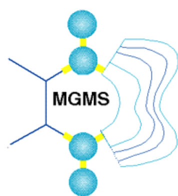

MGMS and RSC MMG

Young Modellers' Forum 2010

POSTER PRESENTATIONS

Poster 1

Theory vs. experiment in probing a prototypical protein-protein interaction

Richard T. Bradshaw, Bhavesh H. Patel, *Ian R. Gould, Edward W. Tate, Robin J. Leatherbarrow*

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

The central role of protein-protein interactions in a wide range of cellular processes makes them a target for research and drug discovery. The computational MM-PB/GBSA approaches are widely used to predict and rank relative binding free energies (ΔΔG) for complexes and their mutants, however the conformational flexibility of the interface means it is often difficult to obtain converged free energy estimates using snapshots from independent molecular dynamics simulations.

In this study, we first experimentally assayed binding affinities for an example protein-protein complex between trypsin and a nine-residue synthetic peptide. We then compared how accurately, precisely and reliably two different MM-PB/GBSA methods replicated these experimental results.[1] We find that the "post-process alanine scanning" protocol[2] of a single native complex trajectory gives results with better accuracy than running separate molecular dynamics trajectories for individual mutants. Importantly, compared across ten independent simulations we find that results from post-process alanine scanning are also more precise, and are obtained over five times faster than their equivalent with the "full MD" protocol. Our results suggest that, whilst not suitable in every case, post-process alanine scanning can be a reliable tool in predicting residues at protein interfaces with potential for modulation, and may be used in preference to more traditional MD approaches.

References:

1. R.T. Bradshaw et al. *Prot. Eng. Des. Sel.,* 2010, in press, DOI: 10.1093/protein/gzq1047
2. I. Massova and P.A. Kollman, *J. Am. Chem. Soc.*, 1999, **121**, 8133-8143

Poster 2

Large scale conformational changes in aromatic transporters in presence of benzene

Pragya Chohan & Mark S.P. Sansom

Department of Biochemistry, University of Oxford.

Proteins such as TodX and TbuX mediate transport of hydrocarbons across the outer membranes in many Gram negative bacteria. This is of interest in understanding mechanisms of bacterial biodegradation of hydrocarbons. Transport of hydrocarbons across the outer membrane is hindered by the anionic lipopolysaccharide chains which form the outer leaflet of the bilayer. A number of studies have indicated that the TodX may capture and transport hydrophobic ligands such as benzene and toluene. Crystallographic structure of TodX reveals a β-barrel architecture and suggest that the L3 external loop may be associated with regulation of ligand access and binding. An N-terminal domain (the 'hatch' domain) folds into the barrel and blocks the pore, a feature seen in some related outer membrane proteins. Upon ligand binding, conformational changes in N-terminal domain of TodX have been proposed to open the pore and allow transport of aromatic compounds to the periplasm. A similar mechanism has been proposed for the related protein FadL.

We have performed MD simulations of TodX embedded in a lipid bilayer, focussing on loop conformations and β-barrel stability. In atomistic simulations (100 ns duration) of apo-TodX embedded in POPC bilayer, we observed large conformational changes in the L3 loop. The inner hatch domain remains within the barrel and restricts the motion of water through the TodX pore. Simulations of mutants with the hatch domain deleted show that water is able to fully permeate the pore but the stability of the β-barrel is decreased. Loop flexibility may open a lateral pathway for entry of hydrocarbons into the interior of the pore.

In the crystal structure, Loop L3 is in a 'closed' conformation, restricting access to the central pore. We have generated models of TodX with L3 in an 'open' conformation and have identified possible sites of ligand (benzene) binding. Further simulations based on these models may help to reveal the pathway(s) of hydrocarbons and through into the central pore.

References:
Hearn, EM, van den Berg B, Patel DR. Outer-membrane transport of aromatic hydrocarbons as a first step in biodegradation, *Proc Natl Acad Sci USA*. (2008) **105**,860

Poster 3

A Computational Study into the Role of Defects in $BaTiO_3$

<u>J. A. Dawson</u>, B. L. Bin, C. L. Freeman, *J. H. Harding and D. C. Sinclair*

Department of Engineering Materials, University of Sheffield, Sheffield, United Kingdom

It is well known that the $BaTiO_3$ perovskite structure can accommodate a wide range of both isovalent and aliovalent dopant ions and that these dopant ions have a significant effect on the electrical properties of the material. For instance, a small concentration of La on the Ba-site has a dramatic effect on resistivity and causes the material to exhibit *n*-type semiconductivity, however when the La concentration is increased the material becomes insulating [1]. XRD has been used to establish ternary phase diagrams for Rare Earth (RE^{3+}) ion doping of $BaTiO_3$ and to gain insightful information on solid solution limits and substitution sites. Large rare earth ions, such as La have been found to exclusively dope at the Ba-site with a large solid solution limit (up to 25%). Intermediate sized cations such as Gd, show very limited Ba or Ti-site solubility; yet show extensive solid solutions (up to 20%) according to a self-compensation mechanism, i.e. RE^{3+} substitution at both the Ba^{2+} and Ti^{4+} sites. Small dopants such as Yb, doped exclusively on the Ti-site with a solid solution limit of up to 5% for the cubic phase and between 5% and 15% for the hexagonal phase.

Atomistic methods have been used to study the defect chemistry of $BaTiO_3$ at both the dilute limit and beyond. New Ba-O and Ti-O interatomic potentials have been created and calibrated against ab initio simulations. Both cubic and hexagonal phases have been considered. Numerous dopant incorporation methods have been studied for RE^{3+} incorporation and excellent correlation with experimental findings has been observed. We also discuss the implications for the method of semiconductivity in the material based upon the energetics of the dopant incorporation mechanisms.

References:

1. Jonker G. H., *Some aspects of semiconducting barium titanate*, Solid-State Electron., 1964, **7** (12): p. 895-903.

Poster 4

Novel half-bonded model for metalloprotein simulation

Mahmoud A. A. Ibrahim[a,b] and Neil A. Burton[a]

[a]*School of Chemistry, The University of Manchester, Manchester M13 9PL, UK*
[b]*Chemistry Department, Faculty of Science, Minia University, Minia 61519, Egypt*

The use of classical molecular mechanical potentials to study binding to metal ions is a challenge in drug discovery. Two general models for metal ion simulation are known: Bonded and non-bonded models. Each of these models has its own drawbacks. Non-bonded models are known to be very sensitive to the choice of the parameters and suffer from an inability to retain a low coordination number. Bonded models lack the ability for the metal ion to change its coordination number; in addition, a huge effort is needed to generate the missing force field bonded terms for the studied ligands. The aim of the current research is to develop a new model for metalloprotein simulation overcoming the drawbacks of each of the conventional bonded and non-bonded models. A novel approach is proposed in the current study: A half-bonded model, in which the interactions between the metal ion (zinc ion) and the active site's residues have been treated using bonded terms. On the other hand, the interaction between the metal ion and an inhibitor has been treated by specific parameterization of the electrostatic interactions. The new model was applied to study the binding affinities of a set of ligands against Matrix Metalloprotein-3 (MMP3). The results were compared with the data generated with conventional non-bonded models and experimental values. The new model predicts the correct binding affinity ranking of the studied inhibitors.

Poster 5

Quantitative mapping of ligand and target binding spaces of chemical probes

Felix Krüger, *John Overington*

European Bioinformatics Institute, Wellcome Trust Genome Campus, Hixton, Cambridgeshire, CB10 1SD

Chemical probes are an essential component of the toolbox available to chemical biologists. These small molecules are used to perturb biological systems in a defined and reproducible fashion [1]. Increasing availability of pharmacological and high throughput screening data make it possible to globally map the chemical space of bioactive compounds. This study aims to describe quantitatively the ligand and target binding spaces of chemical probes. Large scale analysis of binding affinity data from the ChEMBL data base and literature mining have been combined to identify compounds that are of interest to the chemical biology community and a quantitative analysis of these compounds has been carried out. The overlap and differences between these compounds and commercially available compound libraries were identified and explored in terms of chemical structure, occurrence in the literature and bioactivity profiles. The robustness of
compound binding in different species was examined by comparing affinity data for pairs of homologs and a general correlation between differences in binding affinity and sequence divergence was observed.

References:

[1] Brent R Stockwell, Exploring biology with small organic molecules, *Nature* (2004), **432**, 846-854.

Poster 6

Towards a Multiscale Simulation of Cellular Calcium Signalling in the Cardiac Cell

Daniel Mason, Jonathan Essex, Hans Fangohr

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

Atomistic simulation of a large-scale cellular process is currently a computationally intractable task. To realistically see a simulation through to completion in a reasonable time, the optimum method must be chosen. A compromise between simulating detailed atomistic processes and larger scale cellular processes is achieved through multiscale modelling.

Calcium is an important ion in biology, usually maintained at specific concentrations in different parts of the cell to maintain proper homeostasis. Changes in concentration lead to a variety of different cellular responses. Normal heart function relies on precise cooperation from different parts of the cardiac cell. This could involve regional rises in concentration to induce release of further calcium from a cellular store or to induce apoptosis (programmed cell death). Malfunction in any part of the cellular machinery could have an adverse effect on the ability of the cell to contract, resulting in a disease state.

The model implemented in this project contains simple equations describing flux, pump rates, membrane voltage and compartmental ion concentrations over time as a network of ordinary differential equations (ODEs). The inter-dependence of each equation combined with Markov representation of ion channel kinetics adds a great deal of complexity to the system. Although complex, this is still a simplification of actual cellular function designed to reproduce experimental data.

The framework of this network model will allow the use of information provided by more detailed simulations to investigate the effect a molecular level event can have on a cellular level.

Poster 7

Dynamics based alignment of proteins

Marton Munz, Rune Lyngsø, Jotun Hein, Philip C Biggin

Structural Bioinformatics and Computational Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK

Background: The dynamic motions of many proteins are central to their function. It therefore follows that the dynamic requirements of a protein are evolutionary constrained. In order to assess and quantify this, one needs to compare the dynamic motions of different proteins. Comparing the dynamics of distinct proteins may also provide insight into how protein motions are modified by variations in sequence and, consequently, by structure. The optimal way of comparing complex molecular motions is, however, far from trivial. The majority of comparative molecular dynamics studies performed to date relied upon prior sequence or structural alignment to define which residues were equivalent in 3-dimensional space.

Results: Here we discuss an alternative methodology for comparative molecular dynamics that does not require any prior alignment information. We show it is possible to align proteins based solely on their dynamics and that we can use these dynamics-based alignments to quantify the dynamic similarity of proteins. Our method was tested on 10 representative members of the PDZ domain family.

Conclusions: As a result of creating pair-wise dynamics-based alignments of PDZ domains, we have found evolutionarily conserved patterns in their backbone dynamics. The dynamic similarity of PDZ domains is highly correlated with their structural similarity as calculated with Dali. However, significant differences in their dynamics can be detected indicating that sequence has a more refined role to play in protein dynamics than just dictating the overall fold. We suggest that the method should be generally applicable.

References:

Dynamics based alignment of proteins: an alternative approach to quantify dynamic similarity (Munz M, Lyngso R, Hein J, Biggin PC) BMC Bioinformatics, 2010 Apr 14;11:188.
PDZ domains: folding and binding (Jemth P, Gianni S.)Biochemistry, 2007 Jul 31;46(30):8701-8.

Poster 8

Electrostatic embedding in large scale DFT biomolecular simulations

Chris Pittock[1], Stephen Fox[1], Thomas Fox[2], Christofer Tautermann[2], Noj Malcolm[3], *Chris-Kriton Skylaris*[1]

[1] *School of Chemistry, University of Southampton, SO17 1BJ, UK*
[2] *Boeringer Ingelheim Pharma GmbH & Co. KG, Department for Lead Discovery, 88397 Biberach, Germany*
[3] *Accelrys, 334 Cambridge Science Park, CB4 0WN, UK*

Simulating the activity of biomolecules, such as proteins, is usually performed using molecular mechanical (MM) methods, since their computational cost is much lower than quantum mechanical (QM) approaches, such as Density Functional Theory (DFT). However, there are limitations, such as a lack of polarisation or the inability to describe electron-transfer, and this can reduce the accuracy. A method of eliminating these issues is to partition the biomolecular system to QM and MM regions, which are coupled via some interface. Numerous studies have been made using such QM/MM approaches[1].

To further improve this description and provide a system that is optimal towards accuracy, we ideally would like to treat the entire system at the QM level. With the ONETEP[2] code for linear-scaling DFT, the ability to perform DFT calculations on entire biomolecules, such as proteins aids us towards this goal. We will show preliminary tests of such approaches, which include calculation of interaction energies, where the entire protein and specific parts of the solvent area are described by QM. The system is further embedded electrostatically in MM water.
Future work that will result from this study will be using these supplemented large-scale QM calculations in rigorous statistical mechanics approaches, designed for calculating free energies of binding of ligands in proteins.

References:
[1] Lin, H., Truhlar, D.G., *Theor. Chem. Acc.* (2007) **117**, 185-199.
[2] Skylaris, C-K., Haynes, P.D., Mostofi, A.A., Payne, M.C., *J. Chem. Phys.* (2005) **122**, 084119.

Poster 9

Protein Unfolding Under Force: Crack Propagation in a Network

Gareth Shannon[†], Adam M. R. de Graff[‡], Dan Farrell[‡], M. F. Thorpe[‡], Philip M.Williams[†]

[†]Laboratory of Biophysics and Surface Analysis, School of Pharmacy, University of Nottingham NG7 2RD, UK; [‡]Department of Physics and Center for Biological Physics, Arizona State University,Tempe, AZ

The manner in which proteins respond under an applied force is of direct biological significance, as the mechanical, regulatory and signalling properties of many proteins depend not only on their native state conformations, but also on the nature of intermediate states that become populated when subjected to an applied load. Protein unfolding under force has been studied experimentally using atomic force microscopy and optical tweezers. Many theoretical and computational studies have also been performed; most notably molecular dynamics simulations, as well as other models such as coarse-grained Gö-like models and elastic network models. Here, we have developed a new approach which builds on a recently developed method called Geometric Targeting (1). Proteins are represented as all-atom polypeptides cross-linked by hydrogen bonds, salt bridges and hydrophobic constraints; each modelled as harmonic inequality constraints capable of supporting a finite load before breaking. Additional constraints enforce correct stereochemistry by preventing steric overlap and ensuring favourable Ramachandran and torsion angles. Unfolding pathways were generated for a set of 12 proteins with diverse topologies by minimally overloading the network in an iterative fashion whilst removing bonds that are in violation of their maximum strain. The results were compared with those of molecular dynamics simulations and also experimental results where possible. This simple and intuitive model of protein unfolding was shown to be sufficient to describe the dominant unfolding pathways from the molecular dynamics simulations for 9 of the 12 proteins in this study, capturing both subtle intermediates and intermediates far from the native state ensemble. Of the 4 proteins for which conclusive experimental evidence is available, the technique repeatedly reproduced the intermediates of 2 of them. The success of the technique is surprising as the unfolding model is based purely on strain and does not sample the free energy of the states along the pathway. One might expect strain-based pathways to rapidly deviate from those from MD simulation for which the protein is allowed to diffuse in a detailed energy landscape. Instead the sequential strain-based breaking events can be followed far from the native state and reproduce the order of loss of many units of secondary structure. This provides further evidence that the high forces used in typical MD simulations tilt the energy landscape to such an extent that the unfolding is not the thermally driven process that occurs under experimentally and physiologically relevant forces.

References:

1. Farrell, D. W., K. Speranskiy, and M. F. Thorpe. 2010. Generating stereochemically acceptable protein pathways. Proteins: Structure, Function, and Bioinformatics 78:2908-2921.

Poster 10

Mining of Emerging Structural Patterns for Identification of Toxicophores

Richard Sherhod[1], *Val Gillet*[1], *Philip Judson*[2]

[1]*Department of Information Studies, University of Sheffield, Regent Court, Portobello Street, Sheffield, S1 4DP, UK*
[2]*Lhasa Limited, 22-23 Blenheim Terrace, Woodhouse Lane, Leeds, LS2 9HD, UK*

The design of new *alerts*; collections of structural features observed to result in toxicological activity, can be a slow process and may require significant input from toxicology and chemistry experts. In this poster we present an automated methodology for *toxicophore* identification by mining descriptions of activating structural features directly from toxicity datasets. The method is based on *emerging pattern* mining (Dong and Li 1999), a technique that is well known to computer science, but is relatively new to chemistry. For any data that can be expressed as a series of binary properties, emerging pattern mining can be applied to extract patterns of those properties that occur more frequently in one dataset compared to another. For toxicophore, or indeed pharmacophore, formation mining emerging patterns from structural fingerprints provides a means of identifying structural features that distinguish active compounds from inactives. *Jumping-emerging patterns* are emerging patterns that are unique to the active set and therefore represent structural features that are only present in active compounds. The method described here employs the Horizon-Miner algorithm (Li, Dong & Ramamohanarao 2001) and border-differential operation (Dong & Li. 2005) to generate the minimal and maximal borders of a set of jumping-emerging structural patterns. Using the minimal jumping-emerging patterns it is possible to cluster toxic compounds into groups defined by the presence of shared structural features that occur exclusively in the actives. From these clusters it is then possible to derive larger and more complete descriptions of the distinguishing structural features that will be recognisable to toxicologists. A method has been developed to identify hierarchical relationships between clusters and their associated jumping-emerging patterns, which has enabled families of structural feature descriptions to be arranged into trees. The root of each tree represents the most general and most commonly occurring structural feature description in the family. By inspecting clusters further down the tree, it is possible to extend the significant structural features to further distinguish sets of toxic compounds. This may provide additional structural detail that is relevant to the toxicological endpoint, and useful for compiling a new alert. The methodology has been tested on a number of datasets composed of known Ames mutagens. These tests have shown the method to be effective at clustering the datasets around minimal jumping-emerging structural patterns and finding larger descriptions of the significant structural features. The resulting descriptions of the significant structural features have been shown to be related to some of the known alerts for the Ames mutagenicity endpoint.

References:

Dong, G. & Li, J. (2005). "Mining border descriptions of emerging patterns from dataset pairs". *Knowledge and Information Systems*, **8**, 178-202.

Li, J., Dong, G. & Ramamohanarao, K. (2001). "Making Use of the Most Expressive Jumping Emerging Patterns for Classification". *Knowledge and Information Systems*, **3**, 131-145.

Dong, G. & Li, J. (1999). "Efficient mining of emerging patterns: Discovering trends and differences". pp. 43- 52. ACM.

Poster 11

QM/MM study of Oleamide (OLE) hydrolysis and Oleoylmethylester (OME) hydrolysis catalysed by Fatty Acid Amide Hydrolase.

Jitnapa Sirirak[#], Alessio Lodola[*] , Marco Mor[*], Fred Manby[#] and Adrian Mulholland[#]

[#]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, BS8 2AL*
[*] *Dipartimento Farmaceutico, Università degli Studi di Parma*, *43100, Parma, Italy.*

Due to the unusual serine-serine-lysine triad for catalysis of FAAH which made it be able to hydrolyse amides and esters with similar catalytic efficiencies[1], FAAH can control the magnitude and duration of signals communicated by endogenous fatty acid amides within a complex milieu of structurally similar lipid natural products[5]. Hence, FAAH become one of the attractive therapeutic targets for the treatment of pain and related neutral disorders. Here, QM/MM on adiabatic mapping simulations were carried out to obtain two-dimensional potential energy surfaces of oleamide (OLE) and oleoylmethylester (OME) hydrolysis reaction catalysed by FAAH at BH&HLYP/6-31+G*/CHARMM27, MP2/aug-cc-pvdz//BH&HLYP/6-31+G*/CHARMM27, SCS-MP2/aug-cc-pvdz// BH&HLYP/631+G*/CHARMM27, B3LYP/6-31+G*/CHARMM27, MP2/aug-cc-pvdz//BH&HLYP/6-31+G*/CHARMM27 and SCS-MP2/aug-cc-pvdz//BH&HLYP/6-31+G*/CHARMM27 level. Nudged Elastic Band method (NEB) was also calculated at B3LYP/6-31+G*/CHARMM27 by using the structures from adiabatic mapping results for OLE and OME in order to determine the minimum energy pathway and the location of transition state of these reactions. The results show that the mechanism of both OLE and OME hydrolysis consists of (1) deprotonation of Ser217 by Lys142, (2) formation of the tetrahedral intermediate concerted with proton transfer from Ser241 to Ser217, (3) proton transfer from Ser217 to the leaving group concerted with leaving group elimination, and (4) protonation of the Ser217 alkoxide anion by the ammonium side chain of Lys142 to return to their initial states. The overall energy barriers of these reactions for OLE and OME calculated with NEB were similar at 23 and 24 kcal/mol respectively. The formation of the tetrahedral intermediate was the rate limiting step for OLE whereas proton transfer from Ser217 to the leaving group concerted with leaving group elimination was the rate limiting step for OME.

References:
1. M. P. Patricelli, B. J. Cravatt, *Biochemisty*, 1999, **38**, 14125
2. C. Walsh, *Enzymatic Reactions Mechanisms*, W. H. Freeman and Co., New York 1979.
3. D. R. Corey and C. S. Craik, *J. Am. Chem. Soc.,* 1992, **114**, 1784.
4. M. L. Bender, G. R. Schonbaum and B. Zerner, *J. Am. Chem. Soc.*, 1962, **84**, 2540.
5. M. K. McKinny, B. F. Cravatt, *J. Biol. Chem*., 2003, **278**, 37393

Poster 12

Folding and Dynamics of Lantibiotic Peptide Analogues

Eleanor Turpin, *Jonathan Hirst*

School of Chemistry, University of Nottingham, Nottingham

Naturally occurring antibiotic peptides are a source of new drug leads to help overcome antibiotic resistance. Lantibiotics are antibiotic peptides secreted by gram-positive bacteria, characterised by unusual dehydrated amino acids and thioether rings formed by post-translation modifications. Rings A and B are conserved across type A lantibiotics and have been identified, in NMR experiments with nisin[1], as binding to lipid II by forming a cage around to the pyrophosphate group. In our work we have performed molecular dynamics simulations to study analogues of the first 12 residues of nisin modified so that the thioether bonds have been replaced with disulfide bonds between cysteines. The chirality of the four cysteine residues was varied to see how this influences disulfide bond formation.

For six of the eight analogues with D-Cys3 there is an interaction between S_3-S_7 (ring A). The sulfur atoms in Cys3 and Cys7 are brought together by a type-IV β-turn between residues D-Cys3-Ile4-Ala5-Leu6, characterised by dihedral residue angles (φ, ψ) of (-90° ± 20°, -65° ± 31°) for Ile4 and (-95° ± 21°, -77° ± 32°) for Ala5. L-Cys8 with L-Cys11 favours conformations corresponding to a S_8-S_{11} (ring B) interaction. The interaction between S_8-S_{11} is due a type-IV β-turn between Cys8-Pro9-Gly10-Cys11, stabilized by a hydrogen bond between 8(O) an 11(HN)/12(HN). Nisin analogue 12 (L-Cys3-L-Cys7-D-Cys8-D-Cys11) has a global minimum on the relative energy surface corresponding to the simultaneous formation of the S_3-S_7 and S_8-S_{11} disulfide bond. Comparisons of representative conformers of this analogue with the NMR structure of wild-type nisin in complex with lipid II found backbone RMSD between 0.6 and 1.7 Å, suggesting that the disulfide analogues are worth pursuing further as possible new peptide antibiotics[2].

References:

(1) S. T. D. Hsu, E. Breukink, E. Tischenko, M. A. G. Lutters, B. de Kruijff, R. Kaptein, A. M. J. J. Bonvin and N. A. J. van Nuland, *Nature Structural & Molecular Biology* (2004) **11**, pp.963-967.

(2) E. R. Turpin, B. B. Bonev and J. D. Hirst, *Biochemistry* (2010), in press, DOI: 10.1021/bi101214t