly as a correction to the final reduction free energies. In this study, $[AnO_2]^{VI/V}$ reduction potentials of $[AnO_2L]^{n+}$ [An=U,Np and Pu] [L=H_2O, OH^-, Cl^-,CH_3COO^-,CO_3^{2-}] complexes in aqueous solution and UO_2L[L=Organic ligands] in non-aqueous solutions (DMSO, DCM, DMF, MeCN and Pyridne) are computationally estimated and compared with experimentally determined reduction potentials. The effects of cavity models, explicit water molecules, and counter ions on the computationally estimated reduction potentials are explored.

Poster 8

The Relationship between Drug-Lipid Interaction Energy and Non-Specific Binding

Callum Dickson;[1] *Lula Rosso;[1,2] Antony Gee;[1,3] Ian Gould[1]*

[1]*Department of Chemistry, Imperial College London, South Kensington, London, SW7 2AZ, UK*
[2]*Department of Neuroscience, Imperial College London, South Kensington, SW7 2AZ, UK*
[3]*Division of Imaging Sciences, King's College London, St Thomas' Hospital, London, SE1 7EH, UK*

Positron emission tomography (PET) scanning is a powerful non-invasive molecular imaging modality, involving the introduction of a radiotracer into the body, which binds to the target of interest allowing a PET image to be obtained. Non-specific binding of radiotracers, whereby tracers bind to cell membranes or other non-target species, is a significant issue with PET radiotracers, as signal from non-specifically bound radiotracer may obscure the target signal creating PET images with poor contrast.

Experimental work has uncovered that drugs interact with lipid molecules when permeating through a cell membrane, a possible mechanism of phospholipidosis [1]. It has been postulated that since drugs interact with cell membranes, the computationally calculated drug-lipid interaction energy may correlate with *in vivo* non-specific binding. This hypothesis holds for a data set taken from the literature [2].

The *ab initio* drug-lipid interaction energy has been calculated for a larger, more consistent data set of twenty-two PET radiotracers; each with *in vivo* non-specific binding values measured in pig using consistent methodology at the GSK Clinical Imaging Centre, with the aim of further validating the correlation between drug-lipid interaction energy and non-specific binding. Faster semi-empirical quantum chemical methods are also assessed (AM1, PM3, PM6, PDDG) for their ability to reproduce *ab initio* results.

Good correlation (r^2=0.80) is found between *ab initio* drug-lipid interaction energy and *in vivo* non-specific binding for those radiotracers which undergo passive transport into the cell from a set of twenty-two PET radiotracers. Semi-empirical quantum chemical methods were unable to reproduce such correlation between drug-lipid interaction energy and *in vivo* non-specific binding.

Insight into the poorly understood phenomenon of non-specific binding may be gained using computational tools as they allow the study of drug-lipid interactions at the atomistic level. Assessment of the predictive power of drug-lipid interaction energy is required and subsequent development may allow a virtual screening tool to be constructed. However semi-empirical quantum chemical methods may not be used in such a tool given their inability to reproduce *ab initio* results.

References:
[1] Baciu, M., Sebai, S.C., Ces, O., Mulet, X., Clarke, J.A., Shearman, G.C., Law, R.V., Templer, R.H., Plisson, C., Parker, C.A. and Gee, A., Degradative transport of cationic amphiphilic drugs across phospholipid bilayers, *Philos Transact A Math Phys Eng Sci* (2006) **364**, 2597 – 2614.
[2] Rosso, L., Gee, A.D. and Gould, I.R., In Silico Prediction of in Vivo Non-specific Binding of Positron Emission Tomography (PET) Radiopharmaceuticals, *J Comp Chem (2008)* **29**, 2397.

Poster 9

Understanding the Competing Reactions of Polyurethane Foam Formation: A Computational Study

J. N. Gibb, *J. M. Goodman*

Unilever Centre For Molecular Science Informatics, Department of Chemistry, Lensfield Road, Cambridge, UK., CB2 1EW

Polyurethane foams are used for a wide range of applications in the modern world, often in structural roles dependent upon their mechanical properties. The relative rates of the foam formation reactions are responsible for many of the foam's final physical properties. Knowledge of the reaction mechanisms would be valuable when trying to predict rates and would reveal any autocatalytic effects, the latter having been suggested by experimental studies. [1]

Investigation of mechanisms using a MM/DFT toolkit suggest uncatalysed gelling and blowing to proceed via concerted ring-containing transition structures involving alcohol and water mediation, respectively. The blow-to-gel catalytic activity ratios of various common commercially available diamines and triamines were explained by consideration of transition structure geometries and energetics.

The kinetics and thermodynamics of various commercial diisocyanates' cyclo-oligomerisation have been investigated to allow prediction of isocyanate consumption with time, and to gain insight into the nature of the isocyanates which may be left unreacted before thermal curing. The mechanism of uncatalysed isocyanate cylotrimerisation to the highly stable isocyanurate product, which introduces much crosslink density into polyurethane foams, has been deduced to preferentially occur via the uretdione kinetic product, acting as a geometric and energetic stepping stone.

Conversely, the superbase proazaphosphatrane [2] has been shown to catalyse isocyanurate formation via stepwise addition of isocyanate to linear activated oligomers. Catalyst transannulation between the phosphorus and basal nitrogen acts to stabilise the positive charge developed upon isocyanate activation and to ensure almost complete isocyanate cyclotrimerisation, providing an accessible route to high purity isocyanurate. The catalytic route to the isocyanurate trimer bypasses the uretdione dimer, in sharp contrast with the uncatalysed system.

References:

[1] Schwetlick, K., Noack, R. And Stebner, F., Three fundamental mechanisms of base-catalysed reactions of isocyanates with hydrogen-acidic compounds, *J. Chem. Soc., Perkin Trans. 2*, (1994), 599-608.

[2] Raders, S.M. and Verkade, J.G., A New Electron Rich Proazaphosphatrane for Isocyanate Trimerization to Isocyanurates, *J. Org. Chem.*, (2010), **75**, 5308-5311.

Poster 10

Exploring the mycolic acid potential energy surface: the ups and downs

Wilma Groenewald, *Anna K Croft*

School of Chemistry, Bangor University, Bangor, LL57 2UW, United Kingdom

Human immunodeficiency virus (HIV) and tuberculosis (TB) are a lethal combination and are taking their toll on developing countries where HIV prevalence is high and public health systems are not efficient. However, even the industrialised world with effective public health systems is at risk of losing control over TB due to the development and spread of extremely drug-resistant strains of TB that are not curable with current TB drugs[1]. Mycolic acids (MAs) form part of the impermeable cell wall of *Mycobacterium tuberculosis*, the bacterium that causes TB. MAs are very long fatty acids, differing at three functional groups to give rise to different classes, namely alpha-, methoxy- and keto-MAs. Mycolic acids alone can mimic certain aspects of TB infection[2]. The various classes are recognised differently by antibodies, and elicit varied inflammatory responses and foam cell formation. Previously, we studied the conformational flexibility of single MAs with molecular dynamics (MD) in vacuum and found that MAs from different classes favour different folding pathways and conformations. In the present study we are extending this work to a comprehensive study of the effect of both intrinsic factors (such as functional group and stereochemistry) and external factors (such as solvent and packing effects) on MA conformations. With quantum mechanics calculations we systematically search conformations of chiral fragments of MAs and find that dispersion forces are important in folding. MD simulations were carried out on varied structures, both in vacuum and in water. Analysis showed similar flexibility in molecules of different subclasses and in different environments. Preliminary conformational analysis shows clearly distinguishable clusters of open, unfolded and compactly folded MA conformations. A course grain model is being optimised for studying the packing of MAs. By using different levels of computational theory, we gain a better understanding of how these factors influence MA conformations and unravel their structure-function relationships.

References:

[1] Russel, G., Barry, C., and Flynn, J. Tuberculosis: What we don't know can, and does, hurt us. *Science* (2010), **328**, 852-856.

[2] Korf, J., Stoltz, A., Verschoor, J., De Baetselier, P., Grooten, J., The Mycobacterium tuberculosis cell wall component mycolic acid elicits pathogen-associated host innate immune responses. *Eur. J Imm* (2005), **35**, 890-900.

Poster 11

In silico target predictions using machine-learning methods: Naïve Bayes & Parzen-Rosenblatt Window

Alexios Koutsoukas[1], Robert Lowe[1], Hamse Y. Mussa[1], John B.O. Mitchell[2], Werner Klaffke[3], Andreas Bender[1a], Robert C.Glen[1b]

[1]*Unilever Centre for Molecular Science Informatics, Chemistry Department, University of Cambridge, Cambridge CB2 1EW, United Kingdom*
[2]*Biomedical Sciences Research Complex and EaStCHEM School of Chemistry, Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Scotland KY16 9ST, United Kingdom*
[3]*Unilever Research Vlaardingen Laboratory, Olivier van Noortlaan 120, NL-3133AT Vlaardinger, The Netherlands*

Data mining methods have been successfully applied to study mechanism of actions, de-orphanize chemical compounds and suggest novel protein-drug interactions [1]. In silico target prediction employs computational methods to predict protein targets of small molecules. These methods are based on the principle of chemical similarity, which states that similar chemical molecules tend to present similar biological activities more often than not [2],[3]. Data mining methods are able to identify patterns in huge chemogenomics datasets and cluster similar molecules in multi-dimensional space that otherwise would be impossible considering the size of the chemogenomics data.

In this work two machine learning algorithms; the Naïve Bayes, a successful and established algorithm in the field of *in silico* target prediction, and the newly proposed Parzen-Rosenblatt Window method were investigated and compared for target prediction using bioactivity data extracted from ChEMBL database. As molecular descriptors the extended connectivity fingerprints (ECFP) were used. In total more than 100,000 molecules covering 894 human protein targets were used in this study, covering broad chemical and biological space. The performance of both methods was measured based on their ability to retrieve and rank active biological targets higher in ordered lists among all possible targets. Furthermore, the use of score cut-offs was explored in this study. Different score cut-offs were applied on both methods in order to measure their selectivity and sensitivity over broad range of scores. Initial findings suggest that using scores present a number of advantages when compared to ranking.

Our future work will focus on developing novel cutting-edge methods for *in silico* bioactivity prediction by improving our current methods. Furthermore, we intend to apply these methods to study bioactivities of small molecules, metabolites and natural products.

References:
[1] Koutsoukas A., Simms B., Kirchmair J., Bond P.J., Whitmore A.V., Zimmer S., Young M.P., Jenkins J.L., Glick M., Glen R.C., Bender A. From in silico target prediction to multi-target drug design: Current databases, methods and applications. *J Proteomics*,(2011), (Article in Press)
[2] A. Bender and R.C. Glen, Molecular similarity: a key technique in molecular informatics, *ORG Biomol Chem **2** (2004), pp. 3204-3218*
[3] Y.C. Martin, J.L. Kofron and L.M. Traphagen, Do structural similar molecules have similar biological activity? (2002), *J Med Chem **45**, pp. 4350-4358.*

Poster 12

An *In Silico* Scaffold Hopping Protocol for Identifying Novel Kinase Inhibitors

Sarah R. Langdon, Nathan Brown, Julian Blagg

Cancer Research UK Cancer Therapeutics Unit, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey, SM2 5NG, UK

Kinases have become one of the most pursued classes of drug target in recent years. As a result of this, kinase inhibitor chemical space has been extensively mined leading to a crowding in intellectual property space. Scaffold hopping is a technique used to identify compounds with similar activity to a known bioactive compound, but with a novel structure.[1] An *In Silico* scaffold hopping protocol, which utilises the Scaffold Tree,[2,3] has been developed to identify novel kinase inhibitors using a known kinase inhibitor as a probe.
An initial scaffold hopping protocol was developed and validated retrospectively. The protocol underwent several different implementations to improve performance and suitability for prospective validation. The optimised protocol represents each compound of a small molecule library with a scaffold generated from the Scaffold Tree.[2,3] This library is interrogated by a probe scaffold derived from a known, well characterised, active compound. Scaffolds, which mimic the probe, are selected using a variety of methods. A diverse selection of compounds represented by the selected scaffolds is then selected for biochemical screening.

A control experiment that uses common virtual screening techniques to select compounds for screening was designed. The control experiment was intended to show whether or not the scaffold hopping protocol performed better than a standard existing virtual screening technique that has previously been shown to find scaffold hops.

The scaffold hopping and control protocols have been validated both retrospectively and prospectively. These validations show that the scaffold hopping protocol has the ability to identify kinase inhibitors with different scaffolds to the probe and also selects more structurally diverse actives than the control protocol, as measured by objective methods. For the prospective validation of the scaffold hopping protocol 98 compounds selected from a library of over 2 million compounds were screened. We have confirmed biochemical activity for 9 compounds against the specific target of interest and X-ray co-crystal structures have been obtained for 5 of these compounds.

References:
[1] Langdon, S. R., Ertl, P. and Brown, N., Bioisosteric Replacement and Scaffold Hopping in Lead Generation and Optimization, *Mol. Inf.*, (2010) **29**, 360-365.
[2] Schuffenhauer, A., Ertl, P., Roggo, S., Wetzel, S., Koch, M. A. and Waldmann, H., The Scaffold Tree - Visualization of the Scaffold Universe by Hierarchical Scaffold Classification, *J. Chem. Inf. Model.*, (2007) **47**, 47-58.
[3] Langdon, S. R., Brown, N. and Blagg, J., Scaffold Diversity of Exemplified Medicinal Chemistry Space, *J. Chem. Inf. Model.*, (2011) ASAP.