

6. Martin Noble

Adhesive interactions, the cell cycle, and NAT enzymes

Work in this group addresses structure-function relationships of medically important proteins from three different areas: adhesive cellular interactions, the eukaryotic cell cycle, and the enzyme arylamine N-acetyltransferase (NAT). We study these proteins by both experimental and theoretical approaches. Experimentally, proteins are subject to biochemical and biophysical characterisation, as well as structural analysis by X-ray crystallography and nuclear magnetic resonance. Development of theoretical methods has centred on developing tools for molecular analysis. These include programs to study protein surfaces in terms of their chemical character as indicated by sequence conservation, electrostatic potential, or other potentials calculated by the program GRID. We have also been exploring the use of this approach to identify probable sites of protein-protein interaction.

Adhesion and signalling from adhesive complexes

The cell membrane is a location of bidirectional signalling. External events may activate signalling networks within the cell, while internal signals may be transduced through modulation of the properties of cell-surface proteins that mediate cellular adhesion to the extracellular matrix or to other cells. We are studying two prototypical signalling systems that involve both outside-in and inside-out signalling. These are CD44, a transmembrane protein found on the surface of many cell-types in mammals, where it acts as the major receptor for the glycosaminoglycan hyaluronan (HA), and focal adhesions, the cellular substructure that forms where integrins bind to proteinaceous components of the extracellular matrix. To date we have explored the structure of the HA binding domain of CD44, as well as various aspects of the structure and function of different parts of the focal adhesion kinase (FAK) molecule. We have also looked at proteins involved in signalling from activated T-cell receptors.

CD44

(With Dr. D. Jackson, IMM, Dr. A.J. Day, MRC Immunochemistry)

Adhesive interactions involving CD44, the cell surface receptor for hyaluronan, underlie fundamental processes such as limb morphogenesis, wound healing, leukocyte migration and tumor metastasis. Critical to these events, the regulation of CD44's hyaluronan-binding activity is known to be effected by changes in N-glycosylation, switching the receptor "on" under appropriate circumstances. How glycosylation influences CD44 function has until now been unclear. Like many hyaluronan-binding proteins found in extracellular matrix, CD44 contains a conserved lectin-like domain termed the Link module. However, CD44 is unique in that regions of the extracellular domain additional to the Link module are required for receptor function, and evidence suggests these "extensions" are involved in regulation. We have shown using X-ray crystallography and NMR spectroscopy that sequences flanking the Link module form a supplementary structural lobe that extends the main hyaluronan-binding surface. Moreover, the location of key N-glycosylation sites revealed for the first time how such glycans might regulate CD44 function. This year we have pursued crystallographic studies with mouse CD44 with a view to producing crystals that lend themselves to binding studies. These have resulted in two crystal forms that have provided high resolution structures of the complex between CD44 and an 8mer derived from its natural HA ligand. This complex explains much about the activity and regulation of CD44, and offers a basis for structure-based drug design targeting CD44-HA interactions.

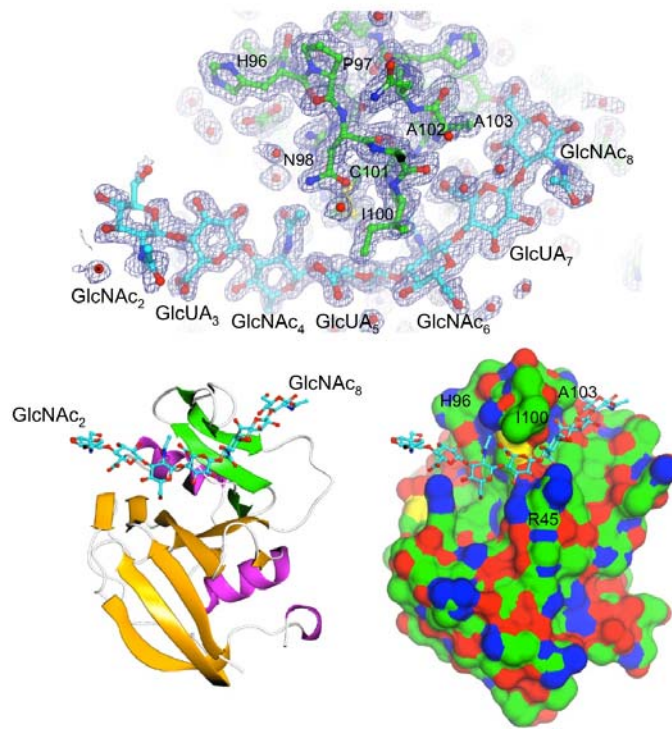


Figure 1 The HA binding site of mCD44.

A) Final refined 1.3 Å resolution 2Fo-Fc electron density for the binding site in crystal I, calculated using SIGMAA weighted map coefficients generated by REFMAC, is contoured at 0.25 e-Å⁻³ (equivalent to the standard deviation of the final map). The refined structure is shown in ball and stick representation coloured by atom type (green:carbons, blue:nitrogens, red:oxygen), with the exception of carbon atoms of HA that are coloured cyan.

B) A ribbon diagram of mCD44 is shown, coloured as for figure 1,

C) Surface representation of the HA binding site in crystal II. The cleft between the two subdomains of CD44 into which HA binds is shown in a molecular surface representation. A similar representation of crystal I shows similar features, but lacks the lower platform for the HA interaction offered by reorientation of Arg45.

The Focal adhesion targetting domain.

Maria Hoellerer and Sonja Lorentz (With Dr. S. Arold, CNRS Montpellier, Prof. I.Campbell, Biochemistry)

FAK is localised to focal adhesions via its C-terminal Focal Adhesion Targeting (FAT) domain. FAT performs this function by binding to paxillin and talin, both of which are in turn associated with the cytoplasmic tails of integrins. FAT is also required for binding of FAK substrates: a phospho-tyrosine motif within FAT binds to the Src homology (SH) 2 domain of the adapter protein Grb2. A proline-rich sequence immediately upstream of the FAT domain binds to the SH3 domains of p130 CAS, Graf, and the p85 subunit of phosphatidylinositol-3 kinase.

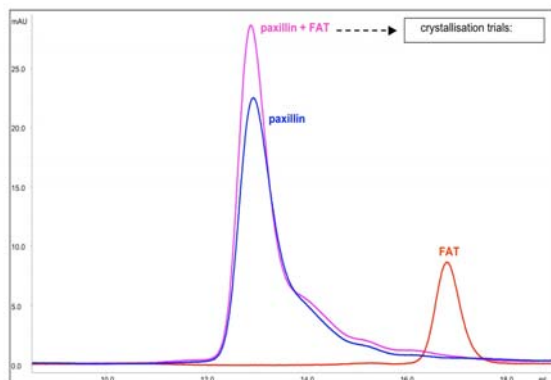


Figure 2: Size exclusion chromatography of FAT/paxillin N-terminal domain fragment

This year we have established that the FAT domain binds to long-forms of all 5 LD motifs of paxillin, indicating a highly relaxed specificity, the implications of which we are exploring by further binding studies. Preliminary indications suggest that we can form stable complexes of FAT with paxillin fragments that contain more than one LD motif. Structural and biophysical studies of these complexes will be pursued.

T-cell signalling

Ewa Pilka

As is the case in many other intracellular signalling systems, fidelity in T-cell signalling emerges from the formation of complexes of many proteins, each one of which may have comparatively poor selectivity in its interactions. In order to understand the mechanisms underlying this observation, it is necessary to explore the structural basis of the communication between the multiple interactions that stabilise such complexes. To this end, we have this year determined the structure of a tandem SH2 domain construct derived from phospholipase C-gamma. Two copies in the asymmetric unit of this structure show similar but subtly differing relative orientations of the constituent SH2 interactions, indicating that while each is able to independently engage ligands, there is likely to be some amount of stereochemical constraints involved in their simultaneous binding of partners.

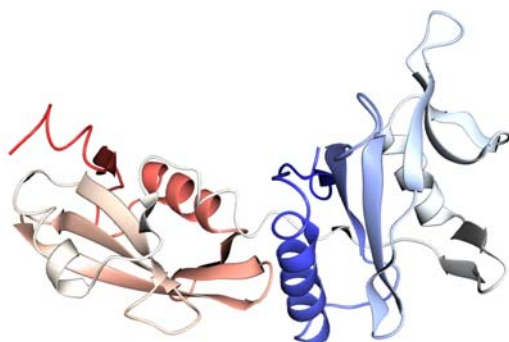


Figure 3. Structure of the tandem SH2 domains of PLC-gamma 1.

Cell cycle

We have collaborated with other groups in the LMB on various aspects of cell-cycle regulation over recent years, the results of which are discussed elsewhere in this report (see reports from Prof. Louise Johnson, Dr. Jim McDonnell, and Prof. Jane Endicott).

NAT enzymes.

With the group of Professor Edith Sim, Pharmacology

The NATs (arylamine N-acetyltransferases) are a well documented family of enzymes found in both prokaryotes and eukaryotes. NATs are responsible for the acetylation of a range of arylamine, arylhydrazine and hydrazine compounds. This year, we have published an investigation into the catalytic triad of residues (Cys-His-Asp) and other structural features of NATs using a variety of methods, including site-directed mutagenesis, X-ray crystallography and bioinformatics analysis, in order to investigate whether each of the residues of the catalytic triad is essential for catalytic activity. The catalytic triad of residues, Cys-His-Asp, is a well defined motif present in several families of enzymes. We mutated each of the catalytic residues in turn to investigate the role they play in catalysis. We also mutated a key residue, Gly126, implicated in acetyl-CoA binding, to examine the effects on acetylation activity. In addition, we have solved the structure of a C70Q mutant of *Mycobacterium smegmatis* NAT to a resolution of 1.45 Å (where 1 Å=0.1 nm). This structure confirms that the mutated protein is correctly folded, and provides a structural model for an acetylated NAT intermediate. Our bioinformatics investigation analysed the extent of sequence conservation between all eukaryotic and prokaryotic NAT enzymes for which sequence data are available. This revealed several new sequences, not yet reported, of NAT paralogues. Together, these studies have provided insight into the fundamental core of NAT enzymes, and the regions where sequence differences account for the functional diversity of this family. We have confirmed that each of the three residues of the triad is essential for acetylation activity

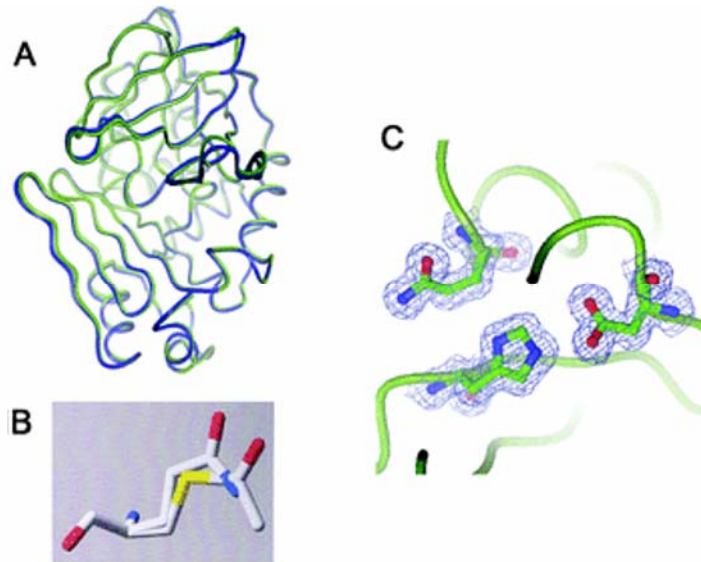


Figure 4: A) Comparison of the peptide backbone of wild-type *M. smegmatis* NAT (blue) and the C70Q mutant (green), showing the catalytic triad as a ball and stick representation. B) Superposition in the model of acetylcysteine and glutamine. C) Close-up view of the modified catalytic triad in the C70Q *M. smegmatis* NAT. Maps are contoured at 1.5 s.

Methods

Visualisation/analysis.

Jan Gruber, Alexander Zawaira, Rhodri Saunders

We have developed and implemented algorithms for the analytical calculation of protein surfaces, and for the evaluation of electrostatic potential through solving the Poisson Boltzmann equation by finite difference methods. These algorithms have been implemented in the programs AESOP, ccp4mg, and coot. We have also designed a novel algorithm for assigning amino-acid conservation from analysis of aligned sequences that promises to reveal important functional surfaces.

This year, we have begun to explore the utility of these methods in predicting the character of protein:protein interactions. Our electrostatic calculations have demonstrated that electrostatic complementarity is a defining property of authentic protein:protein complexes, and we have further demonstrated that the mean sequence conservation of surface buried in a putative protein:protein complex is a sensitive discriminator of whether such interactions can occur for a particular protein sequence.

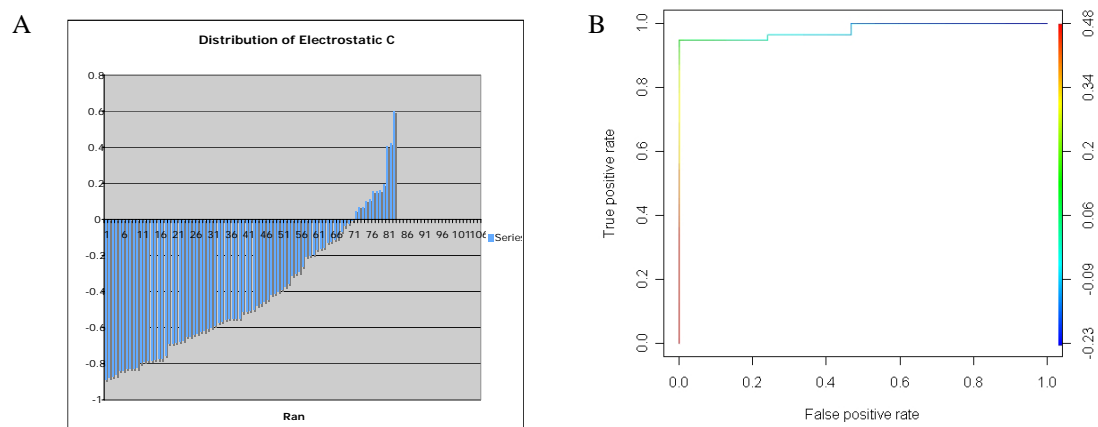


Figure 5: Analytical tools: A) Electrostatic complementarity of a test set of protein:protein interaction complexes (negative correlation indicates that surfaces are complimentary) B) Receiver Operator Curve analysis indicating the effectiveness of conservation-burial as a discriminator of cyclin dependent kinases from other serine threonine kinases.

Protein Lattices.

John Sinclair, Eugene Valkov

We have been exploring the use of protein fusions to generate stable lattices for application in diverse biotechnological and bionanotechnological projects. This approach has begun to demonstrate the feasibility of exploiting binary systems to generate ordered assembly into higher order structures.

References.

J. Sandy, A. Mushtaq, S.J. Holton, P. Schartau, M.E. Noble, E. Sim, Investigation of the catalytic triad of arylamine N-acetyltransferases: essential residues required for acetyl transfer to arylamines. *Biochem J* 390 (2005) 115-23.

C.P. Barrett, M.E. Noble, Dynamite extended: two new services to simplify protein dynamic analysis. *Bioinformatics* 21 (2005) 3174-5.

J. Sandy, S. Holton, E. Fullam, E. Sim, M. Noble, Binding of the anti-tubercular drug isoniazid to the arylamine N-acetyltransferase protein from *Mycobacterium smegmatis*. *Protein Sci* 14 (2005) 775-82.

C.P. Barrett, M.E. Noble, Molecular motions of human cyclin-dependent kinase 2. *J Biol Chem* 280 (2005) 13993-4005.

C.P. Barrett, B.A. Hall, M.E. Noble, Dynamite: a simple way to gain insight into protein motions. *Acta Crystallogr D Biol Crystallogr* 60 (2004) 2280-7.

I.M. Westwood, S.J. Holton, F. Rodrigues-Lima, J.M. Dupret, S. Bhakta, M.E. Noble, E. Sim, Expression, purification, characterization and structure of *Pseudomonas aeruginosa* arylamine N-acetyltransferase. *Biochem J* 385 (2005) 605-12.