MS01 O1

Comparative and Structural Genomics to Explore the Evolution of Protein Function. Corin Yeats, Gabrielle Reeves, Oliver Redfern, Juan Ranea and <u>Christine Orengo</u>. Department of Biochemistry and Molecular Biology, University College London, UK.
E-mail: orengo@biochem.ucl.ac.uk

Key words: domain families, genome analysis

How can structural genomics initiatives target relatives from protein families in a manner that increases our understanding of the evolution of protein structures and functions within families? There are now nearly 100,000 domain structures in the CATH database which can be classified into approximately 2000 evolutionary superfamilies. Using HMM based methods to predict structural relatives in completed genomes we observe that more than half of the domain sequences can be assigned to known structural families in CATH [1]. This structural mapping allows us to probe more deeply into the evolutionary history of these families and their differential expansion in the genomes. Although, there are about 140 structural families that are common to all kingdoms of life, a small proportion of these (<20) are highly recurrent accounting for nearly 50% of domain structure annotations in the genomes. Furthermore, many of these very large families are observed to be highly structurally and functionally divergent, though functional divergence is generally limited to changes within a COG major functional class rather than a complete change of functional class. Structural analyses of the most divergent enzyme families reveals a mechanism whereby small accretions of secondary structural elements along the polypeptide change during evolution, are amplified in their impact on the structure through co-localisation in 3D. These secondary structure embellishments often modify the geometry of the active site or the structural characteristics on the surface of the protein promoting different protein-protein interactions [2]. Whilst local structure comparison methods and 3D-templates based on functional sites have difficulty in distinguishing functional subgroups within a structural superfamily, template methods based on global structural comparison show increased specificity and selectivity and reflect the ability of these approaches to capture a broader range of surface characteristics. Sequence based methods for predicting functional subgroups within superfamilies identifies functionally distint subfamilies with no close structural relatives available for homology modelling. These can be targetted by the structural genomics initiatives to improve our understanding of structure-function space.

[1]Comprehensive genome analysis of 203 genomes provides structural genomics with new insights into protein family space. Marsden RL, Lee D, Maibaum M, Yeats C, Orengo CA. (2006) Nucleic Acids Res 34, 1066-1080.
[2]Structural Diversity of Domain Superfamilies in the CATH Database. G.A. Reeves, T.J. Dallman, O.C. Redfern, A. Akpor & C.A. Orengo. (2006) Journal of Molecular Biology 360, 725-41.

MS01 O2

What is the Value of Automatic Protein Structure Prediction? <u>Johannes Söding</u>, *Max-Planck-Institute for Developmental Biology, Tübingen, Germany.* Present address: *Gene Center, University of Munich, Germany.*

E-mail: soeding@lmb.uni-muenchen.de

Keywords: homology modeling, fold recognition, remote homology detection, structure prediction

This presentation will give a general introduction of automatic structure prediction methods and show up the possibilities and limitations of these powerful new methods. After explaining the basic procedure for manual structure prediction, I will give an overview of state-of-the-art automatic prediction methods that are easy to use via web servers. The bi-annual, blind structure prediction benchmark CASP (Critical Assessment of Structure Prediction) will be introduced and results for automatic methods in the 2006 competition will be summarized. An overview of some of the most successfull servers is given, including our own HHpred sever for protein structure and function prediction (http://hhpred.tuebingen.mpg.de). Several examples of applications will demonstrate their potential for structural biology.

MS01 O3

Structural Assignment of Spectra by Characterization of Conformational Substates in MbCO. <u>M. Devereux</u>, M. Meuwly, *Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056, Basel, Switzerland.* E-mail: Michael.Devereux@unibas.ch

Keywords: Molecular Dynamics; Ligand Binding; Protein Modeling

Residue motions of the distal heme pocket of the oxygen-storing protein Myoglobin have been shown to influence protein function, control ligand rebinding rates [1], and have been implicated in ligand recognition [2]. In Myoglobin systems binding NO (MbNO), experiment indicates that rebinding from different conformational substates follows distinct kinetics [1], which is likely to also hold true for carbonmonoxy Myoglobin (MbCO). In contrast to the former, for MbCO both the ligand bound (MbCO, A-state) and unbound (Mb···CO, B-state) have been characterized by x-ray crystallography. Because ligand binding and unbinding are transient processes in nature, it is difficult to experimentally characterize both structural and dynamic properties of the system. Atomistic simulations using validated force fields provide additional insight [3,4]. In further studies [2,5], Molecular Dynamics simulations have linked theoretical motions of residues within the heme pocket to changes in observed spectroscopic A-states. The distinct A-states of the bound CO ligand were attributed to different orientations and protonation states of the adjacent HIS64 moiety. Here we characterize the bound states A_0, A_1 and A_3 using a combination of molecular dynamics simulations and Density Functional Theory calculations. Calculated absorption bands were derived from different configurations for comparison with experimental results. The experimental data is then used to guide refinement of the original CHARMM model. Particular attention is paid to the charge model of the bound CO ligand, key to representing interaction with the local binding site.

[1]   S. Kim, M. Lim, *J. Am. Chem. Soc.* 2005, *127*, 8909.
[2]   A. Loccisano, O. Acevedo, et al., *J. Mol. Graph. Model.,* 2004, *22*, 369.
[3]   D. Nutt, M. Meuwly, *PNAS,* 2004, *101*, 5998.

[4]   M. Meuwly, *Chem. Phys. Chem.*, 2006, *7*, 2061.
[5]   C. Rovira, B. Schulze et al., *Biophys. J.* 2001, *81*, 435.

MS01 O4

Origin and Evolution of One of the Most Ancient Rossmann Folds. <u>W. L. Duax</u>[a], R. Huether[a], Q. Mao[a], V. Pletnev[a], T. Umland[a] and C. M. Weeks[a], [a]*Hauptman-Woodward Medical Research Institute, Buffalo, NY 14203.* E-mail: duax@hwi.buffalo.edu

Keywords: Evolution, β-k-ACPR, Rossmann Fold

The Rossmann fold is one of the most ancient and commonly encountered protein folds. Members of the short-chain oxidoreductase (SCOR) enzyme family, which contains a Rossmann fold, are present in the genomes of all species sequenced to date. The SCOR family is part of a much larger superfamily of Rossmann folds that use NAD(H) or NADP(H) as cofactors. Crystallographic and biochemical studies have revealed that, in the vast majority of Rossmann folds, NAD or NADP cofactor preference is dictated by the presence of an Asp residue in a specific sequence position in the $\beta_2\alpha_3$ turn or an Arg in the adjacent sequence position, respectively. NAD binding is usually associated with enzyme preference for oxidation, and NADP binding is associated with a reductase preference.

The β-ketoacyl [acyl carrier protein] reductase (β-k-ACPR) enzymes, a 690-member subset of the SCOR family, are essential to fatty acid synthesis in bacteria and plants. By focusing analysis on the sequence and structure of the 690 β-k-ACPRs in the gene bank, we have discovered that (1) the most primitive member of the family was an NADP reductase, (2) that NADP binding was originally contingent upon a Ser or Thr residue in the $\beta_2\alpha_3$ turn (not an Arg), (3) that a specific dimer assembly is stabilized by the stacking of aromatic groups at specific sites on the α_5 and α_6 helices and (4) that a previously undetected GGMYM sequence at the C-terminus, conserved in all species of γ-proteobacteria and most species of β-proteobacteria, stabilizes the functionally required tetramer by multiple hydrogen bonding and aromatic ring stacking crosslinking the four monomers together. Our analysis indicates that the primordial members of the β-k-ACPR family probably arose in the GC-rich γ-proteobacteria and that they are distinguished by the presence of multiple open reading frames (MORFs), an extreme codon bias in their DNA and an amino acid bias in their protein composition. The β-k-ACPRs in α- and β-proteobacteria resemble the γ-proteobacteria in having a high degree of conservation of the 40 residues characteristic of the SCOR folds, the 9 residues that are specific to β-k-ACPRs and the conserved residues that are critical to dimer and tetramer assemblies. All β-proteobacteria for which genomes have been reported are GC rich and have MORF's and GC codon bias. However, α-proteobacteria are AT-rich, do not have MORFs, do not exhibit a significant codon bias and their amino acid composition is more divergent. Further analysis should make it possible to determine if β-k-ACPR genes in α-, β-, and γ-proteobacteria evolved from a common ancestor, if γ-proteobacteria are the ancestor of the others, or if the β-k-ACPR gene was horizontally transferred from GC-rich proteobacteria to AT-rich ones. We can now characterize fully and unambiguously the β-k-ACPRs in the gene bank according to cofactor preference and mechanism of recognition, catalytic residues, residues that directly or indirectly form the proton wire via bound waters, residues controlling β-face hydride transfer, and substrate-defining residues. We can also describe, in detail, the stereochemistry of the tetrameric form required for activity and the residues that control tetramer formation as well as any conserved residues on the surface of the tetramer that may be related to specific protein/protein interactions of the β-k-ACPR. We can identify conserved sequence differences between β-k-ACPRs in major classes of bacteria that may make it possible to design inhibitors that can selectively block specific families of bacteria without interfering with mammalian biochemistry. We will use similar techniques to characterize all 10,000 SCOR genes in the databank and determine which residues specifically bind each of over 150 probable substrates.

MS01 O5

Importance of local model quality in Molecular Replacement method <u>Marcin Pawlowski</u>, Janusz M. Bujnicki, *Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, Poland.* Email:marcinp@genesilico.pl

Keywords: molecular replacement, homology modeling, local model quality

Computational models of protein structure have been shown to be useful as search models in Molecular Replacement (MR), a common method to experimentally determine protein structures by X-ray crystallography. It was shown that the success of MR depends on the high accuracy of the models, a parameter that remains unknown as long as the final structure is not available. Interestingly, during the last two years several methods (termed MQAPs) were developed to predict the local accuracy of theoretical models. We were interested in analyzing whether the application of MQAP can improve the utility of theoretical models in MR.

We focused our analyses on four known protein structures with resolution better than 2 Å, for which we generated theoretical models based on 'ideal' sequence alignments to homologous proteins obtained from the SCOP database. After superposition of models onto the corresponding native target structures we measured the local deviation between each atom in the model and its counterpart in the native structure. We also used MQAPs to 'predict' this value for models, using methods that do not take into account any information about the true difference between the model and the real structure. The known or predicted deviation was then used as the B-factor in MR calculations with experimental structure factor of the target protein. We found that the known or accurately predicted deviation of individual atoms in the search model can have significant impact on success of MR. In particular, we demonstrate that a MQAP-evaluated theoretical model that is relatively diverged from the real structure (GDT-TS score 69.1) can be used to obtained a correct MR solution, while models with unknown /non-estimated local quality typically have to exhibit close similarity (GDT-TS score > 80) to generate a comparable solution. These results indicate that theoretical modeling in combination with accurate prediction of quality of models can provide useful search models for crystallographic structure solution by MR.