



MGMS and RSC MMG Young Modellers' Forum 2007

ORAL PRESENTATIONS

Talk 1

Molecular Dynamics Study of the Structural Stability and Materials Properties of DNA-Intercalated Layered Double Hydroxides

Mary-Ann Thyveetil, Chris H. Greenwell, James L. Suter, and *Peter V. Coveney*

*Centre for Computational Science, Department of Chemistry, University College London,
20 Gordon Street, London, WC1H 0AJ*

The intercalation of DNA into layered double hydroxides (LDHs) has various applications, including drug delivery for gene therapy and origins of life studies. The nanoscale dimensions of the interlayer region make the exact conformation of the intercalated DNA difficult to elucidate experimentally. We use molecular dynamics techniques, performed on high performance supercomputing grids, to carry out large-scale simulations of double stranded, linear and plasmid DNA up to 480 base pairs in length intercalated within a magnesium-aluminium LDH. Currently only limited experimental data have been reported for these systems. Our models are found to be in agreement with these experimental observations, according to which hydration is a crucial factor in determining the structural stability of DNA. Phosphate backbone groups are found to align with aluminium lattice positions. At elevated temperatures and pressures, relevant to origins of life studies which maintain that the earliest life forms originated around deep ocean hydrothermal vents, the structural stability of LDH-intercalated DNA is substantially enhanced as compared to DNA in bulk water. We have previously determined the Young's moduli of montmorillonite¹ and LDH intercalated with chloride ions² by carrying out uniaxial deformations, the simulations presented in this study use these same techniques to determine how the materials properties of the LDH are modified due to DNA intercalation.

References:

1. J. L. Suter, P. V. Coveney, H. C. Greenwell, and M.-A. Thyveetil. *J. Phys. Chem. C*, 111:8248–8259, 2007.
2. M.-A. Thyveetil, P. V. Coveney, J. L. Suter, and H. C. Greenwell. *Chem. Mater., Emergence of Undulations and Determination of Materials Properties in Large-Scale Molecular Dynamics Simulation of Layered Double Hydroxides*, in press, 2007.

Talk 2

Binding Free Energy Calculation from Molecular Dynamics Simulation

Floris Paul Buelens, *Gabriel Waksman*

School of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX

The free energy difference between receptor-bound and -unbound states of a ligand is a quantity of critical interest to pharmaceutical science, as a key determinant of drug efficacy. Many empirical approaches to free energy / affinity scoring have been developed and are widely applied, but these methods by necessity ignore many key determinants of ligand binding and fall far short of quantitative accuracy. In contrast, force field-based methods can be used to sample the free energy profile of a ligand-protein association from basic principles of physical chemistry and thermodynamics. These methods promise both quantitative accuracy and insight into the dynamic properties of a ligand-protein pairing beyond the reach of empirical methods, at the cost of much increased computational expense. Inherent weaknesses of most empirical scoring potentials including protein flexibility, solvation energy and association and conformational entropy are implicitly accounted for by force field-based methods.

Despite the potential of force field-based free energy calculation, which has been recognised for over two decades, there are few reports of practical implementations. This can largely be attributed to the amount of computing power required to obtain consistent results, which has not been practically available until recently, but in addition unanswered questions remain with regards to the detail of implementing the relevant principles of statistical thermodynamics [1].

The work presented here is the implementation of a ‘theoretically exact’ framework for ligand-protein binding free energy calculations, based on a thermodynamic integration protocol sampled from molecular dynamics (MD) simulation, using the NAMD molecular simulation package. Practical aspects of the implementation, including computational details and modifications applied to standard MD force field potentials, will be presented. I will consider the strengths and weaknesses of the protocol with reference to the above mentioned failings of empirical scoring potentials. Finally,

experiences with calculating affinities for a series of inhibitors of kinase PDK-1 will be discussed.

References:

[1] Gilson, MK, Given, JA, Bush, BL and McCammon, JA. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: a Critical Review. *Biophysical Journal* (1997) **72**, 1047-69.

Talk 3

Molecular Spam: Use of a Modified Spam Filter for Classification of Bioactive Molecules and Drug Target Prediction

Florian Nigsch, *John B.O. Mitchell*

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

With 90% of all global email traffic being spam, it is a considerable challenge to find those messages that are legitimate and contain useful information. Computational methods in support of drug discovery are faced with the same challenges: Which out of tens of thousands of possible molecules will show a similar biological activity to a handful of known and promising examples?

We have converted a highly accurate spam filter to discriminate between biologically active and inactive molecules from large *in silico* virtual libraries. The adaptations we have introduced allow us to use the method for the prediction of the likely protein targets of small druglike organic molecules, and for the identification of the molecular fragments giving rise to each observed bioactivity.

Our results reveal an error rate of below 4% on a dataset of 8,500 molecules in 11 classes, including GPCR-binding ligands, HIV protease inhibitors and protein kinase C inhibitors. Addition of 95,000 inactive molecules as decoys (“molecular spam”) results in a decrease in accuracy to just below 90%. We have followed this up with an even more challenging experiment, in which we show that our method can predict the correct protein target, from almost 200 possibilities, for 70% of 43,000 molecules.

The nature of the underlying algorithm [1] permits straightforward identification of important molecular fragments contributing to the bioactivity of each class. We show how this can be used to colour-code molecules in a chemically intuitive way, displaying the propensity of each fragment to confer a given bioactivity.

Reference:

[1] Littlestone, N. Learning Quickly When Irrelevant Attributes Abound: A New Linear-threshold Algorithm. *Machine Learning* **1988**, 2, 285–318.

Talk 4

Homology Model Based Virtual Screening for GPCR Ligands Using Docking and Target-Biased Scoring

Sebastian Radestock, *Steffen Renner, Tanja Weil*

*Merz Pharmaceuticals GmbH, Merz Pharmaceuticals GmbH, Chemical R&D,
Altenhöferallee 3, 60438 Frankfurt am Main, Germany*

G-protein coupled receptors (GPCRs) are one of the most important drug targets for the pharmaceutical industry. For instance, metabotropic glutamate receptors (mGluRs) have attracted interest due to their role as modulators of major neurotransmitter systems in the central nervous system. Detailed structural information about GPCRs is lacking. As a consequence, computational design of modulators for GPCRs can only be accomplished by using structure-based approaches grounded on homology models, or by using ligand-based virtual screening methods.

In the present study, we investigated the combination of two recently reported techniques for the improvement of homology model-based virtual screening. First, we applied ligand supported homology modelling.^[1] Clues to infer the binding modes of the ligands were provided by data from mutagenesis studies. Second, to rank order docking solutions, we developed a scoring scheme that exploits the patterns of interactions between ligands already known to bind to the target, and the binding site. As reference ligands, the compounds that have already been employed to support homology modeling were used. Patterns of interactions were modelled using binary ligand receptor fingerprints,^[2] as pioneered by Singh *et al.*^[3] The similarity of two fingerprints was evaluated using the Tanimoto coefficient.

Our methodology, subsequently referred to as interaction fingerprint-based similarity (IFS), has been tested in retrospective virtual screening experiments against mGluR subtype 5. It is expected that the identification of negative allosteric modulators of mGluR5 will open up new therapeutic possibilities to treat pain, anxiety, or Parkinson's disease.^[4] To put the results into proper perspective, docking solutions were also rank ordered using conventional scoring functions (D-Score, PMF-Score, G-Score, Chemscore, and FlexX-Score). Using IFS, the enrichment rates could significantly be improved. We also show that the power of IFS to discriminate between active and inactive compounds is superior to the discriminatory power of the conventional scoring

functions. Our results indicate that the presented approach might serve as a general setup for successful GPCR virtual screening.

References:

- [1] A. Evers, H. Gohlke, G. Klebe, *J. Mol. Biol.* (2003) **334**, 327-345.
- [2] S. Renner, S. Derksen, S. Radestock, *submitted*.
- [3] Z. Deng, C. Chuaqui, J. Singh, *J. Med. Chem.* (2004) **47**, 337-344.
- [4] C. J. Swanson, M. Bures *et al.*, *Nat. Rev. Drug Disc.* (2005) **4**, 131-134.

Talk 5

Study of the Lysozyme Protein: Use of a Novel Method to Sample Conformational Change

J. Fernandez Carmona, *J. W. Essex*

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

Lysozyme is an opsonin protein, known to be bound to several ligand, each of which involves a different conformation of the F loop. Several Molecular Dynamic simulations have been performed to compute the relative free energies of this ligand with good agreement with experiment [1]. However these simulations could not reproduce the change in the F loop conformation.

Difficulties in reproducing the change in conformation lie in geometric problems such as the existence of different rotamers of the side chain, hydrophobic clustering, or the small difference in the Ψ/Φ angles of the residues within the loop.

Using Monte Carlo methods to investigate loop motion in proteins requires specific algorithms, such as the Concerted Algorithm with Angle [2] (CRA). This specific move allows local concerted motion of several degrees of freedom over the protein, and hence enhances the sampling of phase space. In this presentation, the application of the CRA algorithm to the F loop will be described. The use of CRA led to insight into the binding mechanism for two specific ligands, indole and isobutylbenzene. Several sets of simulations using the CRA move on the F loop have been run at various temperatures to capture the conformational change of the F loop, as well as to compute the relative binding free energy between the two ligands.

References:

- [1] Y. Deng, B. Roux, *J. Chem. Theory. Comput.*, 2,5, **2006**, 1255-1273.
- [2] J. P. Ulmschneider, W. L. Jorgensen, Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias, *J. Chem. Phys.*, **2003**, 118, 4261-4271.

Talk 6

A Core Fragment Based Look at Chemical Space

Agata Krasowski, Paul Wyatt, *Ruth Brenk*

University of Dundee, Biological Chemistry and Drug Discovery, Dow Street, Dundee, DD1 5EH

High Throughput Screening (HTS) and Virtual Screening (VS) play important roles in modern drug discovery. These methods are used to screen extremely large numbers of compounds against biological targets. New hits selected from HTS and VS can be used as leads for further optimisation and drug development. Since screening collections can only cover a small fraction of chemical space they should be as diverse as possible. To explore the diversity of compounds within such a collection, a core-fragment based approach was developed. This method was tested by analysing different publicly available databases and subsets of compounds found within them. The two questions addressed were: ‘How many core fragments are excluded when concentrating on more constrained sub-sets?’ and ‘What is the nature of these excluded fragments?’

To answer these questions we analysed the ZINC database (<http://zinc.docking.org>) fragment-like and lead-like sub sets, in terms of the diversity and overlap of the contained structures. In spite of the lead-like sub set being 15 fold larger than the fragment-like sub set, it was observed that one quarter of the lead-like core fragments were also found in the considerably smaller fragment-like sub set. Additionally, most of the lead-like core fragments excluded from the fragment-like set did not satisfy the size and chemical property criteria commonly used in screening libraries. This result suggests that increasing the size of screening sets does not guarantee improved overall output, while conversely smaller, well-designed screening sets may in some cases be advantageous.

Talk 7

In Silico Identification and Characterisation of Ion Binding Sites in Glutamate Receptors

Ranjit Vijayan and *Philip C. Biggin*

Structural Bioinformatics and Computational Biochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU

Ionotropic glutamate receptors (iGluR), activated by the amino acid L-glutamate, form a large family of ligand gated ion channels that mediate the majority of excitatory neurotransmission in the brain. Pharmacologically, the iGluR family is broadly categorized as α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA), N-methyl-D-aspartate (NMDA) and kainate based on selectivity. Ions, including Ca^{2+} , Mg^{2+}

and Zn^{2+} , play a modulatory role in many receptors. Recently, it was shown that kainate receptors, but not AMPA and NMDA receptors, require both Na^+ and Cl^- to function [1]. Whilst only the anion binding site was identified in the above study, it was shown that both anions and cations bind in the ligand binding domain. Using computational approaches, we have identified two cation binding sites located symmetrically opposite the anion binding site and within the same dimer interface cavity. Multiple molecular dynamics simulations and relative binding free energy calculations using thermodynamic integration have been performed to study the anion and cation binding sites in detail. Simulations confirm that the identified locations are indeed cation binding sites. The rank order of binding for halides and alkali metals were determined which is indicative of their binding affinities. The computational results agree very well with mutagenesis [2] and crystallographic studies [3].

References:

- [1] Plested AJ and Mayer ML (2007) Structure and mechanism of kainate receptor modulation by anions. *Neuron* 53:829-841.
- [2] Paternain AV, Cohen A, Stern-Bach Y and Lerma J (2003) A role for extracellular Na^+ in the channel gating of native and recombinant kainate receptors. *J Neurosci.* 23:8641-8648.
- [3] Weston MC, Schuck P, Ghosal A, Rosenmund C and Mayer ML (2006) Conformational restriction blocks glutamate receptor desensitization. *Nat. Struct. Mol. Biol.* 13:1120-1127.

Talk 8

The UV Spectrum of $[\text{Zn}(\text{pyridine})_4]^{2+}$

Caroline Norris and *Hazel Cox*

Chemistry Division, Department of Chemistry and Biochemistry, University of Sussex, Falmer, Brighton, BN1 9QJ

Many of the important metal-ion complexes in condensed phase chemistry and biochemistry are multiply charged. Therefore, the study of similar metal-containing structures in the gas-phase offers a unique opportunity to probe the fundamental metal-ligand interactions that underpin the stability and structure of these complexes.

A recently recorded UV spectrum on cold (~ 100 K) $[\text{Zn}(\text{pyridine})_4]^{2+}$ molecules appears to contain vibrational structure [1]. In this exciting experimental development, theory is critical in understanding the new observations. Time-dependent density functional theory (TDDFT) is able to account for the observed electronic excitations, and by coupling the dominant transition to the allowed vibrational modes quantitative agreement with the spacings between the most prominent peaks is established.

The dominant electronic transitions are ligand based $\pi^* \leftarrow \sigma$ and $\pi^* \leftarrow \pi$, and a comparison with the electronic transitions of the isolated pyridine molecule shows that the presence of the Zn dication red-shifts these excitations. TDDFT calculations on $[\text{Zn}(\text{pyridine})]^{2+}$ were also used to assess the effect of solvation in this important class of compounds.

Reference:

[1] A. J. Stace *et al.*, University of Nottingham, *Private Communication*

Talk 9

An Application of Reaction Vectors in *De Novo* Design

Hina Patel¹, Val Gillet¹, Beining Chen² and Michael Bodkin³

¹*Department of Information Studies, University of Sheffield, Western Bank, Sheffield, S10 2TN*

²*Department of Chemistry, University of Sheffield, Western Bank, Sheffield, S10 2TN*

³*Eli Lilly UK, Erl Wood Manor, Windlesham, Surrey, GU20 6PH*

A number of *de novo* design tools have been described with the aim of generating novel molecules for drug design, however, they are limited in their ability to propose molecules which are synthetically feasible. Here we investigate the use of reaction vectors for the design of novel molecules that are of sufficient complexity to be non-obvious and which are also synthetically accessible.

Broughton *et al.* [1] have recently described the reaction vector which captures the changes that take place at the reaction centre, without the need for complex reaction mapping procedures. The individual components of a reaction are described by vectors (such as atom pairs) and the overall reaction vector is generated using:

$$\text{Reaction Vector} = [\text{Sum of product vectors}] - [\text{Sum of reactant vectors}]$$

Here we show how the reaction vector can be applied in the forward direction to suggest novel molecules for synthesis. Its use in both simple transformations involving, for example, a simple functional group substitution and to more complex multi-component reactions of the form $(\text{R1} + \text{R2} \rightarrow \text{P1} + \text{P2})$ is also demonstrated. Reactions were extracted from the Lilly database, a cleaning algorithm applied to correct incomplete reactions (e.g. missing reactants, missing products) and a database of reaction vectors was created. A database of reactants was obtained from a database of commercially available compounds. By mixing and matching reaction transforms and reactants it was possible to

generate novel compounds. We demonstrate the application of the algorithm to the design of known drugs from simple starting materials.

We are currently incorporating the method into a fully automated multi-objective application for design of novel molecules.

Reference:

[1] Broughton, H. B. et al. Methods for Classifying and Searching Chemical Reactions. *United States Patent Application 367550* (2003). 25 Sept.

Talk 10

Searching New Inhibitors of *Trypanosoma cruzi* Triosephosphate Isomerase (TcTIM). Docking and Virtual Screening Studies.

Liliana Moreno-Vargas (1), Jorge Estrada (2) (3), Javier Sancho (2) (3), Arturo Rojo-Domínguez (1)

(1) *Departamentos de Química y de Ciencias Naturales, Universidad Autónoma Metropolitana-Iztapalapa y Cuajimalpa, San Rafael Atlixco 86, México D. F., 09340 (Mexico)*

(2) *Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza, Pedro Cerbuna 12, Zaragoza, 50009 (Spain)*

(3) *Instituto de Biocomputación y Física de Sistemas Complejos, Universidad de Zaragoza, Corona de Aragón 42, Zaragoza, 50009 (Spain)*

Chagas disease, occurring in America, is caused by the parasite *Trypanosoma cruzi*, and it may do fatal damage to the heart and digestive tract. One research line for the discovery of new drugs against this parasite focuses on specifically inactivating its triosephosphate isomerase homodimer enzyme, possibly an essential one in its glycolysis pathway. Though there exists a crystal structure of TcTIM in complex with an in vitro inhibitor agent, it was also found that two molecules of the agent are needed per enzyme for inactivation. The present work located potential sites of binding for the agent using the docking software Autodock 3.0.5 in a blind docking experiment. From the first three best scored sites, two were selected, and the active site excluded, since the inhibition is non-competitive. Later on, virtual screenings of a commercial diverse chemical library of 10,000 compounds were conducted on those two sites, using the docking software FRED v. 2.2.3. Virtual screening setup consisted of preparing nine test sets using published literature lists of actives for six different protein targets, and choosing the best overall scoring functions, based on the area under the curve of the receiver operating characteristic plot for each test set and scoring function. Inactives for each test set were chosen from the 10,000-compound library. The selection method tried to create an

inactive test set with properties very similar to those of the active set. The properties compared between the active and inactive sets were such as the mean and variance of the molecular weight of all molecules, and the mean and variance of the number of hydrogen bond donors in all molecules. Considering the results of the nine tests, two of the scoring functions were chosen for the production virtual screening. We will also show the analysis of the best scored molecules.

Talk 11

Seeing the Trees Through the Forest

Craig L. Bruce,¹ Stephen D. Pickett,² *Jonathan D. Hirst*¹

¹ *School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

² *GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK*

Several factors contribute to a useful QSAR model. Often a good model hinges on one aspect - predictive accuracy. However, this alone is only part of the model developed. A more important question is what chemical insight can the model offer? This interpretation is dependent on the classifier. For example, a decision tree is simple to interpret, but does not produce the most predictive models. Similarly, support vector machines offer excellent predictive capability but an uninterpretable model. They are an example of Occam's razor, where a balance of simplicity and accuracy must be found with the chosen classifier.

Previously we have shown random forests obtain comparable predictive accuracies to support vector machines.¹ A decision tree is easier to interpret than a random forest; Breiman gave them an 'A' and 'F' for interpretability respectively.² It is the different tree construction and number of trees present in a forest that makes their interpretation exceedingly complicated. One cannot simply glance through the forest and readily see the model, as one can with a decision tree.

Therefore, we present new tools to allow an increased understanding of the random forest classifier. This entails alternative tree representations using SMILES and SMARTS. Increased interoperability means we can use pre-existing SMILES based applications, e.g. to perform similarity comparisons. Descriptor analysis at the tree and forest level highlights the frequency and distribution of descriptors. Pathway analysis will determine popular and recurring sub graphs present in the trees. With this framework now in place around the Weka machine-learning package we can continue to extend this functionality.

References:

1. Bruce, C. L.; Melville, J. L.; Pickett, S. D.; Hirst, J. D. Contemporary QSAR Classifiers Compared. *J. Chem. Inf. Mod.* **2007**, *47*, 219-227.
2. Breiman, L. Statistical Modeling: The Two Cultures. *Statist. Sci.* **2001**, *16*, 199-231.

Talk 12

Cytochrome P450 reactivity and specificity from QM/MM mechanistic modelling

Richard Lonsdale, *Jeremy N. Harvey, Adrian J. Mulholland*

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

Cytochrome P450 is an important metalloenzyme found in plants, animals and bacteria. It activates molecular oxygen and catalyzes stereoselective and regioselective oxygen insertion reactions of a wide variety of organic compounds. These enzymes are of particular interest in drug metabolism, because they play important roles in drug disposition, and in their pharmacological and toxicological effects. A key aim is the prediction of selectivity of reactions of drugs with different P450 isozymes. Combined quantum mechanics/molecular mechanics (QM/MM) methods now allow reactions in P450 enzymes to be modelled, and provide a promising approach to analysing determinants of selectivity in drug metabolism. Comparison with experimental selectivity and reactivity data helps to validate these methods. The oxidation of cyclohexene by the bacterial P450cam from *Pseudomonas putida* is an important example. It produces C=C bond epoxidation and allylic hydroxylation products in approximately equal amounts. In contrast, oxidation of propene yields exclusively epoxidation-type products. QM/MM calculations of barriers for these reactions, combined with molecular dynamics simulations, give relative barriers for these processes that are consistent with experimental findings.

References:

Zurek, J., Foloppe, N., Harvey, J.N., Mulholland, A.J., Mechanisms of reaction in cytochrome P450: hydroxylation of camphor in P450cam, *Org. Biomol. Chem.* (2006) **4**, 3931-3937.

Cohen, S., Kozuch, S., Hazan, C., Shaik, S., Does Substrate Oxidation Determine the Regioselectivity of Cyclohexene and Propene Oxidation by Cytochrome P450? *J. Am. Chem. Soc.*, (2006) **128**, 11028-11029.
