



Joint Departmental / ISMB Retreat 19-20 June 2007

Posters abstracts

#1

Biophysical studies of Klentaq DNA Polymerase

William Allen

The ability of DNA polymerases to select correct nucleotide for incorporation is central to fidelity in DNA replication and repair, and thus to genome stability. Crystal structures of DNA polymerase trapped at different stages of the incorporation cycle show a large inwards movement of the nucleotide binding (fingers) subdomain upon binding of correct nucleotide. This movement is thought to be crucial to nucleotide selection, as a complete active site, with steric constraints excluding incorrect base pairs, is only present in the closed form. However the kinetics of this fingers closure have remained obscure to study at the ensemble level – studying the average behaviour of many molecules. The aim of this project was to design a FRET system capable of monitoring movement of the fingers subdomain at single molecule level, and ultimately to investigate its role in nucleotide selection.

#2

Drug Development for Huntington's Disease

Dr Simon Armitage

Huntington's disease is a rare but devastating inherited neurological disorder affecting around 8 people per 100,000. The typical onset of the disease is around the ages of 30-50 years old and the symptoms include chorea, a lack of coordination, cognitive dysfunction, and behavioural changes. Huntington's disease is caused by a mutation that causes an expansion of a polyglutamine repeat in the protein huntingtin. This leads to the formation of insoluble aggregates of the N terminal fragment of huntingtin that contains the polyglutamine repeat.

Elevated levels of the molecular chaperone heat shock protein 70 have been shown to reduce or dampen protein aggregation in polyglutamine diseases such as Huntington's disease. Over expression of heat shock protein 70 can be achieved via inhibition of heat shock protein 90. However the two classic heat shock protein 90 inhibitors Radicol and Geldanamycin have proven to be unable to cross the blood-brain barrier. Many heat shock protein 90 inhibitors have been investigated and reported as potential chemotherapeutic agents. Our project involves looking for a heat shock protein 90 inhibitor which can cross the blood-brain barrier and inhibit the formation of huntingtin aggregates. In this poster we report the initial work of this project, the synthesis of some known heat shock protein 90 inhibitors, which will next be tested for brain penetration and heat shock protein induction in vivo.

#3

Poster title and Abstract not available online

Mrs Chandrakala Basavannacharya

#4

***Cis* and *trans* regulation of hepcidin expression by Upstream Stimulatory Factor**

Henry K. Bayele¹, Harry McArdle², and Surjit Kaila S. Srail¹

¹Department of Biochemistry & Molecular Biology, University College London, London NW3 2PF, United Kingdom; ²The Rowett Research Institute, Greenburn Road, Bucksburn, Aberdeen AB21 9SB, Scotland, United Kingdom.

Hepcidin is the presumed negative regulator of systemic iron levels; its expression is induced in iron-overload, infection, inflammation and by cytokines but is suppressed in hypoxia and anaemia. Although the gene is exquisitely sensitive to changes in iron status in vivo, its mRNA is devoid of prototypical iron-response elements and it is therefore not obvious how it may be regulated by iron flux. The multiplicity of effectors of its expression also suggests that the transcriptional circuitry controlling the gene may be very complex indeed. We show here that members of the basic helix-loop-helix leucine zipper (bHLH-ZIP) family of transcriptional regulators control hepcidin expression. The Upstream Stimulatory Factor USF2, previously linked to hepcidin through gene ablation in inbred mice, exerts a polar or cis-acting effect while USF1 acts in trans to control hepcidin expression. In addition, c-Myc and Max synergize to control the expression of this hormone, supporting previous findings for the role of this couple in regulating iron metabolism. Transcriptional activation by both USF1/USF2 and c-Myc/Max heterodimers occurs through E-boxes within the promoter. Site-directed mutagenesis of these elements

rendered the promoter unresponsive to USF1/USF2 or c-Myc/Max. Dominant-negative USF1 and USF2 mutants reciprocally attenuated promoter transactivation by the cognate wild-type isoforms. Promoter occupancy by the transcription factors was confirmed by mobility shift and chromatin immunoprecipitation assays. Taken together, these members of the bHLH-ZIP family of transcriptional regulators may synergistically subserve an important role in iron metabolism as well as other pathways in which hepcidin may be involved.

#5

Molecular Mechanisms Specifying Cell Identity in the Regenerating Adult Urodele Limb

Robert Blassberg

Urodele amphibians such as the newt are able to regenerate a variety of missing body structures, one of the most anatomically complex being the limb. When lost, sections are replaced such as to regenerate all and only those missing in order to reconstruct the limb. How it is that the cells remaining in the amputated stump of the limb encode positional information able to direct growth of appropriate structures is poorly understood, however one candidate molecule exists which may prove central to the dissection of the processes involved.

Prod1 is a molecule which is found expressed on the surface of limb cells in a gradient, being highest proximally, and when overexpressed in blastemal cells is able to re-specify cells to more proximal identities. Prod1 is anchored to the membrane by a Guanosyl-Phosphatidyl-Inositol anchor and as such is not directly linked to the cytoplasm making the mode by which it signals unclear.

Prod1 is a member of the Ly-6 family of proteins, of which a number associate with the Epidermal Growth Factor Receptor. On this premise an investigation has been undertaken to determine whether Prod1 also associates with the EGFR and furthermore, whether it signals via the EGFR phosphorylation cascade in order to mediate cell identity.

#6

Poster title and Abstract not available online

Dr Katherine Bowers

#7

Small Molecule Inhibitors of Bacterial Adhesive Pili

Floris Buelens

Adhesive pili are key components of many bacterial virulence systems. In our work we are developing inhibitors of the Type 1 and P pilus systems, which have been identified as key virulence factors for the pathogenesis of urinary tract and kidney infections, from two different angles. Our first approach seeks to characterise and develop a series of lead compounds which interfere with the assembly at the bacterial outer membrane of the pilus organelle. Secondly, we are looking for novel classes of small molecules to target the adhesive receptor domains of the pilus tip.

#8

Nerve-dependence and positional identity in amphibian limb regeneration

James Godwin, Phillip Gates, Nooreen Shaikh, Anoop Kumar, Azara Janmohammed, Jean-Paul Delgado, Robert Blassberg and Jeremy Brockes

Salamanders are able to regenerate their limbs after injury or limb loss. The regeneration of the limb proceeds by formation of a blastema, a mound of mesenchymal stem cells which accumulates under the wound epidermis at the end of the stump. The limb blastema is an autonomous structure and the mesenchymal stem cells are imprinted with the positional cues to restore the limb. We identified a cell surface protein called Prod 1 which is expressed in the adult newt limb in an exponential gradient and is implicated in positional identity. Recently we identified a secretory protein, newt Anterior Gradient (nAG), as potential ligand of Prod 1 and the candidate protein apparently mediates the nerve dependence of limb regeneration. Molecular, structural and functional characterisation of these two proteins are in progress.

#9

Analysing the Oral Metagenome Using Phage Display

Samantha Easton and John Ward

Department of Biochemistry and Molecular Biology, University College London, Gower Street, London

Microbial infections are dependent on a series of changes in both host and bacterium, in response to bacterial attachment. Commensal bacteria also develop and maintain intimate relationships with mammalian cells, without triggering invasion mechanisms.

In the human mouth, millions of bacteria populate every surface and crevice. Oral mucosal surfaces present micro-organisms with numerous habitats in which to colonize, feed, multiply and interact with other bacteria. The provision of warmth and water, and the ubiquitous presence of nutrient-rich saliva on oral surfaces mean that the mouth is readily colonized by commensals and pathogens which maintain an equilibrium that occasionally favors the progression of disease.

Bacterial cell-surface adhesins enable recognition and binding of specific receptors on human cells (1), and are responsible for the establishment of close relationships between microbe and host. The primary objective of this research project is to locate examples of these bacterial adhesins from micro-organisms previously bound to the human tongue. This has so far required the employment of a metagenomic strategy (2) to process a large volume of DNA, and the investigation of this DNA using a molecular tool called Phage Display.

Phage display is a simple genetic technique for the identification of protein-ligand interactions, such as microbial attachment to mammalian cells, and is currently used successfully in epitope mapping, antibody tailoring and receptor agonist and antagonist screening. It is well established that the adhesion of enteric, oral and respiratory bacteria is required for colonization and, once bound, the bacteria are less likely to succumb to host defences. This study aims to identify bacterial adhesins which are vital for bacterial colonisation of the human tongue, using Phage Display. Screening of the Phage Display library will be carried out using a technique called Bio-panning, against ligands human IgA and fibronectin, and is expected to yield one or more bacterial adhesins.

Results are based on tongue scrapings taken from 9 healthy volunteers, who have no record of antibiotic use for 6 months previous to sampling.

1. Azzazy, H.M. & Highsmith, W.E., Jr. (2002) Phage display technology: clinical applications and recent innovations. *Clin. Biochem.* 35, 425-445.
2. Ward, N. (2006) New directions and interactions in metagenomics research. *FEMS Microbiol. Ecol.* 55, 331-338.

#10

Developing Scientific Workflows to support Biomedical Applications

Nigel Ferguson

Within the biomedical sciences, advanced molecular biological techniques (e.g. high throughput proteomics, etc.) have enabled many elements of biological systems to be characterised at a rapid pace. Much of this data is now stored in various databases accessible via the Internet. This expanse of data has also resulted in the creation of many software prediction and modelling tools.

This "data landscape" presents a number of computational challenges. The tools and data are spread over abundant sites, new sites are frequently introduced, there are no common interfaces and the underlying data is constantly changing.

Often, a series of these tools and databases are linked together to form a computational pipeline. Until recently, the linkage of such tools would be achieved by joining these components via scripting languages such as Perl. Although many projects have been successfully based on this simple approach, it has many drawbacks, including: slow development time, underlying tools not architected for distributed environments, the need for a high degree of programming knowledge, the fact that complex applications are prohibitively difficult to construct and the issues associated with extensive component reuse.

Recently, the approach to creating these computational pipelines has been radically changed using graphical environments known as Workflow Management Systems (WMS).

This environment enables a user to connect graphically a number of components (e.g. prediction tools) together to solve a problem that each individual application could not solve. Currently, these workflows are limited to relatively straightforward applications. The WMS environment is ideally suited to dynamic

applications as new components can be added with ease and pre-existing components can be readily reutilised.

The project aims to demonstrate how the dynamic qualities of a WMS can be used on complex biological systems. An example of such a system is the prediction of epitopes within the adaptive immune response.

#11

Reasoning with Uncertainty in Biological Databases using Probabilistic Relational Models

Mr Paul Fletcher

Abstract not available

#12

A cryo-EM study of doublecortin stabilised microtubules

Franck Fourniol

Microtubules are crucial for neuronal migration and differentiation occurring during brain development. Doublecortin is the best-described member of a recently discovered family of microtubule-associated proteins. It participates in organisation of the microtubule cytoskeleton in migrating and differentiating neurons and as a matter of fact patients carrying mutations in doublecortin suffer from mental retardation and epilepsy. Doublecortin was previously shown to stabilise and nucleate specifically 13-prot filament microtubules, but the basis for its specificity is still poorly understood. We set out to determine a high-resolution structure of doublecortin bound to 13-prot filament microtubules using cryo-electron microscopy and image reconstruction. Here we report high-resolution cryo images of microtubules bound with doublecortin, and microtubules doubly bound with doublecortin and kinesin motor domain. These data are to be processed through a single particle approach, the early steps of which are described here. This will yield great insight into the precise way in which doublecortin regulates the microtubule cytoskeleton, a role that cannot be compensated for by other microtubule-associated proteins.

#13

Flavin-containing Monooxygenases: Genetic Polymorphisms & Drug Metabolism

Asvi A. Francois and Elizabeth A. Shephard

Flavin-containing monooxygenase (FMO) proteins play an important role in the metabolism of foreign chemicals including therapeutic drugs. Consequently polymorphic variation in FMO genes can lead to changes in drug metabolism, which in turn can cause adverse drug reactions or lack of drug efficacy. Five members of the FMO family of proteins, FMOs 1, 2, 3, 4 and 5 are known to metabolise drugs. Nearly one-third of the world's population is infected with *Mycobacterium tuberculosis* with pulmonary tuberculosis being the most common form of the disease. Tuberculosis is especially common in the world's poorer areas including Africa. This talk will concentrate on the FMO2 protein and how genetic variation in the FMO2 gene influences the metabolism of an important class of drugs used to treat tuberculosis. The FMO2 protein is of particular interest because in the FMO2 gene of Asians and Caucasians there is a nonsense mutation that causes premature truncation of translation. Thus these individuals do not make FMO2 protein. In contrast, 25% of the population of sub-Saharan Africa have a wild type FMO2 gene and thus produce a catalytically active protein. Results will be presented to show that thiacetazone, a cheap and commonly prescribed anti-tuberculosis drug in Africa, is a substrate for FMO2. The consequences of this finding for drug therapy and/or potential adverse reactions will be discussed.

#14

Structural Studies of Mycobacterial Small Heat Shock Proteins

Gemma Gargent

Small heat shock proteins (sHSPs) are a family of chaperones that all contain a conserved C-terminal sequence called the α -crystallin domain. Sequence alignments of the sHSPs found in mycobacterial species have revealed that the proteins are arranged into 3 classes; named acr1, acr2 and acr3.

Previous work on the two *M. tuberculosis* sHSPs has shown that the *acr1* is a dodecamer (contrary to earlier literature reports of a nonomer) and the *acr2* is polydisperse with oligomers between 16 and 28 subunits. A negative stain EM reconstruction has shown *acr1* to be a tetrahedron with a dimer on each edge.

I have cloned and expressed a number of *acr* sHSPs from mycobacterial species that are homologous to the *acr* proteins from *M. tuberculosis*.

Nanospray Mass spectroscopy* revealed that two of these proteins, *acr2* from *M. marinum* and *M. smegmatis* formed monodisperse 24 subunit oligomers, in contrast to the *M. tuberculosis* protein. *Acr3* from *M. leprae* formed a mixture of a 16-mer and a 18mer, whereas, the *acr1* proteins from *M. marinum* and *M. smegmatis* both formed a large range of oligomers.

The *acr2* protein from *M. marinum* and *M. smegmatis* were both crystallised. The *M. marinum* *acr2* crystal was very small and did not diffract further than 10 Å whereas, the *M. smegmatis* *acr2* crystal diffracted to about 7.5 Å. From this data the space group was calculated to be I432, which suggests that the protein would have octahedral symmetry.

Finally I have taken a number of micrographs on the Technai 120 kV electron microscope of both *M. marinum* and *M. smegmatis* *acr2* at a magnification of 52,000 and a range of defocus values from 600-800nm. Preliminary analysis of the 3,165 particles picked from the S25 micrographs has revealed six-fold symmetry. As images taken using an electron microscope are projections of 3D density the six-fold axis may well be two 3-fold axes superimposed on one another; which would be compatible with the space group calculated from the crystallography dataset.

#15

Structural and Functional Analysis of CIN85 and its Role in Receptor Endocytosis

Ragini Ghosh (Supervisor: John.E Ladbury)

Department of Biochemistry and Molecular Biology, University College London, UK

Endocytosis of receptor tyrosine kinases (RTKs) as epithelial growth factor receptor (EGFR) and c-Met/hepatocyte growth factor receptor is mediated by their ubiquitination by an intracellular enzyme ubiquitin ligase, Cbl. This is followed by the recruitment of other principal adaptor proteins such as **CIN85** and endophilins to form a **Cbl-CIN85-endophilin** complex which in turn regulates the endocytosis, lysosomal degradation and endosomal sorting of these RTKs.

Recent evidence shows that inhibition of the Cbl-CIN85-endophilin interaction was sufficient to block EGFR endocytosis and degradation leading to uncontrolled signalling and tumorigenesis. The present work focuses on the protein CIN85. It acts as a scaffold protein and is able to recruit signalling proteins into RTKs-associated complexes that are critical for RTKs endocytosis. Other than interacting with Cbl, CIN85 is found to non-covalently interact with a number of proline rich proteins/peptides as well as ubiquitin in the degradation pathway. The relatively low affinity of CIN85 for binding proteins could allow for rapid exchange of proteins/peptides, depending on their local concentration, cellular compartmentalization, or posttranslational modifications in response to changes in cell signal. Hence, our interest lies in ultimately determining the significance and contribution of these CIN85-binding protein interactions in the signalling cascade and its relation to tumorigenesis/cancer.

Different CIN85-C constructs have been cloned and expressed and its interaction with ubiquitin has presently been validated by GST-pull down assays, isothermal titration calorimetry and NMR experiments. Future work pertains to investigating the interaction of CIN85 with other interacting proteins/peptides and finally determine the structure of the complexes.

#16

Identification and Functional Characterisation of S6 Kinase Nuclear Associated Partners

Eddy Goh

S6 Kinase (S6K) belongs to the AGC family of serine/threonine kinases. It is highly implicated in cancer development and is found to be over expressed in cancer. Recent studies from our laboratory indicated that two isoforms of S6K (S6K1 and S6K2) are involved in controlling gene expression at the level of transcription, in which the mechanisms of this process is still unknown. By establishing a system of large scale nuclear fractionation from S6K expressing stable cell lines, we have shown that a number of nuclear proteins co-immunoprecipitate with nuclear S6K2 but not with S6K1, indicating that S6K2 has distinct nuclear roles from that of S6K1. We have also shown that these S6K2 associated nuclear proteins act as potential S6K2 phosphorylation substrates in an *in vitro* kinase assay.

#17

Towards the understanding of transcription initiation: Insights into the function of transcription factor E

Dr Dina Grohmann

The transcription cycle can be divided into three phases: initiation, elongation and termination. During each phase different regulatory proteins interact with RNA polymerase (RNAP) and thereby control gene expression. The archaeal transcription machinery is closely related to eukaryotic RNAPII-system and it is a highly attractive model system because of its superior biochemical tractability.

Initiation of transcription in the archaeal system involves recognition of promoter sequences by TBP (TATA binding protein) that is followed by the binding of TFB (transcription factor B) leading to a stabilisation of the TBP-DNA complex. The ternary DNA-TBP-TFB complex is able to recruit RNAP resulting in transcription initiation. A third additional transcription factor E (TFE) completes the initiation complex. TFE is involved in but not strictly required for transcription initiation *in vitro*. The modus operandi of TBP and TFB have been characterised in great detail - but our understanding of TFE action is disappointingly lacking. In archaeal extracts TFE has been shown to interact with both TBP and RNAP. TFE stimulates transcription moderately, most likely by stabilising the initiation complex and promoting open complex formation. TFE consists of two structurally discrete domains, a Winged Helix and a Zn-ribbon domain complemented by two flexible regions, the linker and tail. In this study we focus on the molecular mechanisms of *Methanocaldococcus jannaschii* TFE by applying a wholly recombinant archaeal transcription system. In order to understand how the individual TFE domains are involved in the interaction network of the initiation complex and thereby influence transcription, we made use of TFE deletion variants lacking either of the domains. In summary, the integrity of the Winged Helix domain, the Zn-ribbon and the flexible linker region are required for TFE function *in vitro* whereas the C-terminal tail domain is dispensable. TFE stabilises the DNA-TBP-TFB-RNAP complex in a manner that is strictly dependent on RNAP subunits F/E and influenced by the topology of the promoter template. Our results suggest that both TFE domains cooperate to facilitate the stabilisation of the DNA-TBP-TFB-TFE-RNAP initiation complex and the closed-to-open complex transition during transcription initiation.

#18

Transcription Elongation Factor SPT4/5 modulates RNA-binding of RNAP subunits F/E (RPB4/7)

Angela Hirtreiter and Finn Werner

UCL Department of Biochemistry, Darwin Building, Gower Street, London WC1E 6BT, UK

Archaeal RNA Polymerases mirror eukaryotic RNAPII in architecture and use of basal transcription factors. The archaeal RNAP subunits F/E (RPB4/7) form a module that reversibly associates with RNAP and is involved in both initiation and elongation of transcription. During initiation F/E is thought to close the RNAP clamp domain over the DNA binding channel leading to 'open complex' formation (promoter melting). During the elongation phase F/E binds the emerging nascent RNA via an S1 motif (in E) and this interaction has the potential to regulate the processivity and elongation rate of RNAP.

Spt4/5 is a positive transcription elongation factor that is conserved in all domains of life: as heterodimeric complex in eukaryotes and archaea (Spt4/5) and as monomer in bacteria (NusG, homologous to Spt5). Both Spt4 and 5 harbour putative RNA binding motifs, however whether Spt4/5 can bind RNA and how this affects transcription is not clear. Spt4/5 reversibly associates with elongating RNAPs but the subunits and domains that are involved have not been identified yet. In *Sulfolobus acidocaldarius* Spt4 is fused to the C-terminus of RNAP subunit E suggesting a biologically relevant interaction of Spt4/5 and F/E.

The aim of this project is to characterise the interaction network between Spt4/5, RNAP subunits and the nucleic acid scaffold of the elongation complex. We have fused *M. jannaschii* E and Spt4 similar to *S. acidocaldarius*. The E-Spt4 fusion allows the formation of a stable heterotetrameric complex of RNAP-F/E and Spt4/5. Fusion of Spt4 to E marginally decreases the RNA-binding activity of F/E, but the incorporation of Spt5 into the trimeric F/E-Spt4 complex leads to a dramatic stimulation of RNA-binding. Our results show that the N-terminal domain of Spt5 is necessary for the interaction with Spt4. We are discussing the implications of the interaction between F/E and Spt4/5 for the regulation of transcription elongation.

#19

Kinesins in Malaria

Christina Hoey

Kinesins form a large super family of motor proteins ubiquitous across a number of eukaryotic species including mammals and yeast. They utilise energy derived from ATP hydrolysis and interact with microtubules to fulfill their function. Defined by their motor domain, which contains the microtubule binding site, and ATP binding site, a number of kinesins have been identified, and found to perform different roles within the cell. This includes kinesin 5, which is vital for chromosome separation during mitosis and kinesin 13, which depolymerises microtubules. In comparison, very little is known about kinesins in the most virulent of malarial parasites *Plasmodium falciparum*.

The goal of this research project is to try and characterize a number of kinesins from *Plasmodium falciparum* in the hope that this will reveal novel behaviours, and unique structures that are parasite specific, thereby allowing it to be an interesting new drug target. Sequence alignments already show that there are significant differences between malarial and other eukaryotic kinesins, one key difference being that the malarial proteins appear to have large asparagine rich amino acid inserts, whose function is yet to be elucidated.

Initial attempts have been made to express and purify recombinant protein of the kinesin-5 and kinesin-13 motor domains from *Plasmodium falciparum*, and the asparagine rich insert from the kinesin-5 motor domain. Once an expression and purification protocol has been established for the production of these kinesins, biochemical characterization can take place using a variety of standard *in vitro* assays and techniques. Cryo-electron microscopy will also be performed to try and determine a high resolution structure of kinesins bound to tubulin, enabling the visualisation of the interaction between the kinesins and microtubules directly.

#20

Abysis, an antibody database

Jacob Hurst

The Abysis database links all publicly available Ig protein sequence data together. This involves three main data sources kabat, imgt and the pdb. The sequences are then numbered using the kabat numbering scheme, via a program developed by Abhinandan Raghavan and Andrew Martin. This numbering allows antibody sequences to be directly compared with each other. The three data sources have been combined into a common XML format and loaded into a postgres database.

As, well as providing numbering and other annotation, missing information such as the correct heavy and light chain pairing has been mined from available data.

The next step is to provide a facility for structural comparison of known Ig structures, with the standard numbering scheme. All of this data will be publicly available via the Internet.

#21

Structural studies of two novel peroxidases from the protozoan parasite Trypanosoma cruzi

Syed Hussain, Shreenal Patel, Richard Harris and Snezana Djordjevic
Dept. of Biochemistry and Molecular Biology, University College London

The protozoan parasite *Trypanosoma cruzi* is the causative agent of Chagas' disease. In Latin America, 16-18 million people are infected by the parasite creating a major health problem. The disease is fatal in 15-30% of those infected while others are inflicted by serious disabilities. Drugs currently used in the treatment of Chagas' disease often have toxic side effects and fail to clear parasitaemia and their mode of action is unknown. One of these drugs, nifurtimox, undergoes redox cycling with the parasite.

Reactive oxygen species (ROS), such as hydroperoxides, are the unwanted by-products of aerobic metabolism. To protect cells against their potentially lethal effects a series of pathways have evolved that are collectively called the oxidative defence system. Uniquely, trypanosomatidae contain several enzyme-mediated pathways for the removal of hydroperoxides that are centred upon the unusual thiol trypanothione and are potential therapeutic targets.

Work presented here aims to characterise two enzymatic pathways involved in the pathway of hydroperoxide metabolism by *T. cruzi* through functional and structural studies. The enzymes

Ascorbate peroxidase (APX) and Glutathione Peroxidase (GPX) reduce hydroperoxides by utilising ascorbate and glutathione/trypanoxin as electron donors, respectively. Protein constructs have been generated for APX and GPX allowing large scale over-expression for structural studies. Here we report, for both enzymes, the purification and structural information using X-ray crystallography and NMR spectroscopy.

#22

-

#23

Evaluation and integration of protein named entity recognition tools

Renata Kabiljo

Protein named entity recognition (NER) is widely recognized as one of the most important sub-tasks in biomedical text-mining. Our goal was to properly evaluate a number of protein taggers, and to compare their performance on general biomedical abstracts with that on full text articles from immunological domain. To facilitate this evaluation, we have produced two new protein-specific corpora: ProSpecTome and ImmunoTome. ProSpecTome contains 234 abstracts randomly chosen from the JNLPBA evaluation corpus, while ImmunoTome contains 10 full text articles from the Journal of Immunology. To our knowledge, ImmunoTome is the first corpus of full text articles in this domain. Both ProSpecTome and ImmunoTome explicitly annotate names of proteins, and the annotation guidelines used to produce both corpora together with the degree of inter-annotator agreement associated with their production are explicitly documented. In ProSpecTome, general references to proteins are annotated separately from the names of individual proteins and protein families, while ImmunoTome differentiates between protein names that refer to protein objects and these that refer to other entities. All evaluated protein taggers perform significantly worse on full texts (ImmunoTome) than on abstracts (ProSpecTome). This drop in performance is due to a different information content in abstracts and full texts, and not to specific nomenclature of proteins in the ImmunoTome corpus. We show that combining different protein taggers in a Bayesian network framework can improve overall tagging performance by 4%. We are offering an on-line service that tags submitted full text articles for protein names, offering degree of confidence for each annotation.

#24

Identification of Key Residues for Assembly of a Surface Layer Protein

H. Kinns, S. Howorka.

Department of Chemistry, UCL, 20 Gordon Street, London. WC1H 0AJ

Crystalline surface layer (S-layers) form one of the most common boundary layers of many eubacteria and archaea and are composed of self-assembling single protein or glycoprotein species. Although the ultrastructure of S-layers has been well characterised with electron and atomic force microscopy, little is known about the position of individual amino acids in the protein, and no X-ray crystal structure is available. We investigated the S-layer protein SbsB from *Geobacillus stearothermophilus* PV72/p2 to identify key residues for assembly. Twenty-three single cysteine mutants, located on the surface of the SbsB monomer, were subjected to a cross-linking screen using a sulfhydryl-reactive photoactivatable reagent. Gel electrophoretic analysis on the formation of cross-linked dimers indicated that four out of 23 mutants were located at the interface. In combination with chemical surface accessibility data, four other residues were mapped to the outer surface and three residues to the inner peptidoglycan-bound surface of the lattice. The HA epitope tag was inserted at each of the surface positions to confirm the residues position via an assembly inhibition screen as assayed by negative staining and electron microscopy. Identification of interfacial residues will enable the rational design of mutated SbsB protein for X-ray crystallization studies. Furthermore, engineered proteins carrying an insert at the outer or inner surface will provide a uniform epitope display with applications for scanning probe techniques.

#25

Solution structure of the complex between complement C3d and full length complement receptor Type 2

Keying Li¹, Jayesh Gor¹, Jonathan P. Hannan², Rengasamy Asokan², V. Michael Holers² and Stephen J. Perkins¹

¹ Department of Biochemistry and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, U. K.

² Division of Rheumatology, University of Colorado Health Sciences Center, 1775 N. Ursula Street, Room 3102, PO Box 6511, Mail Stop B115, Aurora, CO 80045, U.S.A.

Complement receptor type 2 (CR2, CD21) is a cell surface protein that links the innate and adaptive immune response during the activation of B cells. The extracellular portion of CR2 comprises 15 or 16 short complement regulator (SCR) domains. Its overall domain structure in solution was recently determined by constrained X-ray scattering and analytical ultracentrifugation (AUC) modeling to be a partially folded back but flexible structure (Gilbert et al. (2006) *J. Mol. Biol.* **362**, 1132-1147). To examine the effects of ligand binding to this structure, we have expressed C3d in *E. coli* and CR2 in baculovirus and are studying their complex by both X-ray scattering data and analytical ultracentrifugation (AUC). The X-ray radius of gyration of free CR2 was determined to be 11.5 ± 0.9 nm, and that of its cross-section was 1.8 ± 0.1 nm. An X-ray data set for the 1:1 complex between CR2 and C3d was collected at ESRF. This showed evidence for complex formation. Analytical ultracentrifugation experiments to verify these results are in progress.

#26

Purification and Characterisation of Recombinant Wild Type FGFR2

Chi-Chuan Lin

Apert syndrome, a serious genetic disease resulting in cranial synostosis and syndactyly during foetal development, is caused by point mutations (S252W and P253R) in the region linking Ig-like domains II and III of Fibroblast growth factor receptor 2 (FGFR2). Fibroblast growth factors (FGFs), the ligands of FGFR2, mediate their biological functions by binding to the Ig-like domains of FGFR2, resulting in the dimerisation and phosphorylation of multiple tyrosine residues on the cytoplasmic domain of the receptors. It is now known that the binding of FGF requires the presence of Heparin/Heparan sulfate proteoglycans (HSPGs), which are known to play an important role in the control of FGF receptor signalling by controlling receptor dimerisation. However, the role of differential heparin structures in controlling FGFR2 signalling remains unclear. This study is focused on the heparin/HSPG structures in controlling FGFR2 signalling. The wild type recombinant receptor proteins were expressed in mammalian cells (293T cells); functional studies carried out in 293T cells have shown that the recombinant receptor proteins (wild type) have the ability to be phosphorylated in the stimulation of FGF *in vivo*. Future *in vitro* work is being done in order to show that the recombinant wild type receptor proteins can bind to FGF in the presence of heparin, and this binding can induce the phosphorylation of receptor proteins. Future studies will be focused on the comparison of binding abilities and signalling responses between wild type and mutant FGFR2 using different structures of heparin/HSPG.

#27

Improving Function Prediction Using Patterns of Native Disorder in Proteins

Anna Lobley

Intrinsically unstructured (disordered) proteins adopt little or no stable secondary structure in their native state. Proteins containing long disordered regions are abundant within eukaryotic genomes and can be predicted successfully from amino sequence. Disordered regions have been shown to be important for functional specificity and frequently contain binding motifs or are located at sites of covalent modification often inferring a regulatory role for the protein. Computational methods that predict protein function from sequence rely upon the use of homology information to transfer annotations between proteins. These methods are not applicable to orphan proteins or cases where whole families of protein sequences are not annotated. To address the requirement for protein function prediction methods that are independent of sequence homology and explore the use of information describing protein disorder, we have implemented a machine learning method for predicting protein function from sequence. A set of features for encoding disorder information was designed and their importance in predicting Gene Ontology (GO) categories demonstrated. The addition of disorder features significantly improved prediction of many GO categories. The method has been benchmarked

against a competing method and the practical use of the classifiers demonstrated through the annotation of a set of orphan and unknown human proteins.

#28

Developments of Combinatorial Domain Hunting using Hsp90

Kate Maclagan¹, Loretto H. Puckey², James Reid², Richard Harris¹, Chris Prodromou^{3,4}, Laurence H. Pearl^{3,4}, Renos Savva^{2,4}, and Paul C. Driscoll^{2,4}

*1*Department of Biochemistry and Molecular Biology, University College London, London, UK, *2*School of Crystallography, Birkbeck College, London, UK, *3*Section of Structural Biology, Institute of Cancer Research, London, UK, *4*Domainex Ltd., London, UK

Producing multi-milligram quantities of full-length proteins is often extremely challenging. Instead, researchers routinely express sub-constructs, or domains, of the target protein. Defining the precise domain boundaries is frequently a labour-intensive trial-and-error process. We have developed 'Combinatorial Domain Hunting' (CDH), a methodology designed to identify highly expressed, soluble, recombinant protein constructs from a given DNA template. CDH combines the use of a finely sampled, unbiased fragmented gene library with an expression screen which enriches the output clones for highly expressed, soluble proteins and hence in principle requires no prior knowledge of the encoded protein domain structure. CDH has been applied to the Hsp90 gene, giving rise to a number of soluble protein fragments. Assessment of the outputs of the screen requires biophysical examination of the foldedness of each CDH 'hit', for which we have adopted Thermofluor®, a thermal denaturation, monitored by extrinsic fluorescence. The poster will describe experimental data generated from applying CDH to Hsp90, a description of how we assess 'hits' and current developments to the technology being investigated.

#29

-

#30

Investigation of S-nitrosothiol formation in complex I and model compounds by FTIR spectroscopy

Dr Doug Marshall

In addition to its role as a regulator of vascular tone, synaptic signalling and cellular defence, nitric oxide (NO) has been shown to inhibit complex I and cytochrome *c* oxidase of the respiratory electron transfer chain. It has been suggested that the mechanism by which NO inhibits complex I involves the formation of S-nitrosothiols on the complex which are deleterious to its function (1;2).

In the present study Fourier transform infrared (FTIR) spectra of glutathione and S-nitrosated glutathione were compared and a putative S-N-O-associated signal was identified. Subtraction of the FTIR spectrum of bovine serum albumin in solution from an equivalent spectrum of bovine serum albumin in which Cys-34 had been S-nitrosated yielded a difference spectrum which contained a signal of comparable amplitude and position. The putative S-N-O-associated signal was at 1044 cm⁻¹ with an extinction coefficient of 166 M⁻¹.cm⁻¹ (calculated from the glutathione data). Preliminary attempts to form S-nitrosothiols on a rehydrated layer of complex I using S-nitroso-N-acetylpenicillamine while simultaneously taking FTIR measurements did not yield spectra of adequate signal:noise for detection of a signal of the expected size. In order to increase signal:noise averaging of multiple experiments will be performed.

Protein-S-nitrosothiol formation can also be detected by UV/visible spectroscopy. UV/visible S-nitrosated minus non-S-nitrosated difference spectra of cysteine, bovine serum albumin and glutathione are presented and show a clear S-N-O-associated signal at ~550 nm, consistent with previously published circular dichroism measurements (3). For cysteine, the extinction coefficient of this signal was calculated to be 0.016 mM⁻¹.cm⁻¹.

Experiments are underway to cycle a rehydrated layer of complex I between S-nitrosated and non-S-nitrosated states reversibly while simultaneously recording both UV/visible and FTIR difference spectra. The data obtained from these experiments, in combination with the data presented here, will allow

quantitation of complex I-S-nitrosation. Further experiments, utilising various S-nitrosating agents, will allow the mechanism of complex I-S-nitrosation to be investigated.

References

1. Braiman, M. S., Mogi, T., Marti, T., Stern, L. J., Khorana, H. G., and Rothschild, K. J. (1988) Vibrational spectroscopy of bacteriorhodopsin mutants: Light-driven proton transport involves protonation changes of aspartic acid residues 85, 96, and 212, *Biochemistry* 27, 8516-8520.
2. Dahm, C. C., Moore, K., and Murphy, M. P. (2006) Persistent S-nitrosation of complex I and other mitochondrial membrane proteins by S-nitrosothiols but not nitric oxide or peroxyxynitrite, *J. Biol. Chem.* 281, 10056-10065.
3. Mohny, B. K. and Walker, G. C. (1997) Conformational restriction of cysteine-bound NO in bovine serum albumin revealed by circular dichroism, *J. Am. Chem. Soc.* 119, 9311-9312.

#31

Scoring Conservation in Homo-orthologues

Ms Lisa McMillan

In order to understand and predict the effects of mutations to a protein we wish to identify residues which are highly conserved across species. Such residues are likely to contribute to protein function. However, to define 'high' conservation in a multiple sequence alignment, it is necessary to accommodate biases that are inherent in the population of sequences (from population size and spread across different species) and in the pattern of conservation (some proteins are highly conserved across a wide range of species). Selecting a suitable threshold for 'high' conservation is straightforward (if somewhat arbitrary) for an individual protein, but to generate thresholds automatically for thousands of proteins is difficult. We have developed a conservation scoring system called 'ImPACT' (Improved Protein Alignment Conservation Threshold), which automatically generates a threshold for 'high' conservation for any given protein alignment, by statistically modelling the conservation patterns observed. In addition, we have developed a new method, 'orthofind' which scans and compares UniProtKB/SwissProt annotations to generate sets of functional orthologues (i.e. homologous proteins from different species with an equivalent function) which can be aligned to calculate conservation.

#32

Title: Evidence for a monomer-dimer equilibrium in native human Factor H

Ruodan Nan¹, Jayesh Gor¹, Robert B. Sim² and Stephen J. Perkins¹

¹ Department of Biochemistry and Molecular Biology, Darwin Building, University College London, Gower Street, London, U. K.
² MRC Immunochemistry Unit, Department of Biochemistry, South Parks Road, University of Oxford, Oxford OX1 3QU, U. K.

Factor H (FH) plays a central role in the regulation of complement activation via the alternative pathway. It is composed of 20 short complement regulator (SCR) domains. Previously, we had demonstrated that Factor H is dimeric in solution by X-ray and neutron scattering (Perkins et al. (1991) *Biochemistry*, 30, 2847-2857), while the follow-up study demonstrated that Factor H is monomeric (Aslam & Perkins (2001) *J. Mol. Biol.* 309, 1117-1138). Recently the SCR-6/8 and SCR-16/20 fragments of Factor H have each revealed monomer-dimer self-association with KD values of 40 mM and 16 mM respectively. We have reinvestigated the discrepancy in the earlier studies by the use of c(s) size-distribution analyses of Factor H in sedimentation velocity experiments on the analytical ultracentrifuge. Between 0.1 mg/ml to 0.6 mg/ml of Factor H in either phosphate buffer saline or Hepes buffer, two peaks were observed in the c(s) analyses between 25k r.p.m. and 50k r.p.m. rotor speeds. The major peak at 5.7 ± 0.1 S corresponds to a molecular weight at 142 ± 2 kDa, while the second peak at 9.2 ± 0.5 S corresponds to a molecular weight at 285 ± 24 kDa. The relative intensity of the second peak increased with concentration. We conclude that Factor H exists in a monomer-dimer equilibrium in solution. A tentative explanation of the outcomes of the 1991 and 2001 studies is that these were fortuitously performed in concentration ranges that corresponded to predominantly dimeric and monomeric Factor H. Further data collection is in progress and will be presented alongside a reanalysis of our 2001 data. From the results so far obtained, we conclude that Factor H will be predominantly monomeric in plasma in its physiological concentration range between 0.2 mg/ml to 0.8 mg/ml.

#33

TIP48 and TIP49: united in mitosis or poles apart?

Andrew Niewiarowski

TIP48 and TIP49 are two homologous AAA+ ATPases and are individually essential in all eukaryotes. They have been implicated in a multitude of broad cellular processes including transcriptional regulation, oncogenesis and DNA repair. It is established that TIP48 and TIP49 are found together as components of multi-protein nuclear complexes, including several chromatin remodelling complexes; their chromatin remodelling activities may explain the variety of nuclear functions. It is still unclear whether a simple underlying mechanism of TIP48 and TIP49 links these diverse processes and how these different functions are coordinated. Moreover, it is essential to establish if TIP48 and TIP49 have independent roles and whether these roles are regulated by oligomerisation. We are currently investigating the biochemical properties and sub-cellular localisation of TIP48 and TIP49, with focus on the assembly of TIP48 and TIP49 into their active forms. Several recent reports have suggested that TIP48 and TIP49 are important factors in mitosis and cytokinesis, further increasing their broad repertoire of functions. In contrast with their nuclear activities, current data suggest TIP48 and TIP49 operate independently during mitosis. Investigating the various mitotic roles of TIP48 and TIP49 and how these are regulated is of central importance, and finding an underlying mechanism that connects the multitude of processes in which TIP48 and TIP49 are essential, remains an interesting conundrum.

#34

Solution structures for the regulatory SCR-1/5 and cell-surface-binding SCR-16/20 fragments of Factor H reveal different self-associative properties

Azubuike I. Okemefuna¹, Hannah E. Gilbert¹, Kim Griggs², Rebecca Ormsby², David L. Gordon² and Stephen J. Perkins¹

¹ Department of Biochemistry and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT, U. K.

² Department of Microbiology and Infectious Diseases, Flinders Medical Centre, Bedford Park SA 5042, Australia.

Factor H (FH) plays a central role in the regulation of complement activation via the alternative pathway. It is composed of 20 short complement regulator (SCR) domains. The SCR-1/5 region is required for decay acceleration and cofactor activity while the SCR-16/20 region possesses C3b/C3d and heparin binding sites. X-ray scattering and analytical ultracentrifugation showed that SCR-1/5 is monomeric while SCR-16/20 is monomeric and dimeric. The Guinier radius of gyration R_G of 4.4 nm for SCR-1/5 and those of 4.7 nm and 7.8 nm for monomeric and dimeric SCR-16/20 respectively showed that their structures are not extended. SCR-1/5 has a maximum dimension of 15 nm, while those for monomeric and dimeric SCR-16/20 are 17 nm and 27 nm respectively. The sedimentation coefficient of 2.4 S for SCR-1/5 showed no concentration dependence, while that for SCR-16/20 is 2.8 S for the monomer and 3.9 S for the dimer. Sedimentation equilibrium data showed a single species for SCR-1/5 and a monomer-dimer equilibrium for SCR-16/20. The concentration dependences of the Guinier parameters and equilibrium data resulted in a monomer-dimer dissociation constant of 16 μ M. The constrained scattering modelling of SCR-1/5 and SCR-16/20 showed that partially bent flexible SCR arrangements fit the data better than linear arrangements, and that the dimer can be modelled by the end-to-end association of two SCR-20 domains. It is concluded that the N-terminal and C-terminal regions of FH showed different self-associative properties. The models for SCR-1/5 and SCR-16/20 are consistent with the partially folded back structure for intact wild-type FH.

#35

Structural Studies of the inner DysF domain of Human Myoferlin

Pryank Patel

Muscular dystrophies are a group of inherited muscle diseases that cause muscle weakness and wasting. Mutations in the dysferlin gene, encoding the dysferlin protein, causes disruption to the muscle membrane repair mechanism and leads to two types of muscular dystrophies, limb-girdle muscular dystrophy type 2B (LGMD2B) and Miyoshi Myopathy. Dysferlin is part of a family of proteins known as the ferlins. Myoferlin, another ferlin family member, is required for skeletal muscle generation and is highly expressed during stages of muscle generation. This family of proteins consist of large

proteins of over 2000 amino acid residues with multiple C2 domains and a C-terminal transmembrane helix. In some of the members there is also a domain of unknown structure and function known as the DysF domain, an unusual feature of which is that there are two copies, one 'nested' inside another as a result of an internal duplication event. As a result, the domain is referred to as two parts – the N-terminus region, DysFN and the C-terminus region, DysFC. Several mutations that affect muscle regeneration occur within this small domain. The structure of the 123 residue inner DysF domain of human myoferlin, determined by multidimensional heteronuclear NMR studies, has a unique fold, and the conserved regions of the structure show extensive tryptophan/arginine stacking, which is vital for folding of this domain. Point mutations in these residues give rise to muscular dystrophies.

#36

Overview of Human Protein Disorder

Dr Melissa Pentony

Disorder predictions have become of increasing interest when determining protein structure and function. 'Disordered' is a term used to describe those regions or whole proteins that do not form a stable 3-D structure in their native state. The over-riding paradigm that protein structure directly infers protein function is challenged by the introduction of non-ordered regions. Subsequent studies of disordered regions have indicated that they are important areas, involved in many functions, such as DNA and protein binding, transcription and translational regulation (Iakoucheva, L. M. et al., 2004; Wright, P. E. et al., 1999). Disordered regions / whole proteins have been found within all organisms, but are more common within Eukaryotes. Previous studies on disorder have estimated that 33% of Eukaryotic proteins contain at least one region of > 30 contiguous disordered residues within their protein structure (Ward, J. J. et al., 2004b). Since the first-draft release of the human genome in 2001 (Lander, E. S. et al., 2001), functional analyses of the human proteome have been of increasing interest.

#37

"UCLAUC"- the new analytical ultracentrifugation laboratory at UCL

By Steve Perkins and Jayesh Gor

Analytical ultracentrifuge (AUC) methods are enjoying a resurgence of interest for reason of new sample detection methods and much improved data handling and structural analyses. BBSRC and Charity funding this year resulted in the purchase of a second AUC for UCL. To illustrate the scope of modern AUC velocity and equilibrium experiments, studies of a monomer-dimer system (complement FH-16/20), protein-protein and protein-carbohydrate interactions (complement FH-6/8 + heparin), and protein polydispersity (SC of IgA antibody) will be summarised. The commissioning of the second AUC now makes possible simultaneous velocity and equilibrium experiments so that protein behaviour in solution can be explored in detail. We discuss what is available and encourage new AUC proposals from UCL colleagues.

#38

Microtubules, an example for single particle analysis of filamentous proteins

Carsten Peters and Carolyn Moores

Microtubules are filaments that are present in all eukaryotic cells. As such they take part in many cellular events, among them cell division, organisation of cytoplasm and vesicle transport. These events are mediated by a plethora of proteins, among them the Kinesin motor protein superfamily. Yet the details of these interactions are still poorly understood. Kinesin 8 is of particular importance for mitosis as it facilitates nuclear movement, spindle positioning and chromosome segregation. Kinesin 8 is the only known Kinesin that functions as both a plus end motor and a microtubule depolymerase. We use electron microscopy to study the interaction of Kinesin 8 with the microtubule lattice. Images of microtubules decorated with Kinesin 8 are cut into segments and treated as single particles for 3D reconstruction.

#39

Fab1p and AP-1 are required for trafficking of endogenously ubiquitinated cargoes to the vacuole lumen in *Saccharomyces cerevisiae*

John P. Phelan¹, Stefan H. Millson², Peter W. Piper,² Frank T. Cooke¹ and Katherine Bowers¹.

¹ Department of Biochemistry and Molecular Biology, Darwin Building, University College London, Gower Street, London WC1E 6BT. UK

² Department of Molecular Biology and Biotechnology, Frith Court, Western Bank, University of Sheffield, S10 2TN. UK

The synthesis of phosphatidylinositol (3,5)-bisphosphate (PtdIns(3,5)P₂) by Fab1p in *S. cerevisiae* is required for several cellular processes, including an as yet undefined step in the ubiquitin-dependant trafficking of some integral membrane proteins from the trans-Golgi network to the vacuole lumen. AP-1 is a heterotetrameric clathrin adaptor protein complex that binds cargo proteins and clathrin coats, and regulates bi-directional protein trafficking between the trans-Golgi network and the endocytic/secretory pathway. Like *fab1* cells, AP-1 complex component mutants have lost the ability to traffic ubiquitinated cargoes to the vacuole lumen – the first demonstration that AP-1 is required for this process. Critically, these deletion mutants are distinct from the well-characterised class E mutants, where cargo proteins stall at aberrant endosomes adjacent to the vacuole. Deletion mutants of AP-1 complex components are compromised in their ability to synthesize PtdIns(3,5)P₂, indicating that AP-1 is required for correct *in vivo* activation of Fab1p. Furthermore, wild-type protein sorting can be restored in AP-1 mutants by overexpression of Fab1p, implying that the protein-sorting defect in these cells is as a result of disruption of PtdIns(3,5)P₂ synthesis. We show that Vac14p, an activator of Fab1p, is also required for another AP-1-dependent process: chitin-ring deposition in *chs6* cells. Our data imply that AP-1 is required for some Fab1p and PtdIns(3,5)P₂-dependent processes. Currently, we are working on a genetic screen to identify novel components of the AP-1/PtdIns(3,5)P₂-dependant step(s) in sorting of ubiquitinated cargo to the Multi-vesicular Body (MVB)/vacuole.

#40

Predicting the antibody VH/VL packing angle from interface residues

Mr Abhinandan Raghavan

The antibody packing angle is defined as the angle at which the variable region light chain (VL) and heavy chain (VH) pack with one another. Variability in this angle will affect the topography of the antigen combining site. Getting the angle correct is therefore important in modelling antibodies and, if specific residues are important in defining the packing angle, these must be included in humanization of mouse antibodies.

We have assessed the variability in the packing angle and find that it varies by up to 30 degrees. We hypothesize that residues in the VH/VL interface determine the packing angle and have tried to identify important interface residues.

Neural networks were used to predict packing angle from a description of residues in the VH/VL interface. Since the number of potential interface residues is very large compared with the limited data available for training the neural network, we performed feature selection using a genetic algorithm in which the penalty function is the performance of the neural network.

#41

Heteronuclear NMR investigations of DDAH enzyme structure and dynamics

Masooma Rasheed

Department of Biochemistry and Molecular Biology, University College London.

Dimethylarginine dimethylaminohydrolase (DDAH) is an enzyme found in all mammalian cells. It is involved in the degradation of methylarginines, specifically asymmetric dimethylarginine (ADMA) and NG-monomethyl-L-arginine (MMA). The methylarginines ADMA and MMA inhibit the production of nitric oxide synthase. Inhibition of DDAH activity causes methylarginines to accumulate, which results in the blockage of nitric oxide (NO) synthesis. The dysfunction of DDAH activity results in the elevation of plasma ADMA level and impairment of vascular relaxation observed in humans with increased cardiovascular disease or risk factors (such as hypercholesterolemia, diabetes mellitus, and insulin resistance). The two isoforms of DDAH (1 & 2) are found in mammals. The DDAH from bacterium source, *Pseudomonas aeruginosa* (PaDDAH) was studied by Beatriz Magalhaes (2006) at the laboratory

of Biochemistry and Molecular Biology (UCL) is so far the only structurally tractable homologue of mammalian DDAH isoforms. For the study of interaction between enzyme with substrate and inhibitor bound ligands, she obtained PaDDAH in its monomeric form by substitution of several interface residues which resulted in the shifting of equilibrium position towards the monomer, which allowed her to design a double mutant variant (Agr40→Glu, Arg98→His) that behaved exclusively as monomer and retained 95% catalytic activity more than wild type (WT). The theme of the present research is linked to the research previously done by the above mentioned scientist. The first task is to recapitulate the methods for protein expression, purification, isotope labeling of monomeric WT and C249S PaDDAH for NMR studies. The polypeptide chain of PaDDAH in apo and inhibitor bound states will be observed and study of dynamics of C249S PaDDAH in substrate and product bound states will be performed. The emphasis of research will be more towards the high expression of human DDAH isoforms (1 & 2), optimal for isotope labeling and thus can be utilized for the examining the structural design by using NMR technique. Our approach would be to assess the expression level for 'synthetic genes' that have optimal codon usage for *Escherichia coli* expression where consideration of mRNA secondary structure has been applied. The exploration of comparative NMR characteristics of Mycobacterium tuberculosis DDAH will be done, in order to express the MtDDAH in sufficient yield and form that is tractable for NMR studies. The analysis of data will be performed using CCPN (Collaborative Computing Project) for NMR and the previous available data files will also be converted into CCPN format. The comparative screening of PaDDAH, MtDDAH and human DDAH isoforms (1&2) will help us to differentiate the role of DDAH in bacterium and human which can be interacted with a chemical library as model systems for early stage drug discovery efforts.

#42

CATHEDRAL: a novel method for assigning domain boundaries and folds to multi-domain proteins

Oliver Redfern

We present CATHEDRAL, an iterative protocol for determining the location of previously observed protein folds in novel multi-domain protein structures. CATHEDRAL builds on the features of a fast secondary-structure-based method to locate known folds (GT) within a multi-domain context and a residue-based, double-dynamic programming (DDP) algorithm, which is employed to align members of the target fold groups against the query protein to identify the closest relative and assign domain boundaries. To increase the fidelity of the assignments, Support Vector Machines are utilised to provide an optimal scoring scheme. Once a domain is verified, it is excised and the search protocol is repeated in an iterative fashion until all recognisable domains have been identified.

We have performed an initial benchmark of CATHEDRAL against other publicly available structure comparison methods using a consensus data set of domains derived from the CATH and SCOP domain classifications. This shows superior performance in fold recognition and alignment accuracy when compared to many equivalent methods.

If a novel multi-domain structure contains a known fold, CATHEDRAL will locate it in 90% of cases, with <1% false positives. For nearly 80% of assigned domains in a manually validated test set, the boundaries were correctly delineated within a tolerance of ± 10 residues. For the remaining cases, previously classified relatives were very remote from the query chain and embellishment to the core of the fold caused significant differences in domain sizes; hence, manual refinement of the boundaries was necessary. Since, on average, 50% of newly determined protein structures contain more than one domain unit and typically 90% or more of these are already classified in CATH, CATHEDRAL will considerably facilitate the automation of protein classification.

#43

New methods for Combinatorial Domain Hunting

Dr James Reid

Abstract not available

#44

Combining methods of remote homology detection can increase coverage by 10% in the midnight zone

Adam Reid

A recent development in sequence-based remote homologue detection is the introduction of profile-profile comparison methods. They are much more powerful than previous technologies and frequently detect potentially homologous relationships not captured by structural classifications such as CATH and SCOP. As structural classifications traditionally act as the gold standard of homology, this poses a challenge in benchmarking them. We present a novel approach which allows an accurate benchmark of these methods against the CATH structural classification. We then apply this approach to assess the accuracy of a range of publicly available methods of remote homology detection including several profile-profile methods (COMPASS, HHSearch, PRC) from two perspectives. Firstly in distinguishing homologous domains from non-homologues. Secondly in annotating proteomes with structural domain families. PRC is shown to be the best method for distinguishing homologues. We show that SAM is the best practical method for annotating genomes, whilst using COMPASS for the most remote homologues would maximise data quality. Finally we introduce a simple approach to increase the sensitivity of remote homology detection by up to 10%. This is achieved by combining multiple methods with a jury vote.

#45

The role of resuscitation promoting factors in the control of bacterial dormancy

Giles F. Robertson[‡], Martin Cohen-Gonsaud[§], Brian Henderson[‡], John Ward[†], Nicholas H. Keep[‡]

[‡] School of Crystallography, Birkbeck College, University of London, London, UK.

[‡]Division of Microbial Diseases, UCL Eastman Dental Institute, University College, London, UK.

[§]Centre de Biochimie Structurale, Montpellier, □France.

[†]Department of Biochemistry and Molecular Biology, University College, London, UK.

Many bacteria seem able to persist in a low growth state after exposure to stress. The adoption of a reversible dormant state in pathogenic bacteria has significant health implications due both to the difficulty in detection of low levels of dormant bacteria and the prolonged treatment periods required to kill such bacteria. In some cases, the exit from this growth state requires specific stimuli; the resuscitation promoting factors (RPF) are secreted proteins first found expressed by the gram-positive bacterium *M. luteus*. Dormant *M. luteus* cells could be induced to grow upon treatment with purified native or recombinantly produced RPF. Proteins homologous to RPF have been found in many pathogenic bacteria including five in *M. tuberculosis*. One member of the five, RPF B, was shown first to have sequence similarity to lysozyme and then structural homology when the NMR structure of the RPF domain of this protein was determined. A functional role of cell wall remodelling through a lysozyme like digestion of peptidoglycan seemed likely, however, we were unable to detect a corresponding lysozyme activity. We now report an HPLC assay that shows for the first time Rpf activity for RpfB and *M.luteus* Rpf. However, the activities were found to be orders of magnitude less than lysozyme. Given that the RPF proteins were capable of resuscitating dormant bacteria at picomolar levels, a simple non-specific cell wall remodelling function seems unlikely. Alternate functional mechanisms are discussed.

#46

Improving Template Selection for Homology Modeling

M.I. Sadowski & D.T. Jones

Homology modeling, the most accurate method currently available for predicting the structure of a protein sequence, generates a model using a homologous template of known structure. Although there has been some recent progress the structural similarity between the template and the target is still the limiting factor in the accuracy of the model. Selecting the proper template is therefore paramount to generating models at the highest resolution. It is well known that sequence similarity is a good predictor of structural similarity, however it is not perfect and the most similar sequence is not always the best template to use. We have recently compared alternative template selection methods on a benchmark set of 732 proteins

from 24 CATH families (Sadowski & Jones, /Proteins/, /in press/) and found that for highly similar proteins sequence identity is still the best predictor. More sophisticated profile-based methods improve template similarity only at < 40% sequence identity. Even at high levels of sequence-similarity (50-80%) choosing the most similar sequence leads to errors of 0.5 Angstroms or more for ~25% of our benchmark set.

In this poster we show three examples of incorrect template selections based on sequence similarity and describe how an identity measure based on a subset of sequence positions can be derived using a simple genetic algorithm and used to improve selection.

#47

Biochemical Characterisation of Archaeal TIP

Anne Schreiber

Abstract not available

#48

Study of the Hsp90 chaperone with NMR

Eleanor Smith

Hsp90 is a ubiquitous chaperone protein whose function, unlike other chaperone proteins that aid in de novo folding, is to mature and activate already folded proteins. The protein targets of Hsp90 include steroid hormone receptors and protein kinases.

Hsp90 is a homodimer, each monomer consists of three domains (N-terminal, middle and C-terminal). The N-terminal domain shows ATPase. Hsp90 binds to an incredibly diverse range of substrates with the aid of various co-chaperone molecules. These co-chaperones include Hsp70, Hop, Aha1, p23, Cdc37 and PP5 and have functions ranging from control of the ATPase activity to acting as adaptors and recruiters of substrate proteins. Some co-chaperones, particularly those associated with the C terminal TPR binding domain, are still of unknown function. The exact method of substrate activation remains unknown.

This project aims to investigate this issue using NMR spectroscopy to monitor structural changes in Hsp90 and its kinase substrates during the ATPase cycle.

Previously an NMR assignment of the N-terminal domain has been obtained. We are working to optimise expression and purification of other Hsp90 constructs using a ¹⁵N perdeuterated labeling strategy and spectra have so far been obtained for amino acids 271-551 (middle domains) and 1-551 (N-terminal and middle domains). Both constructs show good peak dispersion which should allow assignments of the backbone amides and methyl groups as a further step towards structural studies of the functional protein, which as a 165kDa dimer is a significant challenge for NMR methods.

#49

The large-scale fermentation and bioprocessing of bacteriophage

Dr Emma Stanley

Bacteriophage M13 and lambda are well characterised viruses, which over many years have been the workhorses of molecular biology laboratories. In recent years, research and commercial interest into these bacteriophage has increased as the plethora of potential uses in the pharmaceutical market has been recognized. Many techniques used for the growth and purification of bacteriophage at small-scale are not transferable to large-scale production of bacteriophage in industrial processes.

In our studies high titre M13 production in E. coli has been optimised for fermentation conditions using growth curves of M13 production. The final titre of M13 produced during a 2 L fermentation was independent of the multiplicity of infection used, however the time required to achieve the maximum titre of M13 was dependent on the MOI.

Initial fermentation studies for lambda production have concentrated on overcoming several of the difficulties of large-scale lambda down-stream processing through the construction of a genetically engineered E. coli host. The E. coli strain produces a non-specific nuclease which is exported to its periplasm. Upon lysis of the E. coli by lambda the nuclease is released into the surrounding media and successfully degrades the E. coli nucleic acid. The use of this engineered host removes the need to add bovine or commercial nucleases when producing pharmaceutical grade lambda bacteriophage.

#50

Spatial and temporal control of early signalling events following Interferon alpha and T-cell Receptor ligation

Claire Stevens

Interferons (IFNs) are pluripotent, antiviral cytokines that signal through ubiquitously expressed dimeric receptors. The JAK-STAT is the best characterised signalling pathway emanating from the interferon receptor, but it is now becoming increasingly clear that this cascade alone is not sufficient to generate all of the biological effects of interferons upon different cell types. For example, in recent years it has been illustrated that in Jurkat T-cell line, ligation of the interferon also activates the MAPK pathway. Furthermore, T-cell receptor associated proteins, including Lck, Zap-70 CD45, are all imperative for this MAPK phosphorylation.

In the present study, we showed that two additional T-cell receptor associated proteins become activated on IFN α R stimulation in the Jurkat T cell lines. Firstly, the 95kDa guanosine nucleotide exchange factor, Vav, is shown to be phosphorylated and its expression forms a pre-requisite for the IFN α -induced MAPK response. In addition, a novel role is ascribed to the 76kDa adaptor protein, SIp-76, at the IFNAR. Absence severely diminishes the IFN α -dependent MAPK response and involvement of SIp-76 in this pathway is also shown to be phosphorylation dependent.

From these, and previous studies, it is clear that the IFN α R utilises a growing number of adaptor and kinase proteins which also are recruited in TCR signalling. This raises the possibility that crosstalk occurs between the IFN α R and TCR. To demonstrate this we showed that the MAPK response observed post-IFN α R ligation, relies on an intact TCR being expressed at the cell surface. These results highlight for the first time an intimate connection between the TCR and IFN α R.

#51

Data Processing in Circular Dichroism Spectroscopy

Tim Stone

Circular Dichroism (CD) spectroscopy is a method for the biophysical characterisation of protein molecules. It is able to tell the user the percentage composition of secondary structure components that are present in the protein molecule. The advantages of this technique are that concentrations can be very low and the protein does not need to exist in any particular special form in order for the process to be undertaken, in contrast to other techniques.

A number of methods of characterisation and comparison of CD spectroscopy of proteins have been developed in the past. I have been investigating ways to extend such methods of comparison of protein CD spectra with each other by the development of a comparison metric. In addition to this work, I have also been using machine learning methods (namely Support Vector Machines and some Gaussian processes) to try to improve the accuracy of prediction of the amount of secondary structure elements in the protein under investigation. Our group has developed a carefully calibrated dataset using synchrotron radiation CD spectroscopy which has extended the quality and wavelength range of training data used in this method. Using this data, I aim to improve upon methods of characterisation that currently exist.

#52

Structural Investigation of Human PP2A Regulation by Protein Methylation

Tsai, Meng-Lin and Snezana Djordjevic,
Dept. of Biochemistry and Molecular Biology, University College London

Protein serine/threonine phosphatase 2A (PP2A), a conserved protein found in many species, plays an important role in various cellular processes including cell growth, differentiation, signal transduction, DNA replication, transcription, protein synthesis and apoptosis. Deregulation of PP2A methylation is linked to Alzheimer's disease and increased susceptibility to pathogen infections. PP2A was identified as an important tumour suppressor protein and hence a potential target for cancer therapeutic strategies. PP2A comprises a core structure of a 65 kDa scaffolding subunit A and a 36 kDa catalytic subunit C, which associates with a variable regulatory subunit B to form a heterotrimeric holoenzyme. Recently, two different x-ray crystallographic structures of PP2A were determined with the B' (B56-r1/PR61) subunit. Different heterotrimeric compositions influence the enzyme's cellular location and substrate

specificity.

Leucine carboxyl methyltransferase 1 (PPM1) and protein phosphatase methylesterase 1 (PME1) are involved in PP2A's methylation and de-methylation respectively. Recent research has demonstrated that methylation of the PP2A C subunit at the carboxyl-terminus, residue Leu309, by PPM1 is essential in facilitating formation of the heterotrimeric PP2A complex. In contrast, PME1 is thought to associate with and stabilize the inactivated form of PP2A complex.

Work presented here aims to determine the crystal structure of PPM1 and PME1, the effect of the putative catalytic residue mutations of PME1, and further research molecular basis of PP2A and PME1/PPM1 interactions.

#53

Structural biology of the P pilus rod component PapA in uropathogenic E. Coli

Dr Denis Verger

P pili are important adhesive fibres involved in kidney infection by uropathogenic Escherichia coli strains. P pili are assembled by the conserved chaperone–usher pathway, which involves the PapD chaperone and the PapC outer membrane usher. During pilus assembly, subunits are incorporated into the growing fibre via the donor–strand exchange (DSE) mechanism, whereby the chaperone's G1 α -strand that complements the immunoglobulin-like fold of each subunit is displaced by the N-terminal extension (Nte) of an incoming subunit. P pili comprise a helical rod, a tip fibrillum, and an adhesin at the distal end. PapA is the rod subunit and is assembled into a super-helical right-handed structure. Here, we have solved the structure of a ternary complex of PapD bound to PapA through donor–strand complementation, itself bound to another PapA subunit through DSE. This structure provides insight into the structural basis of the DSE reaction involving this important pilus subunit. Using gel filtration chromatography and electron microscopy on a number of PapA Nte mutants, we establish that PapA differs in its mode of assembly compared with other Pap subunits, involving a much larger Nte that encompasses not only the DSE region of the Nte but also the region N-terminal to it.

#54

Assembly of a BraF35:Kif4 complex

Matt Webster

BRAF35 was initially isolated from a 2Mda complex also containing the Breast Cancer susceptibility gene product BRCA2. BRAF35 is believed to have a dual role in cell cycle progression and also DNA damage repair by potentially recruiting BRCA2 to recombinogenic sites by virtue of a HMG box domain. In 2003 Lee and Kim utilised yeast two hybrid screens and immuno-precipitation to show that BRAF35 could also bind to truncated Human Kinesin KIF4 in a complex of 540kDa. KIF4 has a poorly defined but essential role in the spindle formation during Mitosis. The interaction between the BRAF35 and KIF4 proteins is believed to be through respective α helical coiled-coil domains and this complex may play a role in regulating cell cycle progression during mitosis.

Ultimately this work aims to determine the structural basis of interaction between BRAF35 and KIF4/BRCA2 to guide future cell biology studies. Towards this end, it has been possible to over-express and purify both BRAF35 and KIF4. It has also been possible to assemble and purify a BRAF35/Kif4 complex. The methods used and challenges faced will be presented.

#55

The interactions of VirB11

Robert Williams

Abstract not available

#56

ValiDichro- Validating Circular Dichroism Spectra

Ben Woollet

Abstract not available

#57

A Uniform Approach to Bioinformatics Workflow and Data Integration

Mr Lucas Zamboulis

Data integration in the life sciences requires resolution of conflicts arising from the heterogeneity of data resources and from incompatibilities between the inputs and outputs of services used in the analysis of the resources. We present an approach that addresses these problems in a uniform way. We also present results from the application of our approach for the integration of bioinformatics data resources within the ISPIDER and the BioMap projects and for the reconciliation of bioinformatics services. The ISPIDER integration setting demonstrates an architecture in which the AutoMed heterogeneous data integration system interoperates with grid access and query processing tools for the virtual integration of a number of proteomics resources, while the BioMap integration setting demonstrates the materialised integration of structured and semi-structured functional genomics resources using XML as a unifying data model. The service reconciliation scenario discusses the interoperation of the AutoMed system with a scientific workflow tool. The work presented here is part of the ISPIDER project, which aims to develop a platform using grid and e-science technologies for supporting in silico analyses of proteomics data.