MGMS and RSC MMG Young Modellers' Forum 2013 PROGRAMME & ABSTRACTS

Programme of Oral Presentations

9.00 – 9.20	Coffee and Registration
9.20 – 9.30	**Welcome and Introduction**
9.30 – 9.50	**Modeling the Phosphoryl Transfer Transition State in the Oncogenetically Indicated GTPase protein RhoA and it's Activating Protein RhoA.GAP** Whitney Kellett, *Purdue University, Indianapolis.*
9.50 – 10.10	**Revealing the selectivity determinants of ternary GPCR-complexes by homology modeling and molecular dynamics simulations** Ralf Kling, *Friedrich Alexander University, Erlangen.*
10.10 – 10.30	**Combining Computational and Experimental Methods to Understand and Develop Asymmetric Methodology in Organic Chemistry** Matthew Grayson, *University of Cambridge.*
10.30 – 10.50	**A Novel Approach to *De Novo* Design using Reaction Networks** James Wallace, *University of Sheffield.*
10.50 – 11.20	**Tea/Coffee break**
11.20 – 11.40	**Atomic-Scale Insights into New Iron-Phosphate Materials for Lithium Batteries** John Clark, *University of Bath.*
11.40 – 12.10	**Modelling Polymer Nanoprecipitation using a Hybrid Atomistic-Coarse-Grain Force Field** Robert Mackenzie, *University of Nottingham.*
12.10 – 12.40	**Lightning Talks** *All poster presenters, in poster number order.*
12.40 – 14.00	**Lunch and Poster Session**
14.00 – 14.20	**Protein Druggability: the JEDI Approach** Rémi Cuchillo, *University of Edinburgh.*
14.20 – 14.40	**Predicting protein conformational change using a simulation swarm approach** Alessio Atzori, *University of Manchester.*
14.40 – 15.00	**Chemogenomics Approaches in Rationalising Compound Action of Traditional Chinese and Ayurvedic Medicines** Fazlin Mohd Fauzi, *University of Cambridge.*
15.00 – 15.30	**Tea/Coffee break**
15.30 – 15.50	**Interaction Energies in Free Energy Calculations** Christopher Cave-Ayland, *University of Southampton.*
15.50 – 16.10	**Modelling the Photochemical Steps involved in Photodynamic Therapy** Therese Bergendahl, *Heriot Watt University.*
16.10 – 16.30	**Interpretation of statistical machine learning models: application to Ames mutagenicity prediction** Samuel Webb, *University of Surrey.*
16.30 – 16.45	**Judges Deliberations/Networking Session**
16.45 – 17.00	**Prize Presentations**
17.00	**End**

Talk 1

Modeling the Phosphoryl Transfer Transition State in the Oncogenetically Indicated GTPase protein RhoA and it's Activating Protein RhoA.GAP

Whitney F. Kellett and *Nigel G. J. Richards*

Department of Chemistry and Chemical Biology, Purdue School of Science, Indiana University Purdue University Indianapolis, Indianapolis, IN 46202, USA.

GTPase enzymes, which hydrolyze guanosine triphosphate (GTP) to guanosine diphosphate (GDP) and inorganic phosphate (P_i), are involved in a large number of critical cellular processes including proliferation. GTPase Activating Protein (GAP) is responsible for the regulation of GTPase. GTPase proteins, when poorly regulated, can signal for uncontrolled cellular growth and are indicated in oncogenesis [1].

I present a computational model of the GTPase protein RhoA in solution and bound to the regulating protein RhoA.GAP. I have modeled a transition state analog of the GTP to GDP phosphoryl transfer reaction based on the crystal structure of the RhoA:RhoGAP protein complex with GDP and MgF_3^- [2]. Hybrid QM/MM Car-Parrinello MD simulations were performed on the protein complex with the transition state analog, all in explicit water. The reaction was modeled using thermodynamic integration by increasing the terminal phosphate-oxygen bond.

These simulations provide evidence for a dissociative transition state and suggest that MgF_3^- is the best adduct to date to model the transition states of these types of enzymatic phosphoryl transfer reactions. This model could be further exploited to design transition state mimics for many families of GTPase proteins. Finally, this model has shown some structural effects of binding the RhoA.GAP protein to RhoA. These preliminary results indicate that we can understand better the RhoGAP:RhoA protein interface using computational methods. This model could be further exploited to identify ways to dissociate the errant GTPase:GAP complex by targeting the GTP binding site or even the protein-protein interface itself.

References:

[1] Sahai, E. and Marshall, C. J. Rho-GTPases and Cancer. *Nat Rev Cancer*, (2002), **2**, 133-142.

[2] Graham, D. L., Lowe, P. N., Grime, G. W., Rittinger, K., Smerdon, S. J., Gamblin, S. J., and Eccleston, J. F. MgF_3^- as a Transition State Analog of Phosphoryl Transfer. *Chem and Bio*, (2002), **9**, 375-381.

Talk 2

Revealing the selectivity determinants of ternary GPCR-complexes by homology modeling and molecular dynamics simulations

Ralf C. Kling [a, b], *Harald Lanig* [a], *Tim Clark* [a, c], *Peter Gmeiner* [b]

[a] *Department of Chemistry and Pharmacy, Computer Chemistry Center, Friedrich Alexander University, Naegelsbachstr. 25, 91052 Erlangen*

[b] *Department of Chemistry and Pharmacy, Emil Fischer Center, Friedrich Alexander University, Schuhstr. 19, 91052 Erlangen*

[c] *Centre for Molecular Design, University of Portsmouth, King Henry Building, King Henry I Street, Portsmouth PO1 2DY.*

G-protein-coupled receptors (GPCRs) are proteins that enable signal transduction through membranes by activating G-proteins. Despite many investigations, the selectivity determinants of this process on the amino-acid level remain to be discovered. A recent publication on the crystal structure of the β_2-adrenergic receptor in complex with its cognate G_s-protein offers important structural insights into the nucleotide-free ternary signaling complex [1].

We use the β_2-$G\alpha_s$-structure as a template to generate a homology model of the D_2-$G\alpha_i$ complex. For both systems, β_2-$G\alpha_s$ and D_2-$G\alpha_i$, microsecond all-atom molecular dynamics simulations in a hydrated lipid bilayer identify distinct amino-acid contact sites within the receptor-G-protein interface. Investigation of these interfaces by Computational alanine scanning reveals amino-acid hot spot residues that presumably contribute to receptor-G-protein selectivity [2].

References:
[1] Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., Kobilka, B. K., Crystal structure of the beta2 adrenergic receptor-Gs protein complex, *Nature* (2011) **477**, 549-555
[2] Kling, R. C., Lanig, H., Clark, T., Gmeiner, P., Active-State Models of Ternary GPCR Complexes: Determinants of Selective Receptor-G-Protein Coupling, *PLOS ONE* (2013) **8(6)**, e67244

Talk 3

Combining Computational and Experimental Methods to Understand and Develop Asymmetric Methodology in Organic Chemistry

Matthew N. Grayson,[a] Silvina C. Pellegrinet,[b] *Jonathan M. Goodman*[a]

[a]*Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom.*
[b]*Instituto de Química Rosario (CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario (2000), Argentina.*

Of the twenty top-selling prescription pharmaceuticals globally in 2012, twelve were chiral organic drug molecules. Therefore, asymmetric reactions, which are chemical transformations that can selectively produce one enantiomer over the other, are required. As the other enantiomer is considered an impurity, in order for these methods to be cost-effective, highly selective reactions are required to reduce waste generation and to avoid time consuming separation and purification steps. This requires a full understanding of the factors controlling selectivity. Moreover, this understanding enables the rational design of further experimental work, eliminating much trial and error, leading to fewer, more productive reactions and greener chemistry.

My research utilises a complementary computational-experimental approach to gain a clear insight into important asymmetric chemical transformations, with a strong focus on the factors which influence selectivity. Computational tools are used to analyse experimental data from which qualitative models are developed to rationalise experimental observations. These models are reported in an intuitive and accessible format to convey the concept for a broad range of applications to the wider scientific community. Predictions are then made using these models and tested experimentally to verify their accuracy. Knowledge gained from this work can then be applied to the rational design of new molecular transformations in other areas of synthetic chemistry.

The synergistic, computational-experimental approach has led to the correction of existing data reported in the literature. An important asymmetric chemical reaction, the addition of an allylborane to aldehydes, was found to yield a molecule whose modification allows access to a wide range of biologically-active compounds. However, my work found the sense of chirality of the product was wrongly reported.[1]

Other work has examined a reaction type for which the selectivity origins have proved controversial for over a decade, the asymmetric addition of boronates to ketones. Experimental work presented in the literature led to misleading conclusions. However, my computational work aided the understanding of this data and added further weight to the hypothesis that selectivity arises in a different manner to that predicted by experiment.[2] Further work led to the correction of two conflicting selectivity paradigms reported in the literature for the reaction of aldehydes with allenylboranes. Based on this understanding, predictions were provided for a previously untested reaction and were then confirmed experimentally, thus demonstrating the model's value in methodology development.[3]

References:
[1] Grayson, M. N., Pellegrinet, S. C. and Goodman, J. M., Mechanistic Insights into the BINOL-Derived Phosphoric Acid-Catalyzed Asymmetric Allylboration of Aldehydes, *JACS*, (2012) **134**, 2716-2722.
[2] Grayson, M. N. and Goodman, J. M., Lewis Acid Catalysis and Ligand Exchange in the Asymmetric Binaphthol-Catalyzed Propargylation of Ketones, *JOC*, (2013) **78**, 8796-8801.
[3] Grayson, M. N. and Goodman, J. M., Understanding the Mechanism of the Asymmetric Propargylation of Aldehydes Promoted by 1,1′-Bi-2-naphthol-Derived Catalysts, *JACS*, (2013) **135**, 6142-6148.

Talk 4

A Novel Approach to *De Novo* Design using Reaction Networks

James E. A. Wallace[a], Beining Chen[b], Michael J. Bodkin[c] and Valerie J. Gillet[a]

[a]*Information School, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK.*
[b]*Dept. of Chemistry, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.*
[c]*Eli Lilly, Erl Wood, United Kingdom.*

One of the main issues with current *de novo* techniques is the number of possible solutions generated that prove impossible to synthesise in the lab. To retain practicability, some form of sampling needs to be employed to permit more structured searching, such as the reaction vector approach [1]. The reaction vector is effectively a list of the changes that occur in a chemical reaction. Reaction vectors can be applied to unseen molecules containing similar structural features, thereby generating potentially novel molecules with increased likelihood of synthetic accessibility. The method in its original form considers single step reactions only, which can lead to difficulties when built into a multi-step *de novo* design tool. For example, it is often the case that the intermediate structures in a sequence do not resemble the endpoint in terms of shape or functionality. This is problematic for scoring functions that assume smooth progressions from the starting molecule to potential products and can lead to many useful routes being discarded.

The aim of this work is to encode the changes that occur in a reaction sequence in a single vector to enable the intermediates to be bypassed in *de novo* design. The first step is to build a network of sequences from a database of single step reactions [2]. A reaction network can be considered as a

graph in which molecules form the nodes of the graph, linked by edges that represent reactions that transform one molecule to another. By connecting molecules in this manner, every path within the network will represent a reaction sequence. Reaction sequence vectors are then generated by tracing the paths and recoding the differences between the product and start molecules. The vectors can now be applied to an unseen molecule as before, generating the desired product in a single step, without the need to generate intermediates. The reaction networks themselves can also be analysed to identify key molecules and reactions and give insights into preferred chemistries.

References:
[1] Patel, H., Bodkin, M. J., Chen, B. and Gillet, V. J., Knowledge-Based Approach to *De Novo* Design Using Reaction Vectors. *J. Chem. Inf. Model.*, (2009), **49,** 1163-1184.
[2] Soh, S., Wei, Y., Kowalczyk, B., Gothard, C. M., Baytekin, B., Gothard, N. and Grzybowski, B. A., Estimating chemical reactivity and cross-influence from collective chemical knowledge. *Chem. Sci.*, (2012), **3**, 1497-1502.

Talk 5

Atomic-Scale Insights into New Iron-Phosphate Materials for Lithium Batteries

John M. Clark and *M. Saiful Islam*

Department of Chemistry, University of Bath, Bath, UK, BA2 7AY

The discovery and development of new materials are crucial for the next-generation of rechargeable lithium batteries for potential use in portable electronics and electric vehicles. In this context, advanced computational methods combined with structural techniques are now powerful tools for probing the properties of solid-state materials on the atomic and nano-scale. The recently synthesised lithium iron pyrophosphate ($Li_2FeP_2O_7$) provides a new platform for lithium battery research with the highest reported voltage of all known iron-phosphates. Hence it could become a potential competitor to the current $LiCoO_2$ and $LiFePO_4$ electrode compounds. It is known that the investigation of fundamental defect and diffusion phenomena is crucial for greater understanding of the macroscopic behaviour of lithium battery cathode materials. This presentation highlights the recent modelling study of $Li_2FeP_2O_7$ [1] providing detailed insights into defect, dopant and ion migration properties relevant to its electrochemical behaviour, extending related work on novel oxide anode compounds [2]. A significant result is that lithium-ion diffusion is found to be two-dimensional with a low energy barrier, which is important for good rates of charge/discharge.

References:
[1] Clark, J.M., Nishimura, S., Yamada, A. and Islam, M.S., 'High Voltage Pyrophosphate Cathode: Insights into Local Structure and Lithium Diffusion Pathways', *Angew. Chem. Int. Ed.*, (2012) **51**, 13149-13153.
[2] Armstrong A.R., Lyness C., Panchmatia P., Islam M.S. and Bruce P.G., 'The lithium intercalation process in the low voltage lithium battery anode $Li_{1+x}V_{1-x}O_2$', *Nature. Mater*. (2011) 10, 223-229.

Talk 6

Modelling Polymer Nanoprecipitation using a Hybrid Atomistic-Coarse-Grain Force Field

Robert Mackenzie[a], *Charles Laughton[a], Martin Garnett[a], Cameron Alexander[a], Jonathan Booth[b]*

[a]School of Pharmacy, University of Nottingham
[b]AstraZeneca, Macclesfield

Nanoprecipitation is used to encapsulate small drug molecules into polyglycerol adipate (PGA) nanoparticles. These drug delivery systems provide targeting, have slow release properties and reduce off target effects of the drug molecules. However, high drug loading levels are hard to achieve and often there is no pattern to the encapsulation efficiency of a particular polymer-drug combination[1].

We propose the use of molecular dynamics simulations to gain a molecular understanding of the polymer-drug interactions involved. Our model replicates the process of nanoprecipitation by simulating the dispersion of an acetone drop containing PGA in water containing drug. To allow sufficient dispersion of acetone a large amount of water is required, thus coarse-graining becomes mandatory. However, we maintain accuracy for our polymer-drug interactions by using a hybrid force field. Atomistic polymer and drug molecules contain coarse-grain virtual sites which facilitate interactions with the coarse-grain solvent particles[2].

Here we present two hybrid simulations of PGA and a C18 functionalised PGA with dexamethasone phosphate. The accuracy of the hybrid simulations was assessed using atomistic backmapping studies. We also discuss the potential applications of this model for use with different drug-polymer combinations.

References:
[1] Puri, S., Kallinteri, P., Higgins, S., Hutcheon, GA. and Garnett, MC., Drug incorporation and release of water soluble drugs from novel functionalized poly(glycerol adipate) nanoparticles, J. Control. Release, (2008), **125(1)**, 59-67.
[2] Rzepiela, AJ., Louhivuori, M., Peter, C. and Marrink, SJ. Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites, J. Phys. Chem. Chem. Phys., (2011), **13(22)**, 10437-48.

Talk 7

Protein Druggability: the JEDI Approach

Rémi Cuchillo and *Julien Michel*

University of Edinburgh, EaStCHEM school of Chemistry, Joseph Black Building, West Mains Road, Edinburgh, Scotland EH9 3JJ

Several scoring functions have been developed over the last decade to evaluate the druggability -or ligandability- of a protein structure. The majority of existing methods, such as Fpocket, were designed to assess the druggability of crystallographic structures and were not developed to be tightly coupled to molecular dynamics (MD) simulations, in spite of the fact that post-processing of MD trajectories is possible. [1] We present JEDI, a novel computational approach for druggability assessment using a combination of empirical descriptors that can be collected "on-the-fly" during MD simulations.

The Druggable Cavity Directory (http://fpocket.sourceforge.net/dcd) was used to build a data set of 64 diverse proteins in order to parameterize the JEDI scoring function. JEDI is a grid-based approach able to perform the druggability assessment of a binding site in only a few seconds making it one of the fastest methodologies in the field. Agreement between computed and experimental druggability estimates is comparable to literature alternatives. In addition, our estimator is less sensitive than existing methodologies to small structural rearrangements and gives consistent druggability predictions for similar structures of the same protein.

To facilitate evaluation of the druggability of a target at each step of a MD simulation, the JEDI scoring function has been integrated within the PLUMED free energy plugin that supports a broad range of MD packages.[2] Preliminary results show that druggability estimates can be computed for each step of a MD simulation with modest overheads. A unique feature of the approach is that because our druggability function is sufficiently rapid and continuously differentiable, it is possible to: 1) supplement typical potential energy functions with a druggability potential, and 2) exploit information provided by the druggability force to bias molecular dynamics simulations with a variety of free energy methods. Progress towards the identification of cryptic druggable conformations in a range of systems will be reported.

References:
[1] Schmidtke, P. and Barril, X., Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites, *J. Med. Chem.*, (2010) **53**, 5858–5867.
[2] Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., Broglia, R. A. and Parrinello, M., PLUMED: A portable plugin for free-energy calculations with molecular dynamics, *Comput. Phys. Commun.*, (2009) **180**, 1961-1972.

Talk 8

Predicting protein conformational change using a simulation swarm approach

Alessio Atzori, Neil J. Bruce and *Richard A. Bryce*

Manchester Pharmacy School, University of Manchester, Oxford Road, Manchester, M13 9PT, UK

For a growing number of systems, significant protein flexibility appears to be implicated in the execution of biological function. This plasticity also provides opportunities for rational design of selective and potent inhibitors of the protein. Increasingly, computational methods are being applied to predict biomolecular flexibility, although this is a challenging task given the motions involved can be large and occur on time scales generally difficult to achieve with standard simulation methods.
Here, we examine the ability of a method recently developed by us, called swarm-enhanced sampling molecular dynamics (sesMD), to model protein flexibility. The method uses a swarm of interacting replica simulations to improve exploration of conformational space. Specifically, we apply the methodology to examine the plasticity of a protein kinase, p38 mitogen-activated protein kinase (MAPK). We consider the ability of the algorithm to capture fundamental aspects of conformational change in this kinase, and discuss future prospects in relation to structure-based ligand design.

Talk 9

Chemogenomics Approaches in Rationalising Compound Action of Traditional Chinese and Ayurvedic Medicines

Fazlin Mohd Fauzi,[a,b] Alexios Koutsoukas,[a] Rob Lowe,[c] Kalpana Joshi,[d] Tai-Ping Fan,[e] Robert C Glen[a] *Andreas Bender*[a]

[a]*Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, United Kingdom;* [b]*Universiti Teknologi MARA (UiTM) Malaysia, 40 450 Shah Alam, Selangor, Malaysia;* [c]*Blizard Institute of Cell and Molecular Science, Barts and The London School of Medicine and Dentistry, The Blizard Building, 4 Newark Street, London E1 2AT, UK;* [d]*School of Health Sciences, University of Pune, Pune 411007, India;* [e]*Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1PD*

Traditional Chinese medicine (TCM) and Ayurveda have been used in man for thousands of years.[1] While the link to a particular indication has been established in man, the mode-of-action (MOA) of the formulations is relatively unknown. In our recent study,[2] we aim to understand the MOA of formulations used in traditional medicine using *in silico* target prediction, which predicts protein targets (hence, MOAs) given the chemical structure of a compound. We were able to establish several links between suggested MOAs and experimental evidence. In particular, compounds from the 'tonifying and replenishing medicinal' class exhibit a hypoglycemic effect[3] which can be connected to sodium-glucose transporters, SGLT 1 and 2[4], and Protein Tyrosine Phosphatase 1B[5]. Similar results were obtained with Ayurvedic anti-cancer drugs. Here, both primary anti-cancer targets, which directly participate in cancer pathogenesis, *i.e.* steroid-5-alpha-reductase 1 and 2 were predicted, as well as synergistic targets, *i.e.* p-glycoprotein (blocking this efflux pump increases intracellular concentration of primary active ingredient)[6]. Additionally, some targets may point to possible novel MOAs and side effects. Most notably, GPBAR1 which was predicted as a target for both 'tonifying and replenishing medicinal' and anti-cancer classes, suggest an influence of the compounds on metabolism.[7] Understanding the MOA of these compounds is beneficial as it can potentially be developed into drugs, with higher efficacies in the clinic than in the current drug discovery setting. This can be a promising endeavor as the phenotypes of these compounds are well known, indicating both the therapeutic impact and efficacy against a certain disease.

References:

[1] Patwardhan, B.; Gautam, M., Botanical immunodrugs: scope and opportunities. *Drug Discov. Today* **2005,** 10, (7), 495-502

[2] Mohd Fauzi, F.; Koutsoukas, A.; Lowe, R.; Joshi, K.; Fan, T.-P.; Glen, R. C.; Bender, A., Chemogenomics Approaches to Rationalizing the Mode-of-Action of Traditional Chinese and Ayurvedic Medicines. *J. Chem. Inf. Model.* **2013,** 53, (3), 661-673

[3] Zhao, C.-S.; Yin, W.-T.; Wang, J.-Y.; Zhang, Y.; Yu, H.; Cooper, R.; Smidt, C.; Zhu, J.-S., CordyMax™ Cs-4 Improves Glucose Metabolism and Increases Insulin Sensitivity in Normal Rats. *J. Altern. Complement. Med.* **2005,** 8, (3), 309-314

[4] Idris, I.; Donnelly, R., Sodium–glucose co-transporter-2 inhibitors: an emerging new class of oral antidiabetic drug. *Diabetes Obes. Metab.* **2009,** 11, (2), 79-88.

[5] Elchebly, M.; Payette, P.; Michaliszyn, E.; Cromlish, W.; Collins, S.; Loy, A. L.; Normandin, D.; Cheng, A.; Himms-Hagen, J.; Chan, C.-C.; Ramachandran, C.; Gresser, M. J.; Tremblay, M. L.; Kennedy, B. P., Increased Insulin Sensitivity and Obesity Resistance in Mice Lacking the Protein Tyrosine Phosphatase-1B Gene. *Science* **1999,** 283, (5407), 1544-1548

[6] Patel, K.; Tannock, I., The influence of P-glycoprotein expression and its inhibitors on the distribution of doxorubicin in breast tumors. *BMC Cancer* **2009,** 9, (1), 356.

[7] Watanabe, M.; Houten, S. M.; Mataki, C.; Christoffolete, M. A.; Kim, B. W.; Sato, H.; Messaddeq, N.; Harney, J. W.; Ezaki, O.; Kodama, T.; Schoonjans, K.; Bianco, A. C.; Auwerx, J., Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* **2006,** 439, (7075), 484-489

Talk 10

Interaction Energies in Free Energy Calculations

Christopher Cave-Ayland, *Chris-Kriton Skylaris, Jonathan W. Essex*

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

The calculation of free energy differences remains the most rigorous approach to guide the design of novel physiologically active compounds. Due to the extensive degree of sampling required however such calculations require the use of classical potentials to approximate the true potential energy surface of a molecule. Quantum mechanical calculations provide a method to far more accurately represent the potential energy surface however are prohibitively expensive for direct use in free energy calculations. This has led to the development of a number of so called hybrid techniques that attempt to exploit the accuracy of quantum calculations at only a fraction of the computational cost [1][2].

Seemingly inseparable from hybrid free energy techniques is the use of interaction energies within free energy difference estimators, such as the Zwanzig equation. The formal derivation of the Zwanzig equation however is predicated on the use of total potential energies. Although used extensively in hybrid calculations there has been as yet no theoretical justification for the use interaction energies in this context, raising serious questions as to their accuracy. Additionally, the Hamiltonian form that has found widespread use in certain dual topology implementations of free energy calculations is tantamount to the unfettered use of interaction energies. This means that interaction energies are also implicitly used in a wide range of purely classical free energy calculations.

To elucidate the consequences of using interaction energies within the context of hybrid free energy calculations we carry out a theoretical analysis rooted in the underlying statistical mechanics. We reach the interesting conclusion that the equivalence of total and interaction energy calculations is reached only in the hypothetical limiting case of total independence of the inter and intra-molecular degrees of freedom. The use of interaction energies therefore constitutes an approximation, the quality of which varies on a system by system basis dependent on how closely the limit of independence is approached. To ameliorate this situation we propose a simple diagnostic test, based on energy distributions arising naturally in the course of a free energy calculation, capable of highlighting problematic systems.

From this theoretical basis we go on to consider application within a minimal model system. This provides reinforcement for theoretical analysis and confirms the predictive ability of the proposed diagnostic. We then further consider application to a mutation between two inhibitors of the enzyme cyclooxygenase 2, demonstrating the practical applicability of this work in the context of a calculation in a realistic protein-ligand system.

References:

[1] Warshel, A. and Levitt, M., Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabiliziation of Carbonium Ionin the Reaction of Lysozyme, *J. Mol. Bio.*, 1976, **103**, 227-249

[2] Beierlein, F. R., Michel, J. and 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

Talk 11

Modelling the Photochemical Steps involved in Photodynamic Therapy

L. Therese Bergendahl and *Martin J. Paterson*

Institute of Chemical Sciences, Heriot Watt University, Edinburgh EH14 4AS, Scotland

The importance of a deeper understanding of light-induced processes is particularly pertinent when considering the medical application of photodynamic therapy (PDT). In PDT a photosensitiser chromophore located in a cell is photoexcited and subsequently, through an energy-transfer process, generates the reactive and cytotoxic singlet oxygen species, $O_2(a^1\Delta_g)$. This process presents a potential gateway to the induction of cell death via a combination of chemotherapy and radiotherapy, free from many of the harmful side effects associated with either.

General PDT has developed around the use of photosensitiser molecules based on the family of *porphyrin* macrocyclic compounds. As the field of computational chemistry has developed, so has the ability to model the ground state potential energy surfaces and reaction profiles for compounds such as these. However, the modelling of the photochemistry of structures of this size is computationally demanding, and the methods and tools to achieve this have only recently become available [1].

We discuss the challenges involved in the modelling of the various pathways involved in the energy flow from the excitation of the photosensitiser to the sensitisation of singlet oxygen. We show how linear and quadratic response theory can be used to predict the one-photon absorption (OPA) and two-photon absorption (TPA) properties of new photosensitiser chromophores [2]. We also present the challenges associated with the construction of wavefunctions that characterise the excited states and to model the non-adiabatic pathways available to these systems. These include the elusive energy-transfer processes that ultimately generate singlet oxygen.

References:
[1] Bergendahl, L.T. and Paterson, M.J., Computational Modelling of the steps involved in photodynamic therapy, in Molecular Photochemistry, edited by S. Saha, In-Tech (Reijka), (2012) **Ch. 8**, 161
[2] Bergendahl, L. T. and Paterson M.J., Two-Photon Absorption-Molecular Structure investigation using a porphycene chromophore with potential in photodynamic therapy, J. Phys Chem B. (2012) **116**, 11818

Talk 12

Interpretation of statistical machine learning models: application to Ames mutagenicity prediction

Samuel J. Webb[ab], Thierry H. Hanser[b], *Brendan Howlin[a], Paul Krause[a], Jonathan D. Vessey[b]*

[a] *University of Surrey, Guildford, Surrey, GU2 7XH*
[b] *Research Group, Lhasa Limited, 22-23 Blenheim Terrace, Leeds, West Yorkshire, LS2 9HD*

Highly accurate predictions should be a goal of any (Q)SAR modeller; however, these models are often achieved to the detriment of the interpretation of the result. Ames mutagenicity is a popular endpoint given its regulatory importance but many existing models are 'black boxes' that deprive the user of an explanation or the ability to interpret or assess the prediction.

We have developed a new methodology for the interpretation of black box predictions which is: agnostic to the learning algorithm used and agnostic to descriptor choice. The current implementation assumes that the activity is caused by the presence of a feature, inactivity by the absence of a feature or by a features deactivation and that the classification (target variable) is binary.

A feature network is built based on the enumeration of a query structure or its feature vector. Each node in the network is assigned a prediction from the model allowing for the elucidation of the models behaviour towards the prediction for a given query structure. The resultant networks are often large; visualisation and manual assessment strategies can become difficult for complex structures. An assessment algorithm has been implemented that allows for interpretation by assessing the model's predictions for each node by 1 of 6 rules which utilise the position in the network and the prediction from the model.

The developed algorithms have been incorporated into KNIME [1] nodes to allow access to a variety of different machine learning algorithms, descriptors and chemical engines. Components of the ChemAxon toolkit [2] and our own in house chemical engine have also been incorporated to provide various cheminformatics functionality.

The interpretation methodology has been applied to Support Vector Machine, Random Forest and k Nearest Neighbour models for the prediction of Ames mutagenicity. The developed models perform well on public datasets (c. 82% on external validation studies) and when coupled with the interpretation algorithm provide meaningful interpretation in the form of structural motifs causing activity (or the lack thereof).

References:
[1] Berthold, M.R., Cebron, N.C., Dill, F., Gabriel, T.R., Kotter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K. and Wiswedel, B., KNIME: The Konstanz Information Miner, Studies in Classification, Data Analysis, and Knowledge Organization, (2007), Springer
[2] ChemAxon Java API v5 (http://www.chemaxon.com)

Programme of Poster Presentations

Poster 1	**Targeting EB1-SxIP proteins complex: an integrated approach** Teresa Almeida, *University of Liverpool.*
Poster 2	**Using 3D homology models to explore the evolution of enzyme function** Rosanna Alderson, *University of St Andrews.*
Poster 3	**Modelling Zinc In p53: A Comparison of Bonded Zinc Models** Michael Criddle, *University of Southampton.*
Poster 4	**An approach for free energy calculations from large-scale DFT total energies** Chris Sampson, *University of Southampton.*
Poster 5	**Probing the outer membrane of Pseudomonas aeruginosa using molecular dynamics simulations** Jamie Parkin, *University of Southampton.*
Poster 6	**Predicting Porous Molecular Crystals** Jonas Nyman, *University of Southampton.*
Poster 7	**Towards a Complete Computational Site-Directed Mutagenesis Protocol** Pietro Aronica, *Imperial College London.*
Poster8	**Molecular Mechanics Force Field Parameterisation: Comparing QM and Structure-Based Parameters for Small Molecules** Florian Roessler, *University of Cambridge.*
Poster 9	**Application of QM/MM Methods to probe reaction mechanisms** Michael Limb, *University of Bristol.*
Poster 10	**Efficacy of Chemical Hyperstructures in Similarity Searching and Virtual Screening** Edmund Duesbury, *University of Sheffield.*
Poster 11	**Filling the Bioactivity Spectra of Serine Protease Inhibitors using Proteochemometric Approaches** Qurrat Ain, *University of Cambridge.*
Poster 12	**Structure-Based Drug Discovery (SBDD) using an NMR ensemble: the case of S100P** Rama Camara, *University of Hertfordshire.*

Poster 1

Targeting EB1-SxIP proteins complex: an integrated approach

Teresa Almeida[a], *Andrew Carnell[a], Igor Barsukov[b], Neil Berry[a]*

[a]Department of Chemistry, University of Liverpool L69 7ZD
[b]NMR Centre for Structural Biology, University of Liverpool L69 7ZB

Targeting protein-protein interactions (PPI) using small molecules is currently one of the major challenges in Drug Discovery [1]. The surface of interaction between two proteins is usually large, flat or moderately convex and cannot be covered with a small molecule with drug like properties. Unlike other drug targets like enzymes, G-protein-coupled receptors and ion channels there is no convenient natural substrate to act as a starting point for small molecule design [1]. Although, the existence of hot-spot regions as major contributors to PPI is a tool that can be used as an initial approach to find modulators for these interactions [1].

End binding (EB) protein family members bind to microtubule growing ends and mediate the binding of other microtubule plus end tracking proteins (+TIPs). EB1 is considered as a key regulator of dynamic +TIPs interaction networks at growing microtubule ends. The main aim of this project is to design, synthesize and evaluate novel PPI modulators of the EB1-SxIP complex. Inhibiting the formation of this complex, we can possibly stop microtubule polymerisation and therefore control important cell functions coordinated by these celular polymers.

We have used both ligand- and structure-based strategies in order to identify compounds which can modulate this PPI.

Pharmacophore information from natural ligands has been incorporated in the virtual screening process. Selection of molecules has used a multiobjective analysis, considering the results from molecular docking and other parameters such as aqueous solubility and predicted hydrogen bond contacts with the target.
The hydrogen bonding contacts predicted to occur between the hit molecules and EB1 were confirmed through biophysical assays performed using EB1 and natural ligands, giving excellent support forthe computational approach.

Currently, we are validating the virtual screening hit compounds through NMR and ITC tecniques. Structural information obtained will then be used to improve our further computational approaches.

References:
[1] Koes, D. R., Camacho, C. J., Small molecule inhibitor starting points learned from protein-protein interaction inhibitor structure, Bioinformatics, (2012) 28, 1951

Poster 2

Using 3D homology models to explore the evolution of enzyme function

Rosanna G. Alderson[a], Daniel Barker[b], *John B.O. Mitchell*[a]

[a] *Biomedical Sciences Research Complex and EaStCHEM School of Chemistry, University of St Andrews, Scotland KY16 9ST*
[b] *Sir Harold Mitchell Building, School of Biology, University of St Andrews, St Andrews, Scotland, KY16 9TH*

Understanding the evolution of enzyme function and in particular enzyme catalysis has far reaching consequences – for evolutionary theory in general – but also for enzymology and medically relevant associated problems.

As one might expect, enzymes evolve new functions over vast time scales, spanning large parts of evolutionary history. Therefore, tracking this evolution using only extant information, such as protein and nucleotide sequences, is a real challenge. The problem lies in the fact that evolutionary signal of relatedness – homology – is gradually buried in the plethora of mutations that proteins acquire during evolution. However, it is also known that structure tends to be more conserved in evolution than sequence, that is, two proteins that are descended from a common ancestor may have a low level of sequence similarity but still share detectable structural similarities [1].

Such structural information can be used to map functional evolution along with phylogenetic methods. In fact, we can take this further, using homology modelling to infer the likely functions of ancestors of interesting enzymes such as those involved in antibiotic resistance. In this work, we employed this strategy of homology modelling and subsequent alignment to pre-existing catalytic

site templates to ascertain number of times β-lactam hydrolysis activity had evolved in the metallo-β-lactamase superfamily. After predicting100 ancestral sequences of the metallo-β-lactamases from a variety of possible phylogenetic trees it was found that many of the 100 ancestral predictions, although with sequence signals indicative of close homology to extant metallo-β-lactamases, did not have predicted structures with locations of key catalytic residues that were compatible with metallo-β-lactamase functionality. This could be taken as evidence towards multiple independent evolutionary inventions of metallo-β-lactamase function, however, due to our lack of knowledge of the past evolutionary past of these enzymes, we feel the evidence is equivocal.

Another interesting group of enzymes, the nucleophilic halogenases, have relatives with functions that are poorly understood and annotated. It is hoped that much like the metallo-β-lactamase superfamily homology modelling and structural alignments will be helpful in discerning and distinguishing functional groups within this superfamily.

References:
[1] Chothia, C., Lesk., A., The relation between the divergence of sequence and structure in proteins, *EMBO J* (1986), **5**, 823–826.

Poster 3

Modelling Zinc In p53: A Comparison of Bonded Zinc Models

Michael P. Criddle[a], Zohra Ouaray[a,b], *Jonathan W. Essex*[a], *Chandra Verma*[b]

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

The p53 protein is a tumor suppressor protein that causes apoptosis upon cellular stresses, so it is an integral cancer prevention mechanism. However, deactivating mutations in p53 are common, observed in as many as 50% of all human cancers[1]. Understanding the dynamics of p53 in its wild-type form as well as its mutants, is essential if therapeutic methods to either stabilise the wild-type, or restabilise the mutant forms of p53, are to be developed.

Historically, molecular dynamics (MD) simulations on the p53 system have featured both bonded and non-bonded models to parameterise the zinc ion located in the core DNA binding domain (DBD) of the protein. The studies using the bonded models feature the parameters derived by Ryde (1995)[2]. However, these parameters were developed for use at a catalytic zinc site with 4 or 5 ligands at the coordination site. More recently, a Zinc AMBER Force Field (ZAFF)[3] has been developed based on tetrahedral zinc sites with various coordinating ligands, such as those found in p53. Because of the broader range of zinc coordination sites parameterised in ZAFF, and the importance of p53 as a target, a direct comparison of the Ryde and ZAFF force fields is important.

Five independent atomistic MD simulations were run on the wild-type p53 for 500 ns at 298 K, resulting in 2.5 μs of trajectory data for each model. The local environment of the zinc was analysed over the ten simulations to look for any significant difference in the zinc coordination between the models. Root mean square deviation, root mean square fluctuation, principal components analysis, and cluster analysis were performed to look for large changes in the global DBD structure.

The results show key areas of protein fluctuation are conserved over the simulations using both models. Apart from some subtle differences, the overall performances of the two zinc models are similar. It is suggested, however, that the ZAFF model is more valid when considering the p53 DBD system as the parameters were derived from similar coordination sites.

References:
[1] Joerger, A. C., Fersht, A. R., Structural Biology of the Tumor Suppressor p53, *Annu. Rev. Biochem.*, (2008), **77**, 557-582.
[2] Ryde, R., Molecular Dynamics Simulations of Alcohol Dehydrogenase With a Four- or Five-Coordinate Catalytic Zinc Ion, *Proteins: Struct., Funct., Genet.*, (1995), **21**, 40-56.
[3] Peters, M. B., Yang, Y., Wang, B., Füsti-Molnár, L., Weaver, M. N., Merz, Jr., K. M., Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF), *J. Chem. Theory Comput.*, (2010), **6**, 2935-2947.

Poster 4

An approach for free energy calculations from large-scale DFT total energies

Chris Sampson[a], Christofer Tautermann[b], Thomas Fox[b], Christopher Woods[c], *Chris –K. Skylaris*[a]

[a]*School of Chemistry, University of Southampton, University Road, Southampton, Hampshire, SO17 1BJ, United Kingdom*
[b]*Computational Chemistry, Dept. of Lead Discovery, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany*
[c]*School of Chemistry, University of Bristol, Cantock's Close, Bristol, Somerset, BS8 1TS, United Kingdom*

In an ideal world all binding free energies would be calculated using quantum mechanics. However, for biologically relevant systems this is too computationally expensive to be feasible. Because of this classical mechanics is traditionally used, but the accuracy of the binding free energies is then dependent on how well parameterised the force field is for that system.

By making use of the extended thermodynamic cycle [1] we can perturb from a classical ensemble to a quantum ensemble. In traditional methods interaction energies are used for this perturbation, but this is an approximation and total energies should be used according to statistical mechanics. However, when using total energies, we find that the quantum and classical phase spaces often don't overlap well, and because of this we find a lack of convergence in our quantum-corrected free energies. These initial results show that we cannot simply accept every classical structure as a good representation of a structure in a quantum potential. Instead an acceptance criterion must be applied [2].

We demonstrate an approach to generate an NVT quantum ensemble from an MD simulation with a classical potential in such a way that it obeys detailed balance condition. Using such an approach we can use rigorous thermodynamic methods (such as TI) to calculate the free energy of mutating the biomolecular system from classical to quantum.

Our ultimate goal is to use this approach with large-scale DFT calculations with the ONETEP program.

References:
[1]Štrajbl, M., Hong, G., Warshel, A., Ab Initio QM/MM Simulation with Proper Sampling: "First Principle" Calculations of the Free Energy of the Autodissociation of Water in Aqueous Solution, *J. Phys. Chem. B*, (2002), **106**, 13333-13343
[2]C.J. Woods, F.R. Manby, A.J. Mulholland, J. Chem. Phys., **2008**, 128, 014109
Woods, C.J., Manby, F.R., Mulholland, A.J., An efficient method for the calculation of quantum mechanics/molecular mechanics free energies, *J. Chem. Phys.*, (2008), **128**, 014109

Poster 5

Probing the outer membrane of Pseudomonas aeruginosa using molecular dynamics simulations

Jamie Parkin, *Syma Khalid*

University of Southampton, Southampton, Hampshire, SO17 1BJ

Pseudomonas aeruginosa (PA) is a bacterium of particular interest due to increasing antibiotic resistance. PA is a pathogen that can be fatal in persons with compromised immune systems. Increasing resistance to current antibacterial agents has generated a requirement to develop not only novel methods in drug design, but also a greater understanding of the drug delivery process. The resistance of PA, in part, is due to a complex outer membrane (OM) arrangement, which has a low permeability. Drug transport across the outer membrane is mediated by beta-barrel proteins, which exhibit substrate specificity. It is this substrate specificity that provides the innate defense mechanism. Two methods have been used to help understand the barriers that these substrates face.
The research of Eren et al. comments upon the specificity of the some Occ family proteins during mutation studies, as well as a look at the OccD1 protein using molecular dynamics simulations [1]. We present here, using molecular dynamics (MD) and steered molecular dynamics (SMD), an increased understanding of the arginine transport pathway through the PA outer membrane protein, OccD1. Specific binding events thought to be key to transport have been observed; such as OccD1 loop movements and hydrogen bonding network interactions.

Previously, Bemporad et al. have calculated the energetic profiles for small molecules across simple symmetric lipid bilayer systems [2]. To further understand the barrier that membranes impose, we have performed free energy calculations on a range of substrates across the PA OM to determine the energy landscape that must be overcome for permeation. The results provide a key step towards developing novel antibiotics for PA.

References:
[1] Eren E. et al, Toward Understanding the Outer Membrane Uptake of Small Molecules by Pseudomonas aeruginosa, *J. Biol. Chem.*, (2013), **288**, 12042-12053.
[2] Bemporad, D and Essex, J. W., Permeation of Small Molecules through a Lipid Bilayer: A Computer Simulation Study, *J. Phys. Chem. B,* (2004), **15**, 4875-4884.

Poster 6

Predicting Porous Molecular Crystals

Jonas Nyman, *Graeme M. Day*

Department of Chemistry, University of Southampton, University Road, Southampton SO17 1BJ

To be able to predict the crystal structure(s) of any given molecule has been a long standing goal in chemistry. The failure of early attempts made some claim it was impossible [1], and a “continued scandal” in science [2].

The last decade has seen dramatic improvements in the theories and computer algorithms underlying computational Crystal Structure Predictions. It is now possible to reliably obtain the most likely crystal structures of at least simple molecules starting for nothing more than a drawing of the molecule.

We can now go even further and look for rare and exotic kinds of crystals such as zeolitic molecular crystals and inclusion compounds among our predictions and calculate their physical properties, paving the way for the "science of hypothetical materials".

In our poster, we present novel results on the prediction of porous crystals and inclusion compounds, using two fluorophenols as examples. We have performed crystal structure predictions by global lattice energy searches on *ortho*- and *meta*-fluorophenol. The predicted structures have then been analysed for porosity and their likelihood of being stable inclusion compounds. From the hundreds of predicted structures, we have identified a handful of hypothetical crystals with properties suggesting that they are indeed able to host small guest molecules within, allowing them to act as gas storage vessels, molecular sieves and more…

Preliminary results from solid state NMR indicate that we have successfully predicted a previously unknown porous crystal of *ortho*-fluorophenol.

References:
[1] Gavezzotti, A., Are Crystal Structures Predictable?, *Acc. Chem. Res.*, (1994), **27**, 309-314.
[2] Maddox, J., Crystals from first principles, *Nature*, (1988), **335**, 201.

Poster 7

Towards a Complete Computational Site-Directed Mutagenesis Protocol

Pietro 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.

Poster 8

Molecular Mechanics Force Field Parameterisation: Comparing QM and Structure-Based Parameters for Small Molecules

Florian D. Roessler[a], *Oliver Korb* [b], *Peter J. Bond*[a,c]

[a] *Unilever Centre of Molecular Science Informatics, Department of Chemistry, Cambridge, UK*
[b] *Cambridge Crystallographic Data Centre, Cambridge, UK*
[c] *Bioinformatics Institute (A*STAR), 30 Biopolis Str, #07-01 Matrix, Singapore 138671*

A large variety of experimental and theoretical methods are presently used during novel drug development. In principle, molecular dynamics (MD) simulation approaches may facilitate this, helping to understand the key interactions, energetics, and associated dynamics involved in complex formation between proteins and small-molecules. A major challenge in such simulations is the parameterisation of novel small molecules. Thanks to successive parameterisation and optimization over recent years, current high-quality force fields are able to accurately reproduce the conformational dynamics of solvated proteins. In contrast, the chemical space associated with small molecules is vast, hampering accurate parameter derivation. Current methods focus primarily on using *ab initio* quantum mechanical (QM) based methods, as well as using analogy with previously derived parameters of chemically related functional groups. Whilst these methods can produce reliable parameters, they tend to be slow, as well as being highly dependent upon the level of theory used to derive them. Moreover, they are subject to QM-related limitations, such as a lack of consideration of explicit solvent effects and the derivation of inaccurate intermolecular distances due to gas phase calculations. Therefore, we are tackling the need for accurate, rapid generation of structural force field parameters for MD by taking an experimental structure-based approach. We have used the Cambridge Structural Database (CSD) to extract structural geometry preferences of small organic molecules. These are compared to parameters from two major QM-parameterised force fields (CHARMM'S CGENFF[1] and AMBER'S GAFF[2]). Additionally, simulation based parameter distributions are compared to distributions from the CSD to highlight possible impact of solvent effects on structural parameters. For further comparison, liquid densities and heats of vaporisation are calculated for a variety of small molecules, in order to thoroughly assess the quality of structure-based optimised parameters vs. purely QM-derived parameters in the context of thermodynamic properties.

References:
[1] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.," *J. Comput. Chem.*, vol. 31, no. 4, pp. 671–90, Mar. 2010.
[2] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. a Case, "Development and testing of a general amber force field.," *J. Comput. Chem.*, vol. 25, no. 9, pp. 1157–74, Jul. 2004.

Poster 9

Application of QM/MM Methods to probe reaction mechanisms

Michael A.L. Limb, Adrian J. Mulholland

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

QM/MM (Quantum Mechanics/ Molecular Mechanics) methods are increasingly important in analysing 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 HEWL (Hen Egg White Lysozyme). GPU (Graphics Processing Unit) accelerated, long timescale MD (Molecular Dynamics) simulations were performed on the enzyme system allowing new insights into the nature and degree of puckering of sugar ligand when bound in the active site and suitable 'reactive frames' to be generated for the reaction. QM/MM calculations where then used to determine the nature of the catalytic intermediate formed during the enzyme-catalysed reaction. Reactions of mutant enzymes and alternative (fluorinated) substrates were then modelled, for comparisons with experimental studies: 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 analysed the changes caused by these modifications, testing the conclusions drawn from mutant enzymes and non-natural substrates.

References:
[1] Ranaghan K.E., Mulholland A.J., *Int. Rev. Phys. Chem.,* 2010, **29,** 65-133.
[2] Vocadlo D.J., Davies G.J., Laine R., Withers S.G., *Nature*, 2001, **412**, 835- 838.

Poster 10

Efficacy of Chemical Hyperstructures in Similarity Searching and Virtual Screening

Edmund Duesbury, *John Holliday, Peter Willett*

Information School, University of Sheffield, 211 Portobello, Regent Court, Sheffield, South Yorkshire S1 4DP

Two techniques exist in data fusion which have been proven to work in various forms in chemoinformatics: similarity fusion, where different similarity measures are combined; and group fusion, where similarities are combined from multiple reference molecules. The hyperstructure concept however is another form of data fusion, being a hypothetical moleculethat is constructed from the overlap of a set of existing molecules. Initially proposed to reduce the time of database searching, it has also been used directly for virtual screening on two occasions since its inception [1,2], the latter of which showed it to be useful as a 2-dimensional QSAR method. The concept's performance however in 2-dimensional similarity searching has to date not been shown to be effective, and has not been evaluated thoroughly on large sets of compounds.

The work being carried out in this project aims to evaluate hyperstructures as an alternative (if not superior) method for fusion-based similarity-searching, with an emphasis on virtual screening. Current progress on the project will be discussed, including a brief overview of how hyperstructures are constructed, evaluated for virtual screening and compared with existing search methods. Of particular interest will be a comparison with existing data fusion methods. Results in this work show that the hyperstructure concept is not as effective for virtual screening as group fusion using ECFP4 fingerprints in terms of numbers of actives retrieved, but retrieves a greater diversity of molecules. This suggests that the two approaches are complementary, suggesting that it may be beneficial to apply similarity fusion to the two techniques to improve virtual screening.

References:
[1] Brown, N. PhD Thesis. University of Sheffield, 2002.
[2] Palyulin, Radchenko, & Zefirov. (2000). Molecular field Topology analysis method in QSAR studies of organic compounds. *J. Chem. Inf. Comput. Sci.*, **40**, 659–667.

Poster 11

Filling the Bioactivity Spectra of Serine Protease Inhibitors using Proteochemometric Approaches

Qurrat U. Ain, Oscar M. Lucio, *Andreas Bender*

Unilever Centre for Molecular Science Informatics, Department of Chemistry, Lensfield Road CB2 1EW Cambridge United Kingdom

The availability of sequence and structure data of proteins has made it possible to incorporate drug target information in the bioactivity modelling process. The modern chemogenomic approaches are looking forward to use 3D structure and interaction information of a receptor in addition to the ligand properties in predictive modelling. One such chemogenomic approach, Proteochemometric Modelling (PCM), [1] was used in our study to predict the bioactivity of novel compounds against a set of different proteases. These compounds could possibly act as inhibitors for three serine protease target classes and prevent various lethal diseases like myocardial infarction, thrombosis and stroke.
With the available dataset comprising 2,253 compounds, we were able to derive a PCM model (R^2_{test}=0.70, Q^2_{test}= 0.69) for predicting bioactivity against three drug targets (coagulation factor X, thrombin, trypsin) with an overall RMSE of 0.61 log units. Our results imply that, PCM models with binding site residue features worked better than the conventional QSAR model and models that include protein-ligand interaction fingerprints (PLIF) as cross-terms (R^2_{test}: QSAR=0.25, PL (Z-scales3)=0.70, PLxPLIF [2]=0.66; RMSE: 0.91, 0.59, 0.62 log units, respectively). A performance analysis of various alignment dependent and alignment independent target features indicate that using aligned residues responsible for binding interactions yield a better fit of model than the full sequence of protein.
The structural features of compounds and the properties of amino acid residues responsible for bioactivity prediction were also determined by performing a detailed feature analysis on the final model. All the obtained results suggest that PCM models have a powerful ability to exploit binding site sequence information for predicting the bioactivity of novel compounds in an enhanced manner. These models can further be employed to find novel interactions between the compound and its related targets, orphan targets and off-targets.

References:
[1] Van Westen, G. J. P., Wegner, J. K., IJzerman, A. P., van Vlijmen, H. W. T., Bender, A. Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets. *MedChemComm*, (2011), **2**, 16–30.

[2] Chemical Computing Group, MOE Molecular Operating Environment (MOE) **2012**.

Poster 12

Structure-Based Drug Discovery (SBDD) using an NMR ensemble: the case of S100P

Ramatoulie Camara[a], *Sharon Rossiter*[a], Nasir Mahmoud[b], Tatjana Crnogorac-Jurcevic[b], *Stewart B. Kirton*[a]

[a]*Department of Pharmacy, School of Life & Medical Sciences, University of Hertfordshire, Hatfield, Herts, AL10 9AB*
[b]*Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, London*

S100P is a 95-amino acid calcium-binding protein that belongs to the S100 family of proteins and exerts its effect both intracellularly through calcium ion modulation, and extracellularly through binding to the receptor for advanced glycation end-products, RAGE. Numerous studies have linked the protein to many cancers and it is its high expression in pancreatic cancer that has led to proposals for it to be used as a clinical marker in early diagnosis of the disease [1]. Pancreatic cancer is one of the leading causes of cancer deaths in the developed world with a high mortality rate and a less than 5% 5-year survival rate. There is therefore an urgent need to find an effective therapeutic agent for a timely intervention in this lethal disease.

Using the NMR ensemble of S100P (PDB ID 1OZO) and cromolyn – a ligand that has been shown to bind to and inhibit the protein's interaction with RAGE [2] – protein informatics was carried out to identify appropriate S100P conformers for drug discovery, and subsequently, small molecules from the MOE and ZINC databases of lead-like compounds with the potential to bind to and therefore stop S100P from interacting with its receptor.

Virtual screening of the MOE database of lead-like compounds resulted in a hit rate of 0.008%. Of the 52 hits identified, 15 were purchased and five synthesised in the lab. All were tested in an MTS cell migration assay and five compounds from different chemical classes showed inhibitory activity which was at least equivalent to that of cromolyn-S100P interaction. This corresponds to a hit rate of 25% which is a promising start in search of a potential therapeutic agent for pancreatic cancer. Hits from the ZINC database are presently being processed and the results will be presented at this meeting. Compounds identified so far as hits are currently being explored in hit-to-lead studies.

References:
[1] Deng, H., *et al*., Usefulness of S100P in diagnosis of adenocarcinoma of pancreas on fine-needle aspiration biopsy specimens, *Am. J. Clin. Pathol*., (2008), **129**(1), 81-88.
[2] Arumugam, T., Ramachandran V. and Logsdon C.D., Effect of cromolyn on S100P interactions with RAGE and pancreatic cancer growth and invasion in mouse models, *J. Natl. Cancer Inst*., (2006), **98**(24), 1806-1818.