

MGMS and RSC MMG Young Modellers' Forum 2009

ORAL PRESENTATIONS

Talk 1

Reactivity and Interaction Studies of Glycolysed Generation 3.5 PAMAM Dendrimers with TLR4-MD2 Complex

Teresa Barata¹, Steve Brocchini¹, Ian Teo², Sunil Shaunak² and Mire Zloh¹

¹*School of Pharmacy, University of London, 29/39 Brunswick Square, London, WC1N 1AX*

²*Faculty of Medicine, Imperial College London, Hammersmith Hospital, Ducane Road, W12 0NN.*

Polyamidoamine (PAMAM) dendrimers are hyper-branched molecules and their glycolysed analogues have been reported to have immuno-modulatory and antiangiogenic activity by preventing the LPS (lipopolyssacharide) triggered inflammatory response (Shaunak *et al*, 2004). The saccharide loading on dendrimer end groups has been estimated experimentally, but the position and the distribution of sugars has yet to be determined. Hence it is not possible to determine detailed structure-activity correlations. The aim of these studies is to understand the structural features of glycosylated dendrimers that contribute to their activity and to identify the interaction sites with the biological targets to aid the design of more potent dendrimers derivatives.

We have developed a method that allows the *in silico* generation of initial 3D models from the sequence of modified dendrimers. The study of the conformational flexibility and consequent availability of the sugars for possible interactions with the protein of the interest is studied by the molecular dynamics of fully solvated molecules. The molecular properties of the resulting modified dendrimers are determinant in their interaction with the biological targets and were estimated for all generated structures. Comparative studies of models with different loadings of sugars and different distributions were also performed.

Simultaneously, surface studies and docking studies using methods implemented in GRID and PatchDock have been performed with the components of the LPS recognition system with which the dendrimer is believed to interact. The 3D models of the dendrimer's conjugates are now being used to study the interactions between the dendrimer and TLR4- MD2 complex. These important interactions

should provide information on the molecular mechanism for the interesting biological activities shown by these saccharide modified PAMAM dendrimers.

Reference:

Shaunak, S., Thomas, S. *et al.* (2004). "Polyvalent dendrimer glucosamine conjugates prevent scar tissue formation". *Nature biotechnology* **22**: 977-984.

Talk 2

Molecular Dynamics Simulations of the β_2 -adrenergic Receptor in Complex with a Peptide Derived from the $G\alpha$ Subunit of the Heterotrimeric G_s -protein

Angela Götz¹, Peter Gmeiner¹, Tim Clark², Harald Lanig²

¹*Department of Chemistry and Pharmacy, Emil Fischer Center, Friedrich Alexander University, Schuhstrasse 19, 91052 Erlangen, Germany.*

²*Department of Chemistry and Pharmacy, Computer Chemistry Center, Friedrich Alexander University, Nögelsbachstrasse 25, 91052 Erlangen, Germany.*

The β_2 -adrenergic receptor is a member of the large family of G-protein coupled receptors (GPCRs). It belongs to the subfamily of rhodopsin-like receptors. Ligand binding to the transmembrane domain leads to conformational changes within the receptor and to activation of intracellular G-proteins.

In 2007, the β_2 -adrenergic receptor was crystallized in its inactive conformation and the structure was widely-used for drug design studies. One year later, opsin was crystallized in its G-protein-interacting conformation. It was stabilized by an 11 amino acid synthetic peptide derived from the C-terminus of the $G\alpha$ subunit of the G_t -protein interacting with the cytoplasmic part of the seven transmembrane (TM) helices.

The aim of our studies was to construct a receptor-peptide complex derived from the crystal structure of the β_2 -adrenergic receptor and the C-terminus of the $G\alpha$ subunit of the G_s -protein. Following, the impact of the C-terminal end of $G\alpha$ on the structure of the ligand binding site and the interactions between receptor and the synthetic agonist Isoprenaline should be investigated.

By applying molecular dynamics simulations, the cytoplasmic half of TM6 of the β_2 -adrenergic receptor was shifted outwards of the helical bundle by 6-7 Å to accommodate the C-terminal end of $G_s\alpha$ in the cytoplasmic part of the receptor as observed in the crystal structure of opsin. Following, the peptide-receptor complex as well as the original crystal structure of the β_2 -adrenergic receptor were subjected to molecular dynamics simulations in a DOPC membrane. The simulations in both systems were carried out in presence and absence of Isoprenaline.

Although the peptide, interacting with the cytoplasmic part of the receptor, is not in close proximity to the ligand binding site, it significantly affects the structure of the binding pocket. Hydrogen bond patterns within the receptor and between receptor and ligand as well as root mean square deviation and

fluctuation data are used to characterize the binding crevice. In addition, the binding site is divided into three different layers parallel to the membrane orientation. For each layer the helix-helix distances were calculated to monitor in detail changes in the structure of the binding pocket.

In this presentation, the results of the analyses will be discussed to elucidate the influence of the peptide on structure and dynamics of the ligand binding site.

Talk 3

Minor Groove of DNA as Target for Drug Design

Gudrun M. Spitzer¹, Julian E. Fuchs¹, Adam Biela², Christoph Kreutz¹; Johannes Kirchmair¹, Gerhard Klebe², Klaus R. Liedl¹

(1) Faculty of Chemistry and Pharmacy, University of Innsbruck, Innrain 52a, A-6020 Innsbruck, Austria

(2) Institute of Pharmaceutica Chemistry, Philipps-University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany

Small minor groove binding molecules have been found to influence DNA-dependent processes. Their affinity is high enough to prevent transcription factors from interaction with the DNA. Some of the minor groove binders can be even designed to target only specific DNA sequences. Thereby, they are able to influence transcription, which provides the possibility to systematically regulate the synthesis of proteins. Unfortunately, all highly sequence-selective minor groove binders have the same structural scaffold, which includes a number of amide bonds, and therefore lack metabolic stability. We adopted methods originating from chemoinformatics to merge experimental data with our knowledge on the dynamics of the biomolecular interface of DNA. We demonstrated that we are able to reproduce sequence specificity by a pharmacophore modeling approach [1]. We built a comprehensive database containing all known minor groove interactions within complexes of DNA with different ligands and water molecules [2]. An analysis of the database revealed that the known minor groove binding scaffolds do not optimally fit the hydrogen bonding patterns presented by DNA [3]. Therefore, we applied shape-based screening to discover new ligands. We validated the screening hits by isothermal titration calorimetry to understand the thermodynamics of ligand binding [4].

References:

[1] Spitzer GM, Wellenzohn B, Laggner C, Langer T, Liedl KR, *J. Chem. Inf. Model.* **2007** May;47(4):1580-1589.

[2] Spitzer GM, Fuchs JE, Markt P, Kirchmair J, Wellenzohn B, Langer T, Liedl KR, *ChemPhysChem*. **2008** Sep;9(18):2766-2771.

[3] Spitzer GM, Wellenzohn B, Markt P, Kirchmair J, Langer T, Liedl KR, *J. Chem. Inf. Model.* **2009** Mar;49(4):1063-1069

[4] Spitzer GM, Fuchs JE, Biela A, Wellenzohns B, Klebe G, Liedl KR, to be published.

Talk 4

Fe(H)₂(diphosphine)(diamine) Complexes as Attractive Catalysts for The Hydrogenation of Ketones?

Hsin-Yi Tiffany Chen^a, Chao Wang^b, Devis Di Tommaso^a, Graeme Hogarth^a, Jianliang Xiao^b, and C. Richard A. Catlow^a

^aDepartment of Chemistry, University College London, 20 Gordon Street, London, UK WC1H 0AJ

^bDepartment of Chemistry, University of Liverpool, Crown Street, Liverpool, UK L69 3BX

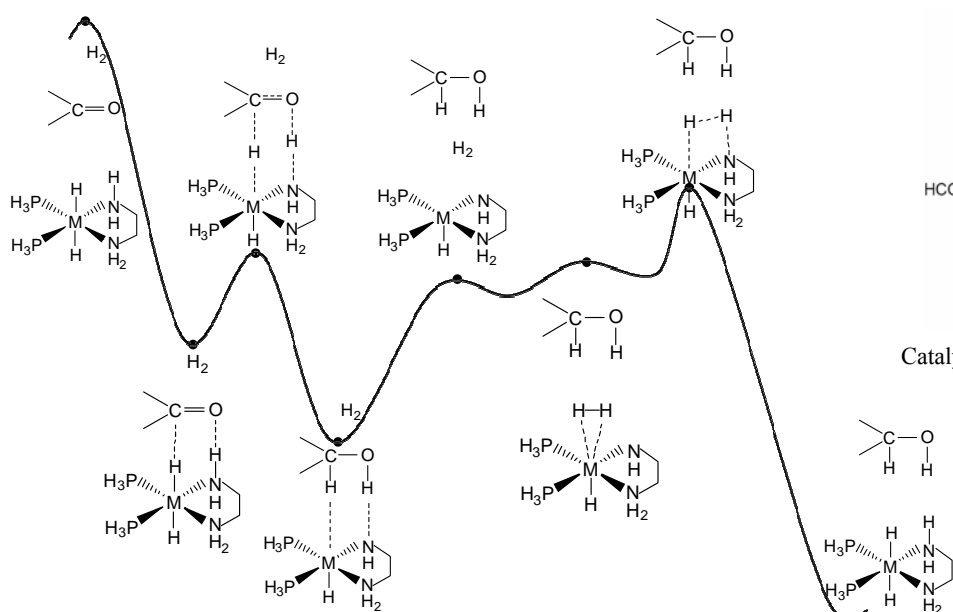
Enantiomerically pure alcohols play a significant role in pharmaceutical and agrochemical industries. The most effective and efficient class of catalysts for enantioselective hydrogenation of pro-chiral ketone to produce chiral alcohols is the *trans*-[Ru^{II}(H)₂(diphosphine)(diamine)] complexes discovered by Noyori and co-workers. This catalyst is composed of one central metal and appropriate ligands. Accordingly, an economical analogue of [Fe^{II}(H)₂(diphosphine)(diamine)] complexes with high enantioselectivity and activity is desirable, and possibly more sustainable. To understand the similarity between [Fe^{II}(H)₂(diphosphine)(diamine)] and [Ru^{II}(H)₂(diphosphine)(diamine)] complexes, the comparison of their structural form, electronic property, and the catalytic cycle is of interest.

The activity of Fe catalysts might be tuned by varying electronic properties of the attached ligands, which may change the hydricity of the hydride. To demonstrate the importance of the electronic properties of ligands, the imine reduction, catalysed by Ir catalysts having ligands with different electronic properties, is investigated due to the obvious change of activity.

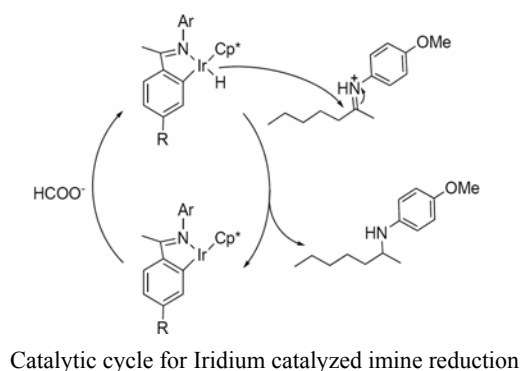
The density functional theory at PBE, PBE0, and B3LYP levels has been employed by using DMol³ and Gaussian program packages.

The results show that *trans*-[Fe^{II}(H)₂(diphosphine)(diamine)] is more stable than *cis*-[Fe^{II}(H)₂(diphosphine)(diamine)] by 5~13 kcal/mol with a low-spin state, which reveals identical characterization to [Ru^{II}(H)₂(diphosphine)(diamine)]. In the Ru- and Fe- based catalytic cycle, the energy barriers in the hydrogen transfer and H₂ splitting processes are similar. This indicates a thermodynamic similarity between the two systems. In the imine reduction catalysed by Ir catalyst, correlation between the electronic properties of the ligands and the catalytic activities has been searched, which might assist in designing ligands to improve Fe catalyzed reduction.

To conclude, *trans*-[Fe^{II}(H)₂(diphosphine)(diamine)] complex tuned by suitable ligands could be a viable substitution for the *trans*-[Ru^{II}(H)₂(diphosphine)(diamine)] complex.



Catalytic cycle catalysed by diphosphine-diamine transition metal complexes



Catalytic cycle for Iridium catalyzed imine reduction

Talk 5

Defect and Transport Properties of LiFePO₄ Battery Materials

Veluz Maria Hart Prieto, Steve C. Parker and M. Saiful Islam

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

The olivine-structured material, LiFePO₄, has attracted considerable interest as a cathode material for rechargeable lithium batteries. The major challenge for generating polycrystalline Li_xFePO₄ cathode materials is to generate the microstructure that produces fast lithium diffusion and high electrical conductivity. We discuss recent atomistic simulations modelling the two end-members, LiFePO₄ and FePO₄. We have begun to study the free energy barrier to lithium diffusion for the first time taking the vibrational properties and temperature into account. The free energy barriers at 300K show the anisotropic nature of the material, as predicted, but we find a lower barrier to diffusion. We have also studied intrinsic disorder which allows us to assess the bulk properties of the material. We predict 'anti-site' disorder where lithium and iron are on opposite nearest neighbour sites are the most common defects. As a result, this prediction suggests a possible blocking mechanism for lithium diffusion. We have also begun simulations to predict the microstructure by initially studying the structures and stabilities of the low index surfaces of the two end-members. This has allowed us to investigate the effect of lithium content on the predicted morphologies. Interestingly, we find that depletion of lithium favours elongated hexagonal prism morphologies, which is in good accord with experiment. Finally, I will describe preliminary work on the electron transfer process. Using the framework of Marcus' theory, we show how the rate of electron transfer at the surfaces differs from the bulk.

Thus the success of these simulations demonstrates that these methods provide a useful complement to experiment.

Talk 6

Assigning Stereochemistry Using GIAO NMR Shift Calculation

Steven G. Smith and *Jonathan M. Goodman*

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

GIAO NMR chemical shift calculation has been applied to the challenging task of reliably assigning stereochemistry to complex molecules such as natural products. ^{13}C and ^1H NMR shifts have been calculated for up to 64 diastereoisomers of over 30 different molecules, including 13 originally misassigned natural products, and the results have been used to guide the development of new methodology for stereostructure assignment using calculated NMR shifts. Two new parameters, CP3 and CDP4, have been developed for quantifying the agreement of a given candidate structure to the experimental spectrum to be assigned; both of these parameters are significantly more successful than the commonly used correlation coefficient and mean absolute error.

The CP3 parameter is designed for the case of assigning two spectra to two possible structures, a situation that occurs commonly in synthetic organic chemistry. By comparing differences in calculated shifts to differences in experimental shifts in a particular way, this parameter simultaneously reduces systematic errors in the calculated shifts and focusses on the particular experimental shifts that are most useful for structure assignment. The results can then be converted into a probability that each proposed assignment is correct using Bayes' Theorem in conjunction with a knowledge of the values of CP3 expected for correct and incorrect assignments. CP3 typically makes correct assignments with over 99% confidence.¹

The CDP4 parameter is designed for the challenging situation in which one unknown compound is to be assigned to one of many possible diastereoisomers. By applying a probability-based approach to each individual ^{13}C or ^1H shift, CDP4 focusses on the key shifts for assignment and directly assigns a probability to each of the candidate structures. As with CP3, correct assignments are generally made with high confidence.

Both of these parameters give excellent results with NMR shifts calculated using a computationally inexpensive procedure in which single point calculations on molecular mechanics geometries replace time consuming ab initio geometry optimisation. Examples of their application to the assignment of several complex and originally misassigned natural products will be presented.

References:

1. Smith, S. G and Goodman, J. M. Assigning the Stereochemistry of Pairs of Diastereoisomers Using GIAO NMR Shift Calculation, *J. Org. Chem.* (2009) **74**, 4597-4607.
2. Smith, S. G., Paton, R. S., Burton, J. W. and Goodman, J. M. Stereostructure Assignment of Flexible Five-Membered Rings by GIAO ^{13}C NMR Calculations: Prediction of the Stereochemistry of Elatényne, *J. Org. Chem.* (2008), **73**, 4053–4062.

Talk 7

Predicting Fragment Positions in Binding Sites Using MCSS (CHARMm)

Kamran Haider^{*}, Hugues-Olivier Bertrand[§], Roderick E. Hubbard^{*}

^{*}York Structural Biology Laboratory,

[§]LS Modelling and Simulations, Accelrys

There has been a huge interest in fragment-based ligand design (FBLD) strategies over the past few years. Fragment-based design involves screening of drug targets against small weakly binding fragments. The fragments that bind the target are then grown and evolved into lead compounds. Alongside experimental methods, computational approaches have been developed to predict binding modes of fragments and to map active sites for their affinity to various functional groups. As with other docking and scoring methods, these approaches have their strength and weaknesses. MCSS (multiple copy simultaneous search)¹ is a well-established computational method for predicting binding orientation of fragments in active sites of proteins. The method finds energetically favourable positions for a functional group by randomly distributing and simultaneously minimizing multiple copies (1,000-10,000) in the binding site. During energy minimization, copies of the functional group do not see each other but they experience the forcefield from the protein. The resulting positions are then ranked according to their interaction energies.

The new experimental datasets for protein-fragment binding provide an opportunity for assessing its efficiency and exploring its utility in the context of fragment screening. Here we present results from MCSS calculations on two different datasets – a set of solvent bound crystal structures of thermolysin (TLN)³ and a set of Hsp90 structures with bound fragments/ligands⁴. Where MCSS was able to reproduce experimental positions in nearly all the test cases, it was noticed that the ranking of such poses was not necessarily correct. For seven different experimental positions of four solvent probes in TLN active site, the MCSS calculations reproduced at least four, at an RMSD of < 2.0Å and ranked them within top 3 poses. For Hsp90, initially, the ranks of experimentally close fragment/ligand positions were quite low but further optimization of the protocol and post-processing showed significant improvement. It was noticed that by introducing pharmacophore models based on conserved protein-ligand interactions, in a post-processing filtering step, a reasonable reordering of poses can be obtained. The discrepancy in the scoring is partly due to the lack of desolvation terms and other factors such as flexibility of the active site. We therefore, intend to further investigate the re-scoring of MCSS clusters based on MM/PBSA calculations and by using more complex pharmacophore queries during post-processing.

References:

1. Miranker, A., Karplus, M. Functionality Maps Of Binding-Sites - A Multiple Copy Simultaneous Search Method. *Proteins-Structure Function and Genetics* **1991**, 11 (1), 29-34.

- English, A. C., Groom, C. R., Hubbard, R. E. Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Engineering* **2001**, 14 (1), 47-59.
- Brough, P. A., Barril, X., Borgognoni, J., Chene, P., Davies, N. G. M., Davis, B., Drysdale, M. J., Dymock, B., Eccles, S. A., Garcia-Echeverria, C., Fromont, C., Hayes, A., Hubbard, R. E., Jordan, A. M., Jensen, M. R., Massey, A., Merrett, A., Padfield, A., Parsons, R., Radimerski, T., Raynaud, F. I., Robertson, A., Roughley, S. D., Schoepfer, J., Simmonite, H., Sharp, S. Y., Surgenor, A., Valenti, M., Walls, S., Webb, P., Wood, M., Workman, P., Wright, L. Combining Hit Identification Strategies: Fragment-Based and in Silico Approaches to Orally Active 2-Aminothieno 2,3-d pyrimidine Inhibitors of the Hsp90 Molecular Chaperone. *Journal of Medicinal Chemistry* **2009**, 52 (15), 4794-4809.

Talk 8

Application of Simulation Methods for the Identification of the Allosteric Binding Site in Human Glucokinase

Nadia Vahdati (1), Jonathan W. Essex (1) and Richard Ward (2)

(1) *School of Chemistry, University of Southampton, SO17 1BJ, UK*

(2) *AstraZeneca, Alderley Park, Macclesfield, Cheshire SK10 4TF, UK*

Allosteric binding sites, in generic terms, refer to binding sites distinct from the active site, where binding of an effector can enhance or decrease enzyme's activity. From a drug design point of view targeting allosteric binding sites can offer advantages in terms of selectivity, saturability, and the opportunity for discovering new chemotypes. Although experimental methods can be successful in identifying allosteric binding sites, the cost and time associated with such experiments limits their use. As such, there is a great interest in computational approaches to identify novel allosteric binding sites.

Human glucokinase (GLK) is an allosteric enzyme that catalyses the ATP-dependent phosphorylation of glucose which plays a critical role in glucose homeostasis. Inactivating mutations in the GLK gene have been linked to maturity-onset diabetes of the young, type 2 (MODY2) (1) and the activating mutations to persistent hyperinsulinemic hypoglycemia of infancy (PHHI) (2). The discovery of allosteric activators for this enzyme in the recent years has offered a promising new therapeutic approach for the treatment of diabetes (3).

Here the application of computational approaches, namely molecular dynamics (MD) and normal mode analysis (NMA) will be presented in the context of GLK, in a quest to predict the allosteric binding site without the bias of using the prior experimental knowledge in the study.

References:

- N. Vionnet *et al.*, *Nature* **356**, 721 (1992).
- A. L. Cuesta-Munoz *et al.*, *Diabetes* **53**, 2164 (2004).
- K. Kamata, *et al.*, *Structure* **12**, 429 (2004).

Talk 9

Screening for Near-Native Protein-Protein Docking Complexes: Use of Residue-Pair Propensity Matrices

Ja-won Song and *Tom Blundell*

Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA

High quality protein-protein interaction information has been difficult to obtain as it depends largely on crystallization of multiprotein complexes. Thus, as an alternative approach, many computational methods (protein-protein docking) have been developed, which aim to predict the structures of protein complexes, given structures in their unbound form. However, the major bottleneck remains the challenge of distinguishing near-native complexes from the decoys. The scoring function needs to be less sensitive to precise atomic positions and at the same time fast enough to be used extensively in conjunction with conformational search algorithms in conventional docking programs. Here we use residue pair propensity matrices to generate a scoring function. This essentially consists of a value which suggests how likely it is to find a particular residue pair at the protein-protein interface. The matrices have been calculated from an in-house protein-protein interaction database, PICCOLO. Further matrices have been then generated according to the type of protein complex, whether it is a heterodimer or a homodimer and whether it is an obligate or a transient complex. Thus, separate matrices have been generated for antibody-antigen and enzyme-inhibitor complexes. Results from using these matrices show improved probability of identifying near-native complexes. For example, from 100 complexes generated for hydrolase calcineurin (PDB ID: 1AUI), careful choice of scoring function matrix allowed 82% of decoy complexes to be filtered out. Furthermore, it has been found that even if all 2000 complexes are generated, the same scoring function allows 67% of decoys to be removed. The scoring function developed here will greatly reduce the time taken in virtual screening and lead identification processes of drug discovery.

Talk 10

QM/MM Modelling of the Reaction Mechanism of Chorismate Synthase

Narin Lawan, Fred Manby and Adrian J. Mulholland

Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS

Chorismate synthase is an enzyme responsible for the biosynthesis of chorismate. Chorismate is a key intermediate in the biosynthesis of many aromatic compounds, including the aromatic amino acids. The reaction mechanism of this enzyme is still unclear: different proposals have been made. We have studied the mechanism with QM/MM on umbrella sampling molecular dynamics simulations and adiabatic mapping calculations, at the AM1/CHARMM27 level. Some selected structures from umbrella sampling and adiabatic mapping calculations were used for higher level (B3LYP/6-31G*/CHARMM27, MP2/aug-cc-pvdz//B3LYP/6-31G*-CHARMM27, and scs-MP2/aug-cc-

pvdz//B3LYP/6-31G*-CHARMM27) QM/MM calculations [3]. A mechanism in which proton transfer from EPSP (5-enolpyruvylshikimate-3-phosphate) to FMNH (flavin mononucleotide) is the first step followed by release of the phosphate group from EPSP is apparent from AM1/CHARMM27 calculations. Higher level calculations suggest that the proton transfer (B3LYP/6-31G*/CHARMM27, MP2/aug-cc-pvdz//B3LYP/6-31G*-CHARMM27, and scs-MP2/aug-cc-pvdz//B3LYP/6-31G*-CHARMM27) could be the first step or that the two may be concerted. The results provide insight into the mechanism of this enzyme.

References:

1. J. Maclean, and S. Ali, *Structure*, 2003, **11**, 1499-1511.
2. A. Osborne, R. N. F. Thorneley, C. Abell, and S. Bornemann, *J. Biol. Chem.* 2000, **275**, 35825–35830.
3. F. Claeysens, JN. Harvey, FR. Manby, RA. Mata, AJ. Mulholland, KE. Ranaghan, Schütz M, Thiel S, Thiel W, Werner H-J, *Angewandte Chemie Intl Edn* 2006, **45**:6856-6859.

Talk 11

Computational Studies of Water-Mediated Interactions in Ionotropic Glutamate Receptors

Michelle A. Sahai and *Philip C. Biggin*

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

Ionotropic glutamate receptors (iGluR) are ligand-gated ion channels which mediate excitatory synaptic transmission. These receptors are of great interest as they are central to fast synaptic neurotransmission in the brain and are thought to be essential in memory and learning. Furthermore, they have been associated with abnormal neuronal activity leading to a wide range of diseases, including Parkinson's, epilepsy, and ischemic stroke. Crystal structures of the ligand-binding domain not only reveal the manner in which ligands interact with iGluRs, but also highlight the importance of water molecules in ligand-receptor binding. Despite the wealth of structural information, the precise role of water molecules in mediating receptor-ligand interactions remains unclear. For example, although the ligands glutamate and AMPA are structurally similar, crystal structures revealed that AMPA adopts a slightly different binding-mode with a water molecule assuming a position that would normally be occupied by the γ -carboxyl group of glutamate [1]. It is possible water could play a significant role in the binding of other ligands as well. Thus, an understanding of the stability of water molecules within the binding site is critical to improved drug design.

This talk will describe a study carried out using MD simulations followed by a theoretical DFT investigation. The study focuses on the interaction energy of the protein-water-ligand complex with the aim of understanding the nature of its stabilization and quantification of the role of dispersion contribution. Interaction energy is determined as a sum of interaction energies of representative fragments containing protein amino acid residue(s), water molecules and ligand, as well as directly for

the whole protein-water-ligand complex. This information will be extremely useful not only in the design of novel drugs for iGluRs but more generally in the field of drug design.

References:

- [1] Armstrong, N.; Gouaux, E. Mechanisms for activation and antagonism of an AMPA Sensitive glutamate receptor: Crystal structures of the GluR2 ligand binding core. *Neuron* **2000**, 28 (1), 165-181.

Talk 12

Evaluation of a Bayesian Inference Network for Ligand-Based Virtual Screening

Christoph Mueller, *Beining Chen*, *Peter Willett*

Departments of Chemistry and of Information Studies, University of Sheffield, Sheffield S10 2TN, United Kingdom

The early stages of a drug discovery programme involve the use of virtual screening methods. A virtual screening method ranks the molecules of a database in decreasing order of similarity to a reference molecule. Molecules which are then found among the top-ranked ones are, according to the similar property principle, more likely to show the same activity as the reference molecule (Johnson and Maggiora, 1990).

In this study a new approach for virtual screening has been developed which makes use of methods from probabilistic information retrieval, in particular Bayesian inference networks (Turtle and Croft, 1991). In information retrieval these networks are used to retrieve relevant documents from large-scale databases given a query formulation. In the present study, experiments were conducted in order to measure the performance of the newly developed method for retrieving chemical molecules from large-scale databases given an active reference structure. The results of these experiments were evaluated with nonparametric tests and compared with the results obtained using a standard virtual screening method (Willett et al., 1998).

Thus far, the study has shown that the Bayesian inference network method achieves better results than a standard method when screening homogeneous activity classes, but that the converse applies with heterogeneous activity classes. This latter case is usually the starting point in a drug discovery programme, meaning that further experiments are needed in order to improve the performance of the Bayesian inference network.

References:

- Johnson, M.A. & Maggiora, G.M. (1990). *Concepts and applications of molecular similarity*. Wiley New York.
- Turtle, H. & Croft, W.B. (1991). "Evaluation of an inference network-based retrieval model". *ACM Transactions on Information Systems (TOIS)*, **9** (3), 187-222.
- Willett, P., Barnard, J.M. & Downs, G.M. (1998). "Chemical similarity searching". *Journal of Chemical Information and Computer Science*, **38** (6), 983-996.