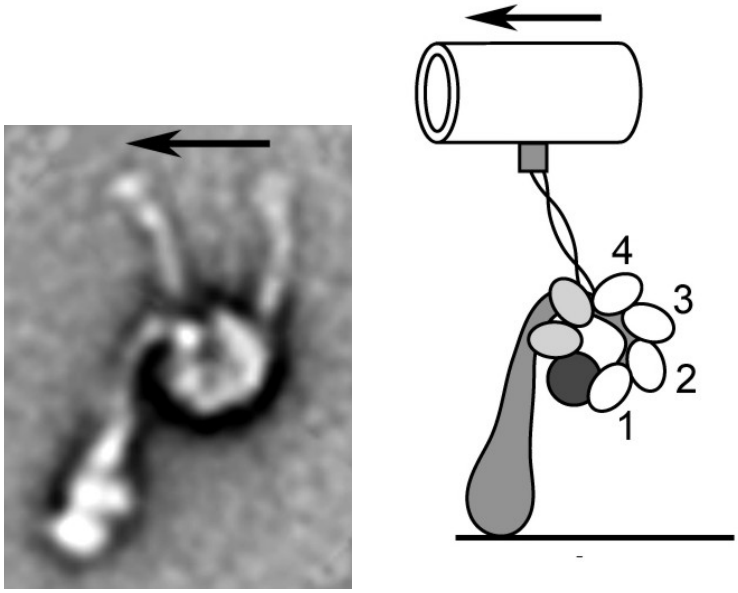
Astbury Centre for Structural Molecular Biology

University of Leeds



Annual Report 2002

Front cover illustration:

The power stroke of the dynein molecular motor.

The left panel shows a superposition of two conformational states of dynein seen for the first time before and after its power stroke. These conformations were produced by the presence and absence respectively of ADP.vanadate. The molecule consists of an elongated stem and stalk connected to a ring-like head. The arrow indicates that a 15 nm displacement of the tip of the stalk is apparent when the other parts are superposed. The images were generated by negative stain electron microscopy, followed by single-particle image processing. This isoform of dynein was purified from flagella of the alga *Chlamydomonas reinhardtii* in the laboratory of Dr K. Oiwa, KARC, Kobe, Japan.

The right panel shows a new model of the dynein structure and how it interacts with its microtubule track (microtubule not drawn to scale). The model combines the new image data with the domain structure deduced from dynein sequences. Four highly-conserved AAA domains are indicated, the first of which is the site of ATP hydrolysis. The ATP-sensitive microtubule-binding domain is at the tip of the anti-parallel coiled coil stalk, 28 nm away.

Burgess, S.A., Walker, M.L., Sakakibara, H., Knight, P.J. and Oiwa, K. (2003) "Dynein structure and power stroke" *Nature*, **421**, 715-718.

Acknowledgement

The Astbury Centre for Structural Molecular Biology thanks its many sponsors for support of the work, its members for writing these reports and Jenny Walker for collating and type-setting this report. The report is edited by Alan Berry.

This report is also available electronically via http://www.astbury.leeds.ac.uk

Mission Statement

The Astbury Centre for Structural Molecular Biology will promote interdisciplinary research of the highest standard on the structure and function of biological molecules, biomolecular assemblies and complexes using physico-chemical, molecular biological and computational approaches.

Introduction

It is once again a great privilege as the Director of the Astbury Centre for Structural Molecular Biology (ACSMB) to write the introduction to our Annual Report (2002). The ACSMB is an interdisciplinary research centre of the University of Leeds founded in 1999 to carry out research in all aspects of structural molecular biology. ACSMB brings together over fifty academic staff from the faculties of Biological Sciences, Chemistry, Physics and Medicine to examine the molecular mechanisms of life. It is named after W.T. (Bill) Astbury, a biophysicist who laid many of the foundations of the field during a long research career at the University of Leeds (1928-1961). Astbury originally identified the two major recurring patterns of protein structure (α and β), took the first X-ray fibre diffraction pictures of DNA (in 1938) and is widely credited with the definition of the field of molecular biology. As I write in 2003, we are celebrating the 50th anniversary of the publication of the double helical structure of DNA by the Cambridge and King's College, London groups, a result that put molecular descriptions at the heart of all Biology. The explosive progress that has occurred following this structural insight fully illustrates Astbury's vision of the new field. As an outsider to the dramatic developments that occurred in our field in 1953, it has always struck me as significant that progress was the result of interdisciplinary interactions. Careful and painstaking experimental work by Franklin, Wilkins and their colleagues, coupled to the relatively new theory behind helical diffraction (Crick & Stokes), together with the deep understanding that the structure must say something profound about genetics (Watson) came together, sometimes controversially, to yield the "big picture". I believe that ACSMB continues to provide an excellent interdisciplinary environment for continued progress to be made. I'm sure that the pages that follow show ample proof that this is the case.

This report describes some of the highlights of our work over the last year. These reports have largely been written by our younger researchers. Their tremendous enthusiasm for this kind of interdisciplinary work augurs well for our future. I was particularly struck by the breadth of activity in the Centre ranging from the sophisticated applications of synthetic organic chemistry to the developments in single molecule biophysics. In between these extremes you will find groundbreaking activity in many traditional areas for structural biology. ACSMB has always been outward looking and this tradition continues with the many external collaborations acknowledged in these pages, both from within the UK and beyond. We would welcome discussions with anyone wishing to collaborate or simply to make use of our facilities, the details of which can be found on our web page (http://www.astbury.leeds.ac.uk/).

These brief summaries, however, only scratch the surface of the work of the Centre. I hope you enjoy reading them, and if you wish to learn more please visit our website or contact the Director.

This annual report is also available as a 7.25MB PDF document that can be download.

Peter G. Stockley
Director, Astbury Centre for Structural Molecular Biology
Leeds, May 2003

Contents

Mass spectrometry facility

Alison E. Ashcroft

Overview of facility

The mass spectrometry (MS) facility has a Platform II electrospray ionisation (ESI) quadrupole instrument with on-line HPLC and CE, a Q-Tof orthogonal acceleration quadrupole-time-of-flight tandem instrument with nano-ESI, a TSQ 7000 tandem quadrupole instrument with ESI, and a surface enhanced laser desorption ionisation/matrix assisted laser desorption ionisation (SELDI/MALDI) mass spectrometer. There is also a MALDI instrument specifically for high-throughput proteomics screening in the department. The facility runs an analytical service as well as being actively involved in several research areas within the Astbury Centre for Structural Molecular Biology and the Faculty of Biological Sciences, in addition to other groups within the University and external collaborators.

Research

The research involves the development of MS techniques to aid the structural elucidation of biomolecules and can be categorised into four main areas:

i). Protein folding (see Fig. 1). Protein folding is an intriguing area of biochemistry and protein mis-folding is thought to be a contributing factor to several diseases. Working with Prof. S.E. Radford's group, ESI-MS is being used to monitor β_2-microglobulin conformations using charge state distribution analysis, enzymatic digestion and H/D exchange to gain insights into folding intermediates and fibril formation.

ii). Protein-ligand non-covalent interactions and macromolecular assembly. In collaboration with Profs. P.G. Stockley and S.E. Radford and Dr N.J. Stonehouse, ESI-MS is being used to investigate non-covalently bound macromolecular structures. Such studies include protein-peptide, protein-protein, and protein-RNA complexes. The latter are important in virus assembly, an area we are investigating with respect to the MS2 and Qβ systems. Protein-protein macromolecular complexes are critical species in fibrillogenesis and are under investigation as an integral part of our β_2-microglobulin folding studies.

iii). Reaction monitoring. A recent project includes measuring the uptake of ATP by the muscle protein myosin in collaboration with Profs. H. White (Eastern Virginia Medical School, USA) and J. Trinick.

iv). Structural elucidation and proteomics. Tandem MS (MS/MS) sequencing of proteins and peptides is an important bioanalytical technique. This is being used in collaboration with Prof. P.J.F. Henderson and Dr R.B. Herbert for the characterisation of labelled membrane proteins. The characterisation of Alzheimer's disease-related peptides is being carried out with Dr C. Exley (Keele University).

A proteomics project is in progress with Dr A. Grierson (University of Sheffield) to identify post-translational modifications of kinesin by the generation of mass maps and MS/MS sequence tags from 2D-gel digests.

Other structural analysis projects include the characterisation of lipopolysaccharides (Dr D. Devine, Oral Biology) and oligosaccharides (Prof. S.W. Homans).

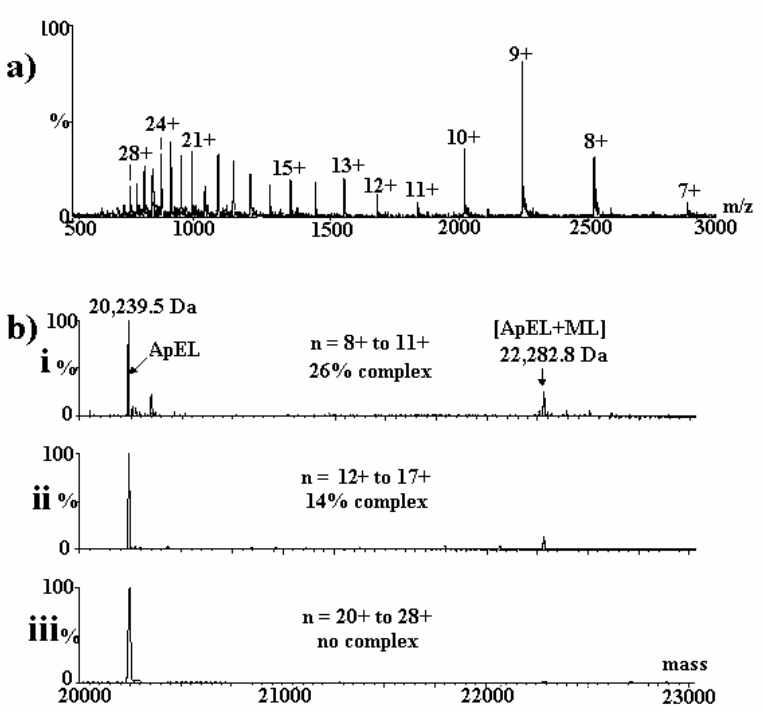


Fig. 1 shows the ESI-MS analysis of the apical domain protein (ApEL) of the chaperonin GroEL. The apical domain is acknowledged to be the site where polypeptides bind during folding to their native state. The m/z spectrum of the apical domain (a) shows three conformations: the folded (n = 8+ to 11+), the partially folded (n = 12+ to 17+) and the unfolded (n = 20+ to 28+) states. Non-covalent attachment of a peptide (ML) can be detected on the folded (b.i) and partially folded (b.ii) ApEL, but not on the unfolded conformation (b.iii).

Publications

Ashcroft, A.E., Brinker, A., Coyle, J.E., Weber, F., Kaiser, M., Moroder, L., Parsons, M.R., Jaeger, J., Hartl, U.F., Hayer-Hartl, M. & Radford, S.E. (2002) Structural plasticity and non-covalent substrate binding in the GroEL apical domain: a study using electrospray ionisation mass spectrometry and fluorescence binding studies. *J. Biol. Chem*., **277**, 33115-33126.

Ramoni, R., Vincent, F., Ashcroft, A.E., Accornero, P., Grolli, S., Valencia, C., Tegoni, M. & Cambillau, C. (2002) Control of domain swapping in bovine odorant binding protein. *Biochem. J.,* **365**, 739-748.

Goodall, J.J., Booth, V.K., Ashcroft, A.E. & Wharton, C.W. (2002) Hydrogen bonding in 2-aminobenzyl-alpha-chymotrypsin formed by acylation of the enzyme with isatoic anhydride: infra-red and mass spectrometric studies. *ChemBioChem*, **3**, 68-75.

Standeven, K.F., Ariens, R.A.S., Whitaker, P., Ashcroft, A.E., Weisel, J.W. & Grant, P.J. (2002) The effect of dimethylbiguanide on thrombin Activity, FXIII activation, fibrin polymerisation, and fibrin clot formation. *Diabetes*, **51**, 189-197.

Ashcroft, A.E., Venter, H., Keen, J.N., Henderson, P.J.F. & Herbert, R.B. (2002) Molecular dissection of membrane transport proteins: mass spectrometry and sequence determination of the galactose-H+ symport protein, GalP, of Escherichia coli and quantitative assay of the incorporation of [ring-2-^{13}C]histidine and $^{15}NH_3$. *Biochem. J*., **363**, 243-252.

Korchazhkina, O.V., Ashcroft, A.E., Kiss, T. & Exley, C. (2002) The degradation of A-beta 25-35 by the serine protease plasmin is inhibited by aluminium. *J. Alzheimer's Disease*, **4**, 357-367.

Treumann, A., Xidong, F., McDonnell, L., Derrick, P.J., Ashcroft, A.E., Chatterjee, D. & Homans S.W. (2002) 5-Methylthiopentose: a new substituent on lipoarabinomannan in Mycobacterium tuberculosis. *J. Mol. Biol*., **316**, 89-100.

Group Members

Antoni Borysic (with Prof. S.E. Radford),
Simona Francese (with Dr N. J. Stonehouse & Prof. P.G. Stockley),
Michelle Morgan (with Dr D. Devine)
Andrew Smith (with Profs. S.E. Radford & P.G.Stockley).

Funding

Financial support from the University of Leeds, the Wellcome Trust, BBSRC, the European Commission, Micromass UK Ltd and AstraZeneca is gratefully acknowledged.

Protein structure and aggregation studied by mass spectrometry

Antoni J. H. Borysik, Sarah L. Myers, Sheena E. Radford and Alison E. Ashcroft.

Characterising the unique three-dimensional structures and folding mechanisms of proteins are intriguing challenges to structural biologists. In particular, protein aggregation, which can arise as a result of protein misfolding and is thought to be a contributing factor to human disease, is of special interest.

Mass spectrometry (MS) is one of the important biophysical techniques being used in this field. The analysis of proteins by electrospray ionisation mass spectrometry produces a spectrum consisting of a characteristic envelope of multiply charged ions which can be used to identify co-existing protein conformers on account of their varying degrees of ionisation. Additionally, non-covalently bound protein and protein-ligand complexes can be sustained throughout the mass spectrometric analysis, thus providing the opportunity to monitor assembly mechanisms on-line and gain insights into assembly intermediates. Other mass spectrometric methods can be used to evaluate the position of protected regions within a protein's structure, for example using limited proteolysis and subsequent tandem mass spectrometric sequence analysis (MS/MS) of the resulting peptides, as well as hydrogen/deuterium exchange procedures.

β_2-microglobulin.

β_2-microglobulin (β_2m) is a small protein, 11,860 Da in mass and 99 amino acids in length with a 7-stranded β-sandwich structure (Fig. 1). The protein is found within serum at low concentrations. However, in patients with renal failure, the concentration of β_2m can increase up to 50 fold, and this can lead to the formation and deposition of amyloid fibrils, characteristically in the musculo-skeletal system. The mechanism by which β_2m forms amyloid fibrils *in vivo* is not known, although biophysical experiments are now beginning to cast light on the mechanism of β_2m amyloid-fibril formation *in vitro*.

Fig. 1. Ribbon diagram of the structure of native β_2-microglobulin

Using monomeric, recombinant β_2m, we have formed amyloid-like fibrils *in vitro* by incubating the protein at low pH. At pH 2.5 β_2m populates an acid-unfolded state, which is a precursor to fibril formation. Limited proteolysis with pepsin has been carried out on both the acid-unfolded monomer and the amyloid-like fibril to determine which, if any, of the protected sites correspond, and to explore whether dynamic/conformational changes between the monomeric and amyloid-like fibril state could be detected. The digest products were analysed by MS/MS to determine the amino acid sequence of the digested protein fragments, thus enabling their unequivocal identification. The data highlight differences in the structures of the monomers compared with the fibrils and also identify the proteolytic-resistant core of β_2m fibrils. Further experiments using limited proteolysis under various pH conditions with different enzymes are underway to give more clues into the key areas of β_2m sequence and structure which are involved in amyloid fibril formation.

The detection and resolution of co-existing conformational states of β_2m are also being investigated using electrospray ionisation MS (Figure 2). We are applying linear deconvolution methods to the m/z spectra to illustrate the relative populations of different conformers. This method offers a unique solution to resolving co-existing conformers and contributes to the characterisation of the fibrillogenic precursors in β_2m amyloid formation *in vitro*.

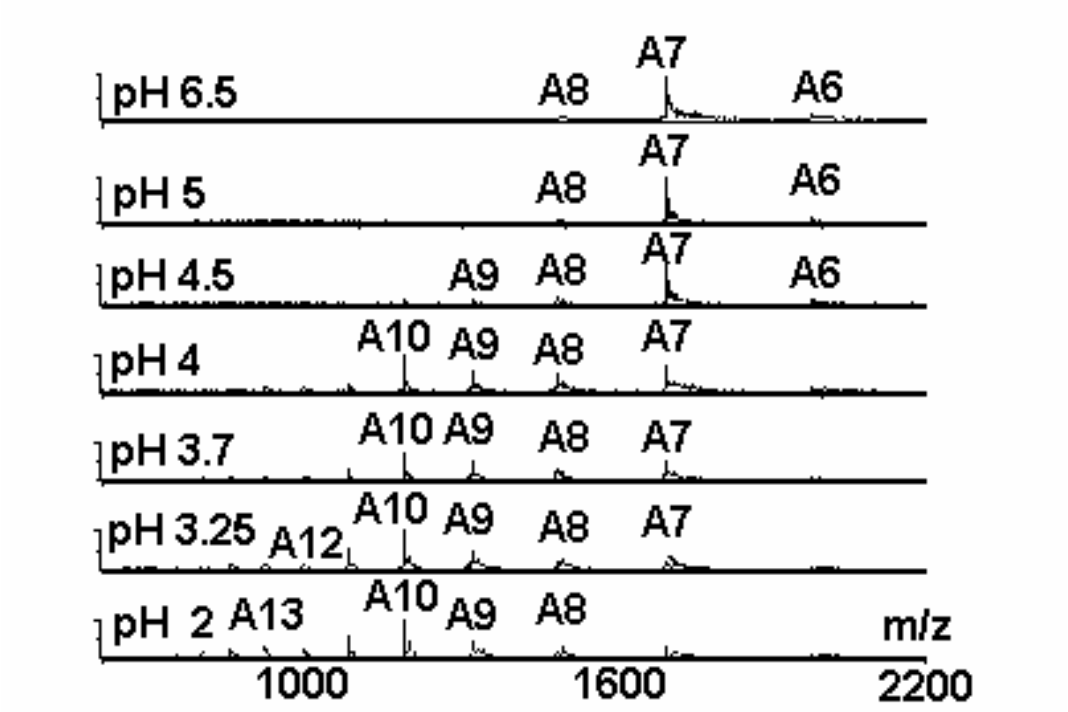


Fig. 2. The m/z spectra of β_2m obtained under varying pH conditions. The spectra indicate the presence of the folded conformer at pH 6.5, the unfolded conformer at pH 2 and intermediate stages of the protein unfolding where several conformers are present from pH 4.5 to 3.25.

Publications

Structural properties of an amyloid precursor of β_2-microglobulin McParland, V.J., Kalverda, A.P., Homans, S.W & Radford, S.E. (2002) *Nature Structural Biology*, **5**, 326-331

Funding

Financial support from the University of Leeds, the Wellcome Trust, BBSRC, Micromass UK Ltd and AstraZeneca is gratefully acknowledged.

Bacterial nucleoside transporters as models for mammalian transporters of chemotherapeutic nucleoside drugs

Hao Xie, Simon Patching, Steve Baldwin, Peter Henderson, Richard Herbert, Adrian Brough, and Simon Phillips

Introduction

In mammals, nucleoside transporters represent the route of cellular uptake for a variety of drugs, such as cytosine arabinoside and azidothymidine, used in the treatment of cancer and viral infections. Gaining a better understanding of the molecular mechanism of these transporters is therefore of clinical importance. Two relevant transporter families have so far been identified in mammals, the equilibrative nucleoside transporters (ENTs) and the sodium-linked concentrative nucleoside transporters (CNTs). We have recently succeeded, in collaboration with researchers at the University of Edmonton in Canada, in cloning and expressing all of the four ENT and three CNT genes present in the human genome, and are using cysteine-scanning mutagenesis, construction of chimaeric molecules and other approaches to probe their solute recognition properties. Unfortunately, the structural study of these eukaryote transporters has been hindered by the difficulty of expressing them at high levels. We have therefore turned to a bacterial homologue of the mammalian CNT proteins, the proton-linked nucleoside transporter NupC from *Escherichia coli*, as a more experimentally-amenable model for structural studies.

Membrane topology and overexpression of NupC

It has proven possible to express NupC in functional form at levels of up to 25% of the total membrane protein in *Escherichia coli*. This has enabled purification of the protein, bearing a Strep-tag II affinity tag, in amounts sufficient for structural analysis by circular dichroism spectroscopy, which has revealed that the protein is largely α-helical (***Fig. 1***).

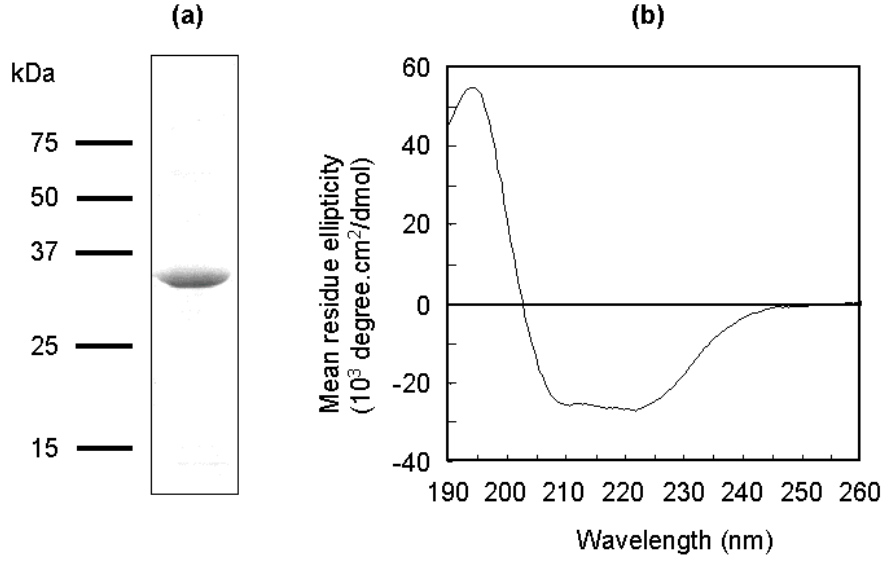


Fig. 1. (a) SDS-polyacrylamide gel of purified, Strep-tagged NupC. (b) Circular dichroism spectrum of the purified protein, revealing predominantly α-helical structure.

Analysis of the protein sequence by the hidden Markov model procedure of Sonnhammer and colleagues suggested that NupC possessed 10 transmembrane (TM) helices, with the N- and C-termini exposed at the periplasmic side of the membrane. We have recently obtained direct evidence for this topology by analysis of the susceptibility of mutants, bearing single cysteine residues, to modification by membrane impermeable, fluorescent maleimide reagents.

Cysteine scanning mutagenesis has also been successful in identifying residues critical for NupC function. For example, the highly conserved residue E149 in putative TM4 was found to be essential for activity, and possibly plays a role in proton conductance (***Fig. 2***). The similarly conserved residue E135 was found not to be essential for transport activity, but modification by methanethiosulphonate reagents of a cysteine residue introduced at this position inactivated the protein, suggesting that it lies on the substrate translocation pathway. Only membrane permeable reagents were effective in whole cells, confirming the cytoplasmic orientation of this residue (***Fig. 2***).

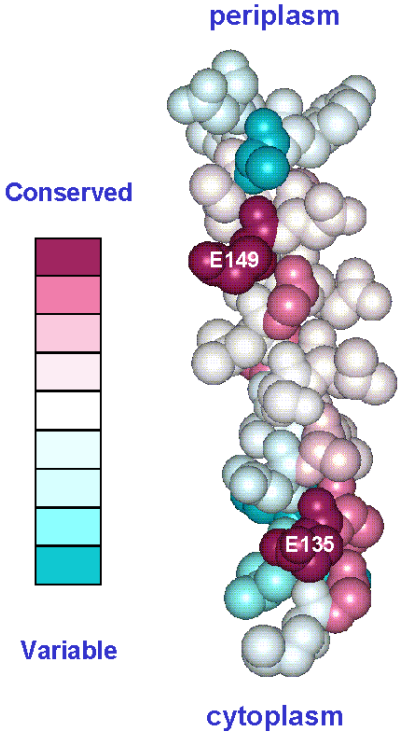


Fig. 2. Pattern of residue conservation in putative TM4 of NupC, determined from comparison with other members of the CNT family. The degree of conservation is colour coded. Residues E149 and E135 are almost invariant, and their function was tested by cysteine mutagenesis.

Structural analysis of NupC by NMR and 2-D crystallisation

Overexpression of NupC has allowed us directly to detect the binding of [^{13}C]-labelled uridine by solid state magic angle spinning NMR, opening the way to mapping the vicinity of the substrate-binding site using labelled protein. In parallel, we have successfully produced large 2-D crystals of the transporter and analysis of these by cryo-EM is currently underway in collaboration with Prof. Per Bullough's group at Sheffield. 3-D crystallization trials are about to be initiated at Leeds.

Collaborators

Maurice Gallagher, University of Edinburgh; Per Bullough, University of Sheffield; David Middleton, UMIST; Jim Young, Carol Cass; University of Alberta, Edmonton, Canada

Publications

Cabrita, M.A., Baldwin, S.A., Young, J.D. & Cass C.E. (2002) Molecular biology and regulation of nucleoside and nucleobase transporter proteins in eukaryotes and prokaryotes. *Biochem. Cell Biol.* **80**, 623-638.

Funding

We thank the MRC and BBSRC for grant support.

Structure and cell biology of a membrane trafficking complex – the exocyst

Sweta Srivastava, Hao Xie, Steve Baldwin, Simon Phillips, Steve Homans,
John Trinick, and Vas Ponnambalam

Function of the exocyst

The exocyst is a large complex of proteins required for polarised exocytosis in eukaryotic cells. It appears to function primarily as a "tether", directing secretory vesicles to specific sites on the plasma membrane, where their fusion is then brought about by SNAREs. The components of the exocyst were initially identified as products of *sec* genes in yeast, but a complex of homologous proteins was subsequently isolated from rat brain and the exocyst probably plays essential roles in the post-Golgi secretory pathways of all eukaryotes. For example, mice homozygous for a mutation in a gene encoding one of the components of the complex die at an early stage during embryogenesis. In polarised epithelial cells the complex is required for the delivery of secretory vesicles to the basolateral membrane and is localised to the tight junctions. Immunocytochemical studies of its distribution in cultured hippocampal neurons suggest that it is also involved in membrane addition at axonal and dendritic growth cones, and plays a key role in synaptogenesis.

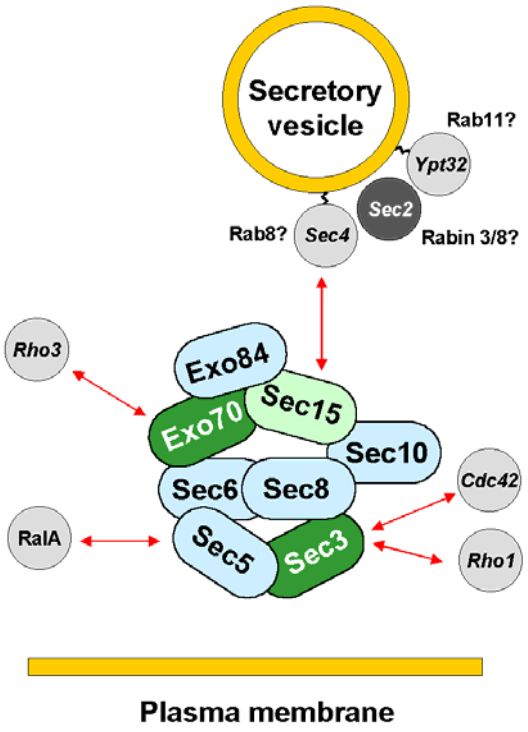


Fig. 1. Possible structure of the exocyst. The arrangement of subunits is that proposed for the mammalian complex. Subunits that may be involved in plasma membrane targeting are in dark green, while that implicated in interaction with secretory vesicles is in light green. Arrows indicate regulatory interactions with yeast (names italicised) and mammalian GTPases. Possible mammalian counterparts of the yeast Sec4p and Ypt32 GTPases and of the guanine nucleotide exchange factor Sec2p are also indicated.

Structure and regulation of the exocyst

The exocyst complex contains single copies of eight subunits (Sec3, Sec5, Sec6, Sec8, Sec10, Sec15, Exo70 and Exo84) that in mammals range in size from 75 to 111 kDa, giving a total molecular mass of 736 kDa. Information on the arrangement of the subunits within the complex has been obtained by yeast 2-hybrid analyses and other approaches, yielding the picture illustrated in Fig. 1. Such studies have identified Sec15 in yeast as the subunit responsible for association of the exocyst with secretory vesicles, via an association with the GTP-bound form of the rab GTPase Sec4. Similarly Sec3, and possibly also Exo70, have

been identified as spatial landmarks for polarised secretion in yeast, targeting the exocyst complex to specific domains of the plasma membrane. The other exocyst components link Sec3/Exo70 and Sec15 via a series of protein-protein interactions, resulting in the polarised targeting of secretory vesicles to appropriate sites of exocytosis. Vesicle targeting and trafficking appear to be regulated by small G-proteins, and the subunits responsible for interaction with a number of such proteins, such as the mammalian GTPase ralA, have recently been identified (Fig. 1).

Recombinant DNA approaches to studying exocyst structure and function

With the support of a grant from the BBSRC MASIF initiative, our long-term objective is to gain an understanding of the mechanism of the exocyst at the molecular level by determining the structures of its subunits and then of the entire complex at high resolution. As the first step, we have been expressing individual subunits/domains from the mouse exocyst complex in *Escherichia coli* in amounts suitable for detailed structural investigations. To check that the expressed proteins are correctly folded, their secondary structures are examined by circular dichroism and Fourier transform infrared spectroscopy. Functionality is also assessed by examining their ability to interact specifically with other components of the complex, or regulatory G-proteins, using surface plasmon resonance and other approaches. For example, we have recently shown that the N-terminal 100 residue region of Sec5 can be expressed and purified on a large scale as a properly-folded, largely β-sheet domain that interacts with the GTP-bound form, but not the GDP-bound form, of ralA.

In parallel with such studies, we are examining the function of individual subunits and their domains *in vivo* by expressing them as green fluorescent protein constructs in mammalian cells. The locations of the proteins, and the effects of co-expressed wild-type and mutant G-proteins is then assessed by fluorescence microscopy. Such studies have been greatly facilitated by access to a Deltavision deconvolution microscope, funded by the Wellcome Trust as part of the newly established Leeds Bioimaging Centre in the Astbury Building.

Future work

Having demonstrated the feasibility of using bacterial expression to produce correctly folded exocyst subunits in amounts suitable for structural analysis, we plan shortly to commence crystallisation trials to enable detailed structural analysis by X-ray diffraction. In addition, NMR spectroscopy will be employed to examine the structure of smaller protein domains. In parallel, we will isolate intact complex from mammalian brain for single particle cryo-EM analyses. In the longer term, we hope to combine information from EM, NMR and X-ray analyses to yield the structure of the entire complex at high resolution.

Publications

Ponnambalam, S. & Baldwin, S.A. (2003) Constitutive protein secretion from the trans-Golgi network to the plasma membrane. *Mol. Memb. Biol.* In Press.

Funding

We thank the BBSRC for grant support. SS is supported by an ORS award and a scholarship from the Faculty of Biological Sciences, University of Leeds.

Analytical centrifuge facility

Andy Baron and Peter Stockley

Introduction

The Centre has a Beckman XL-I analytical ultracentrifuge equipped with absorbance and interference optics, two rotors (4-place and 8-place), and velocity and equilibrium cells with a choice of quartz or sapphire windows. We employ a range of data analysis methods, enabling the determination of properties of macromolecules in free solution including mass, degree of asymmetry, species distribution, and association constants of interacting species.

Work done in 2002

The instrument was used in a wide range of applications, ranging from demonstrating that a 10 kDa protein was monomeric to quantifying RNA-mediated assembly of 28 kDa viral coat protein into 2.5 MDa capsid. The facility was used to study the effect of buffer composition on protein self association in order to optimise crystallisation conditions, to evaluate the effects of small molecules and polysaccharides on protein oligomerisation, and to address problems with protein purification by determining relative masses in solution of partially pure preparations. Sedimentation equilibrium was used to determine true molecular weights of proteins which gave anomalous results in gel permeation chromatography, and to measure affinity constants of associating systems.

Example

This example demonstrates the integration of sedimentation velocity and sedimentation equilibrium analysis. The subject of the study was MobC, a 33.4 kDa dimer, which is one of three proteins involved in the transfer of the mobilisable plasmid pC221 between *Staphylococcus aureus* cells. A 1.2 mg/ml solution of MobC in 50 mM Tris-HCl, 0.5M KCl (pH 8) was subjected to sedimentation velocity analysis at 55000 rpm for 5 hours at 10°C to confirm the degree of oligomerisation. A group of absorbance scans made at 275nm during the middle of the run was analysed by the g(s) method using the program DCDT+ by J. Philo (Fig 1a).

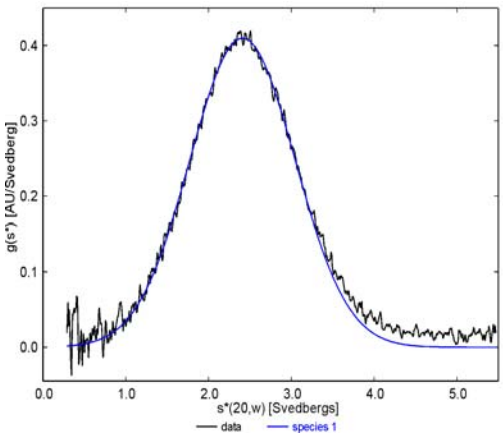


Fig 1a. Sedimentation velocity analysis

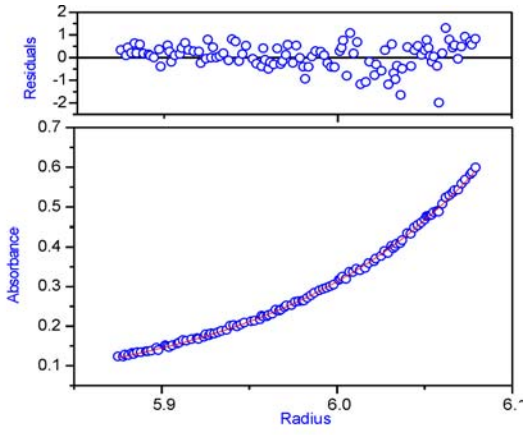


Fig 1b. Sedimentation equilibrium analysis

Parameters resulting from least-squares fitting were:

$s_{20,w}$	2.410	S
$D_{20,w}$	7.28×10^{-7} cm^2 s^{-1}	

A mass of 28.2 kDa was calculated from the ratio of s/D. This value is 14% smaller than dimer of MobC.

To check the true molecular weight of the protein, sedimentation equilibrium analysis was performed. Absorbance scans recorded from different protein concentrations and speeds were analysed globally. An example of one of the fitted scans is shown (Fig 1b.). Fitting to a single species model resulted in a mass of 33.7 kDa. As this value is slightly higher than the predicted MobC dimer mass the results were re-fitted to a dimer-tetramer model. This fit was slightly better than the single species model, and resulted in an equilibrium dissociation constant of 700 μM. The underestimate of mass from sedimentation velocity can now be explained by the reversible weak association of dimer and tetramer resulting in broadening of the sedimenting boundary.

The use of software that enables modelling of sedimentation velocity data to associating systems is currently being tried out, and developments for the coming year include installation of a second XL-I centrifuge as part of the move to the Wellcome Trust funded Biomolecular Interactions Centre.

Publications

Caryl, J.A. & Thomas, C.D. (2002) Initial events in small staphylococcal plasmid transfer. In, Plasmid Biology 2002: International symposium on molecular biology of bacterial plasmids. Frost, L. (ed). *Plasmid* **48**, 232-302.

Funding

This instrument was funded on the JREI scheme by HEFCE.

Directed evolution of aldolase

Gavin Williams, Jijun Hao, Chris Plummer, Adam Nelson and Alan Berry

Introduction

The *E.coli* Zn^{2+}-dependent fructose-1,6-bisphosphate (FBP) aldolase catalyses the reversible condensation of dihydroxyacetone phosphate (DHAP) and glyceraldehyde phosphate (G3P) to form FBP. The enzyme belongs to the $(\beta/\alpha)_8$-barrel family of proteins, a fold which is thought to comprise about 10% of all known protein structures. The structural partitioning of substrate binding residues and catalytic residues in the loops and barrel of $(\beta/\alpha)_8$-barrel proteins makes these enzymes particularly attractive to protein engineers.

We have successfully explored structure-function relationships in FBP-aldolase by using directed evolution to *i*) alter the reaction stereochemistry of the enzyme and *ii*) to improve the thermostability of aldolase.

Altering reaction stereochemistry

We have shown, for the first time that the stereochemistry of an enzyme-catalysed bond forming step can be changed by directed evolution. FBP-aldolase catalyses the formation of a new carbon-carbon bond with hydroxyl groups that are *trans* to each other. A related enzyme, TBP-aldolase, also catalyses the condensation of DHAP and G3P, but produces TBP in which the hydroxyl groups are *cis*. This discrimination arises because the enzymes are able to specifically recognize the *Si* and *Re* faces of the incoming G3P molecule. Using random mutagenesis and colorimetric screening, we evolved a TBP-aldolase into an effective FBP-aldolase. Following three generations of DNA shuffling, a variant was created which showed an 80-fold improvement in the specificity constant for FBP. Using ^{31}P-NMR (Fig. 1), we showed that the evolved aldolase produced FBP and TBP in a ratio of 4:1, compared to <1:>99 by the wild-type TBP-aldolase. This shows that new biocatalysts can be generated that use the same substrates as wild-type enzymes, but catalyse the production of products with opposite stereochemistry. X-ray crystallography is currently underway to understand how the evolved aldolase catalyses the formation of a different stereochemistry.

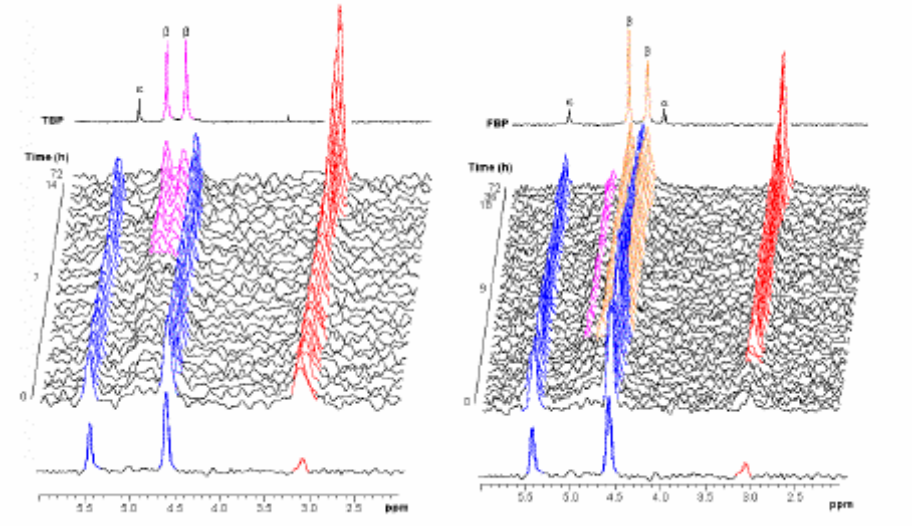


Fig. 1: Evolution of an FBP-aldolase from TBP-aldolase. ^{31}P-NMR spectroscopy shows that the wild-type TBP-aldolase synthesises TBP as a single product (purple) from DHAP and G3P (blue). The evolved aldolase synthesises a new product, FBP (orange) and a small amount of TBP, in a 4:1 ratio.

Evolving thermostable enzymes

The rules that govern the stability of protein folds are still largely unclear. Using the $(\beta/\alpha)_8$-barrel enzyme, FBP-aldolase as a scaffold, we have attempted to understand how protein thermostability can improved using iterative cycles of random mutagenesis and screening.

The FBP-aldolases from *E.coli* and *Edwardsiella ictaluri* are both completely inactivated after heat treatment at 55°C (Fig. 2). The genes encoding these two aldolases were randomly recombined by DNA shuffling and the subsequent mutant enzymes screened for activity after heat treatment at 55°C. Progeny shown to be more thermotolerant compared to the parents were pooled and used as the input for the next generation of shuffling and screening at a slightly higher temperature. Fig. 2 outlines the procedure followed during the stages of directed evolution. In this way, an evolved aldolase was generated which remained active after heat treatment at 64°C, whereas both wild-type aldolases were inactive after treatment at this temperature. The half-life of the mutant enzyme was 100-fold longer than the *E.coli* enzyme at 53°C. DNA sequencing showed that the evolved aldolase was a chimera between the *E.coli* and *Ed.ictaluri* enzymes, but contained sequence mostly originating from *Ed. ictaluri*. In addition, the evolved enzyme possessed an additional 19 amino acid mutations, all of which are located on the surface of the $(\beta/\alpha)_8$-barrel. The structural studies of the improved aldolase are underway using X-ray crystallography.

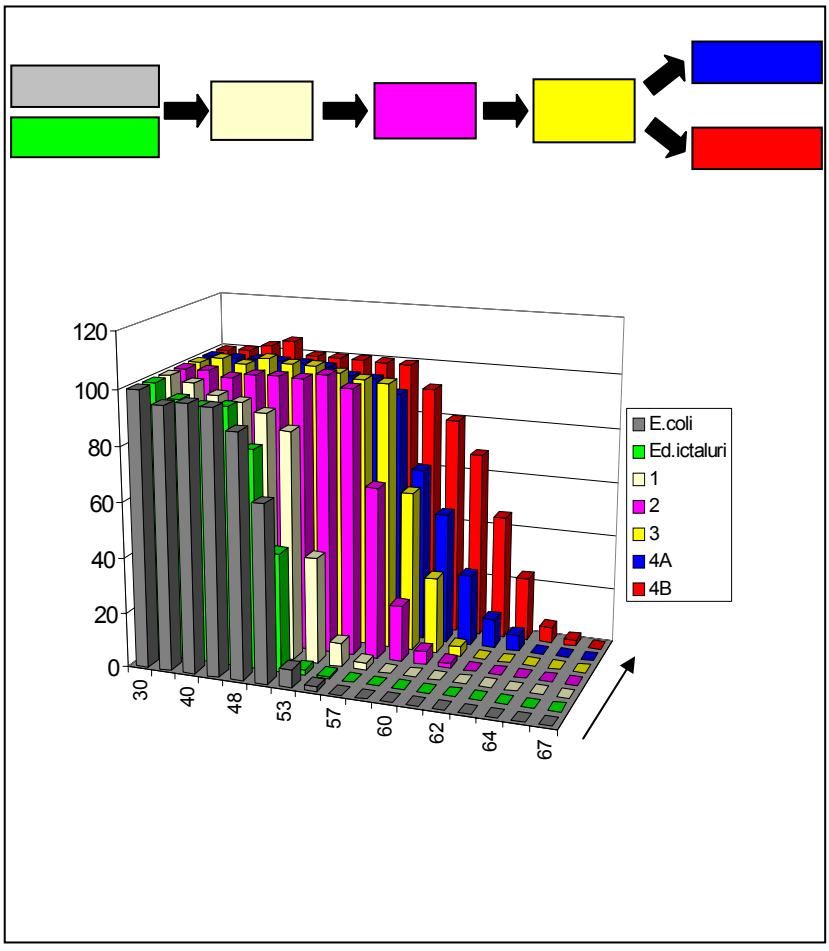


Fig. 2. The E.coli and Edwardsiella FBP-aldolase genes were subjected to DNA shuffling and variants with higher thermostabilities were selected. Further rounds of DNA shuffling, using the 1st generation as parent were carried out to ultimately produce 4th generation FBP-aldolases with significantly increased thermostability.

Other projects using directed evolution

Sialic acid aldolase is another $(\beta/\alpha)_8$-barrel enzyme. It catalyses the formation of sialic acid, an important sugar involved in numerous cell signaling pathways and host-pathogen interaction. We are evolving novel sialic acid aldolase enzymes with new substrate specificities that can be used in the synthesis of sialic acid analogues with medical and pharmaceutical importance. We aim to produce aldolases with both specific activities to new substrates, and also aldolases with broader use.

In addition, we have commenced directed evolution of dihydrodipicolinate synthase. Improved variants with novel substrate specificities will be useful in the synthesis of drugs that can be used to treat HIV.

Publications

Williams, G.J., Domann, S., Nelson, A. & Berry, A. (2003) Modifying the stereochemical course of an enzyme catalysed reaction by directed evolution. *Proc. Natl. Acad. Sciences. USA*. **100**, 3143-3148.

Funding

This work was funded by the BBSRC and The Wellcome Trust.

Partial sequential resonance assignment of the 78kDa Class II fructose-1,6-bisphosphate aldolase and investigation of its dynamics

Christine Hilcenko, Arnout P. Kalverda, Steve W. Homans and Alan Berry.

Enzyme motion and activity

Multiple conformational changes, ranging from loop movements to repositioning of a specific residue in the active site, are often a key part of enzymatic mechanisms, but the role of these conformational changes in the catalytic process is not fully understood. The structure of the enzyme can be optimized for one type of reaction by subtle changes, the binding of the substrate can require a reorganisation of the enzyme overall structure or of the active site, and the substrates can be orientated very precisely to allow the reaction to occur. Multi-dimensional NMR spectroscopy is a very powerful technique to study the flexibility of proteins and the plasticity of their actives sites. It will allow us to understand the relationship between enzyme motion and catalysis and the impact of dynamics on activity.

The challenge with a 78kDa protein

The full or partial sequential assignment of the protein backbone resonances is usually a prerequisite for any analysis of the structural and dynamic characteristics of biomolecules. Resolution and sensitivity have been always a problem for large proteins due to the complexity and high degree of crowding of the NMR spectra, slower tumbling times and shorter transverse relaxation times (T_2). Nevertheless, the combination of TROSY-based NMR techniques introduced by Pervushin and co-workers with the continuing development of multidimensional NMR spectroscopy has considerably increased the range of proteins that can be studied using NMR spectroscopy. Furthermore, the introduction of deuteration has improved both resolution and sensitivity and allowed an increase in the size of proteins which can be investigated.

Sequential resonance assignment strategy

To carry out the sequential assignment of backbone resonances of the Class II fructose-1,6-bisphosphate aldolase, triple labeled [^{2}H, ^{15}N, ^{13}C]-aldolase was prepared and purified. The usual complementary 3D TROSY-HNCA and TROSY-HN(CO)CA experiments were recorded, and 3D TROSY-NOESY experiments as well, at 600 MHz and 900 MHz (Varian NMR Instruments, Oxford, UK). To further aid with assignment and confirm the actual assignment, specifically labeled samples, such as [^{15}N, ^{13}C]-His-FBP-aldolase, [^{2}H, ^{15}N]-Tyr-FBP-aldolase and [^{2}H, ^{15}N]-Glu-FBP-aldolase, were prepared.

Backbone relaxation experiments

Differences in the crystal structures of the FBP-aldolase in the absence and in the presence of a substrate analogue have highlighted the importance of dynamics in the mechanism of the enzyme. Comparison of the 2D [^{1}H-^{15}N]-TROSY and 3D HNCA-TROSY experiments in the absence and presence of dihydroxyacetone phosphate (DHAP) has already suggested the importance of motion in the enzyme mechanism. Thus the dynamics of the protein have been investigated in the absence of DHAP via ^{15}N T_1 and T_2 longitudinal and transverse relaxation times respectively (at 600 MHz and 900 MHz), and heteronuclear [^{1}H-^{15}N]-NOE measurements (at 600MHz). The results of these experiments show a variety of dynamic behaviour in the protein. The catalytic residue Glu-182 is found in a loop of the protein between β5 and α7. Resonances for Gly-179 and Gly-180 (also in this loop) have been identified and their T_1 and T_2 values demonstrate that this region of the protein is very mobile in the absence of DHAP. The nuclear spin relaxation rates T_1 and T_2 have been measured in the presence of DHAP and have been compared with those in the absence of DHAP to provide information on the role of mobile sections of the protein in the catalytic cycle.

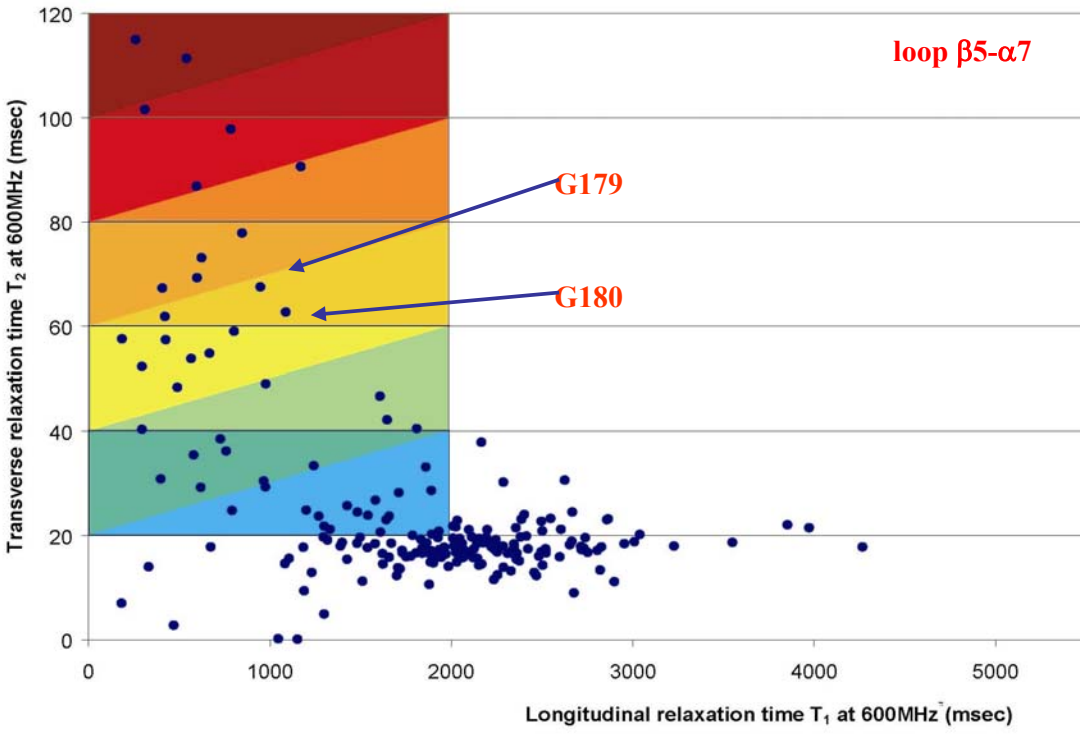


Fig. Comparison of the longitudinal relaxation time T_1 and the transverse relaxation time T_2 at 600 MHz (in the absence of DHAP).

Future work

To complete the sequential assignment of the protein backbone resonance, the perdeuterated labeled aldolase will be unfolded in H_2O using urea to allow all remaining deuterons to be rapidly exchanged by protons.

Collaborators

Dr.Eriks Kupce, Varian NMR Instruments, Oxford, UK

Funding

The University of Leeds, The School Research Fund, BBSRC and the Wellcome Trust are gratefully acknowledged.

Structure/function studies of glucagon-like peptide-1 and exendin 4.

Suleiman Al-Sabah and Dan Donnelly

The GLP-1 receptor and its peptide ligands.

Glucagon-like peptide 1 (7-36) amide (GLP-1) is a 30 amino acid peptide produced in intestinal L cells and released into the bloodstream in response to food intake. It is a potent 'incretin', in that it increases glucose-dependent secretion of insulin by pancreatic β-cells. GLP-1 has also been shown to stimulate pro-insulin gene transcription in the pancreatic β-cells, slow down gastric emptying time and reduce food intake. As a result of these actions, GLP-1 has received much attention as a possible therapeutic agent in the treatment of type II diabetes and obesity. Unlike the sulphonylureas, the actions of GLP-1 are glucose-dependent and do not produce hypoglycemic side effects. However, GLP-1 is rapidly degraded *in vivo* by dipeptidyl peptidase IV (DPP IV), and has a half-life of 2 min. Hence, future therapeutic agents would have to be physiologically more stable than GLP-1.

Exendin-4, found in the venom of the Gila Monster (*Heloderma suspetum*), is also an agonist for the GLP-1 receptor (GLP-1R). It is resistant to DPP IV digestion and, unlike GLP-1, it can be truncated by 8 amino acid residues at its N-terminus without losing receptor affinity. However, the loss of the first 2-8 amino acid residues results in the generation of antagonists.

The N-terminal region of GLP-1 and exendin-4 are almost identical, a significant difference being the second amino acid residue, alanine in GLP-1 and glycine in exendin-4, which gives exendin-4 its resistance to DPP IV digestion. Exendin-4 has an extra 9 amino acid residues at its C-terminus which have been shown to form a 'Trp-cage' by NMR. NMR analysis of exendin-4 also shows that the central region (amino acid residues 10-30) is helical in structure. Interestingly GLP-1 and Exendin 4 only share 8 amino acid residues in this region but since they lie on the same face of the α-helix, we postulate that this face of the helix interacts with the receptor.

The GLP-1 receptor (GLP-1R) has been cloned and is a member of the 'family B' G protein-coupled receptors (GPCRs). Other members of this family include receptors for glucagon, calcitonin, glucose-dependent insulinotropic polypeptide and vasoactive intestinal peptide. It is known that the large amino terminal domain that characterizes the 'family B' GPCRs plays a key role in ligand binding. However the amino terminus is not entirely sufficient to bind the ligand and regions in the exctracellular loops and/or transmembrane helices are also believed to provide additional interactions.

Which regions of exendin-4 contribute to its unique properties?

The aim of this study was to build on the binding site model previously proposed by our laboratory by identifying the regions of exendin-4 that contribute to its ability, relative to GLP-1, to maintain high affinity binding following truncation of its N-terminal residues. This, we hope, will lead us to design more potent GLP-1R agonists and antagonists.

For the purpose of this study, radioligand binding assays were carried out on the wild type rat GLP-1 receptor (rGLP-1R) and a truncated form of the receptor consisting of the N-terminal domain and the first transmembrane helix (rNT-TM1). These receptors were expressed in HEK-293 cells, and binding assays were carried out on membrane preparations. Competition binding assays were carried out using ^{125}I-exendin-4(9-39) and six unlabelled peptide ligands (see Fig.

1A). The six peptides were chosen in order to determine the relative contribution to affinity for the N-terminal region, central helical region and, in the case of exendin-4, the C-terminal "Trp-cage". These peptides are drawn schematically in Fig. 1B.

Peptide	**rGLP-1R**	**rNT-TM1**
GLP-1 (1-30)	7.8 ± 0.10	6.1 ± 0.11
GLP-1 (9-30)	6.4 ± 0.03	6.3 ± 0.08
exendin-4 (1-39)	8.5 ± 0.12	7.9 ± 0.13
exendin-4 (9-39)	8.1 ± 0.07	7.9 ± 0.11
exendin-4 (1-30)	7.1 ± 0.11	6.5 ± 0.11
exendin-4 (9-30)	6.7 ± 0.06	6.8 ± 0.13

Table 1. pIC_{50} values (M) from competition binding assays using radiolabelled anatagonist ^{125}I-Exendin (9-39). The data represent the mean and S.D. for three independent experiments.

As expected, N-terminal truncation of the GLP-1 peptide resulted in a large loss in affinity (25-fold) while the equivalent truncation of exendin-4 to exendin-4 (9-39) resulted in only a small reduction (2.5-fold). The latter observation can also be confirmed by the 2.5-fold reduction in affinity between exendin-4 (1-30) and exendin-4 (9-30). The reduced contribution to affinity by the N-terminal region of exendin-4 is due to the substitution of Ala-2 for Gly. Although this reduces affinity, it actually enhances its resistance to DPP IV and hence actually increases its potency *in vivo*. However, if the N-terminal region of exendin-4 contributes less to affinity than that of GLP-1, how is it that the affinity of exendin-4 is higher than that of GLP-1?

Firstly, it can be seen that the affinity of the core helical region (9-30) of exendin-4 contributes more to affinity than that of GLP-1. In addition, the C-terminal region of exendin-4 enhances its affinity further. The removal of the C-terminal region of exendin-4, yielding exendin-4 (1-30), resulted in a 25-fold reduction in its affinity. This is also confirmed by removing the C-terminal region from exendin-4 (9-39) to yield exendin-4 (9-30). Hence, we can conclude that while the affinity of the core helical region of exendin-4 is only enhanced 2.5 fold by its N-terminus, the addition of the C-terminal region enhances its affinity a further 25-fold. Hence this peptide maintains high affinity as well as DPP IV resistance.

The analysis of the rNT-TM1 receptor highlights the receptor domains to which the three regions of the peptides bind. The removal of the N-terminal regions of the peptides results in no loss of affinity at rNT-TM1, suggesting that this receptor domain is not involved in binding the N-termini of the peptides. This confirms our earlier studies. However, the central helical regions of the peptides maintain affinity for the N-terminal domain of the receptor, suggesting that they do not require sites on the transmembrane and extracellular loop regions of the receptor. The C-terminal "Trp cage" of exendin-4 is essential for high affinity binding to both the full length and truncated receptors. Hence, this region clearly binds to the N-terminal domain of the receptor.

A model for peptide:receptor binding.

These studies provide further evidence for our published model of how GLP-1R binds peptides. In this model, we defined three interactions "N", "H" and "Ex". Fig. 1C shows which of the three interactions are present with each peptide : receptor combination and it can be seen that this is entirely compatible with the binding data shown above. In previous and on-going work, we have identified several specific residues involved in the "N" interaction. In the work described here, we have identified the major components of the "Ex" interaction (the interaction that enhances exendin-4 affinity over that of GLP-1 and removes the dependence of high affinity

upon the N-terminal region) as being the small additional contribution from the helical core domain and a much larger contribution from the C-terminal "Trp Cage".

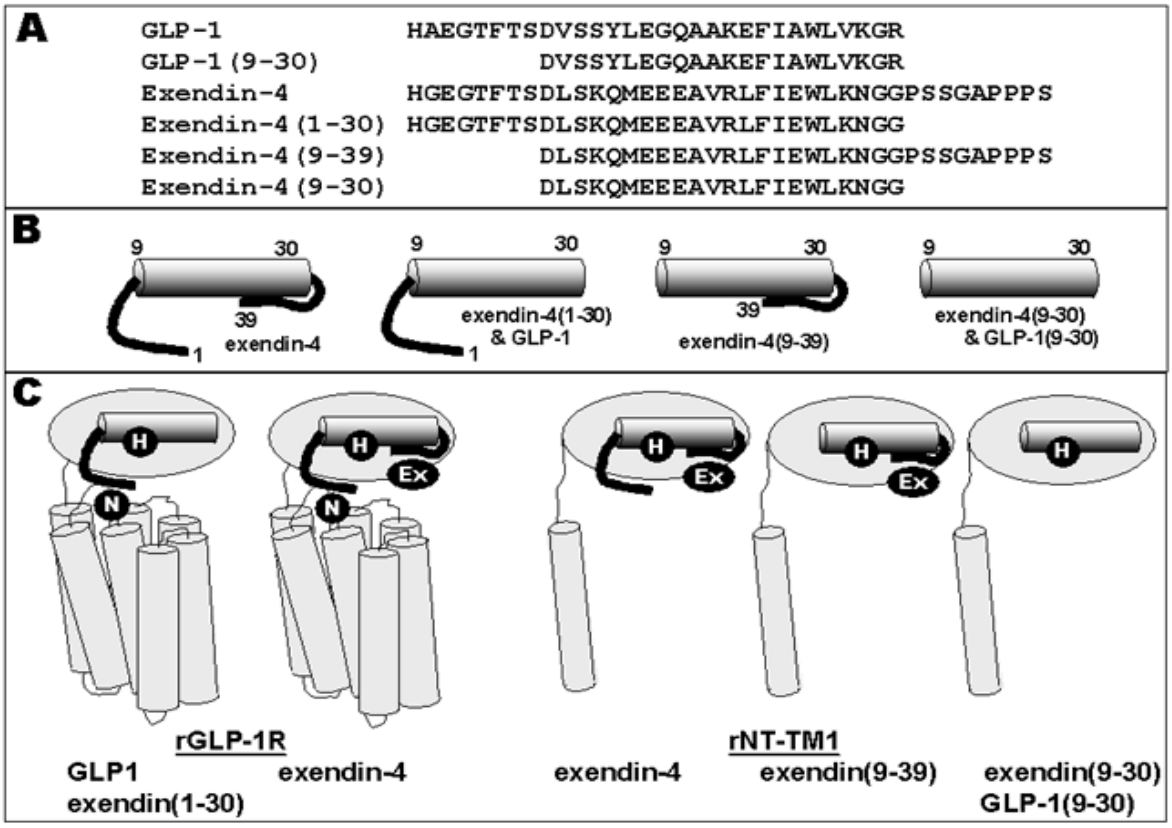


Fig. 1. A, the sequence of the six peptides used in this study. **B**, a schematic diagram of the classes of peptide used. **C**, the interactions possible between each peptide class and the two receptor constructs used.

Publications

López de Maturana, R. and Donnelly, D. (2002) The glucagon-like peptide-1 receptor binding site for the N-terminus of GLP-1 requires polarity at Asp198 rather than negative charge. *FEBS Letters* **531**, 572-.

López de Maturana, R., Willshaw A., Kuntzsch A., Rudolph R. and Donnelly, D. (2003) The isolated N-terminal domain of the GLP-1 receptor binds exendin peptides with much higher affinity than GLP-1. *J. Biol. Chem.* **278**, in press.

Ackowledgements

We thank BBSRC and Diabetes UK for their support.

Structural and functional studies on the Hepatitis C Virus non-structural NS5A protein.

Andrew Macdonald, Andrew Street, Katherine Crowder, Virginie Cazeaux, Chris McCormick, David Rowlands and Mark Harris

Hepatitis C virus (HCV) is an increasingly important cause of liver disease. The virus has a single stranded positive sense RNA genome of 9.5Kb that contains a long open reading frame encoding a single polyprotein of 3000 amino acids. This is cleaved into 10 individual polypeptides by a combination of host cell and virus specific proteases (see Fig. 1). The molecular mechanisms of pathogenesis remain to be elucidated. To this end my laboratory is interested in the potential for the non-structural proteins (expressed from the 3' end of the genome and designated non-structural as they do not form part of the viral particle) to interfere with host cell metabolism and signal transduction pathways.

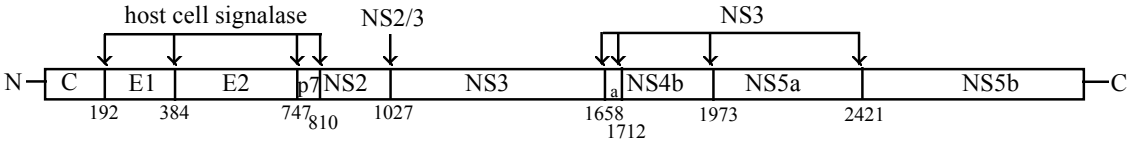


Fig.1: Polyprotein cleavage strategy of HCV

We have shown that two closely spaced poly-proline motifs in NS5A interact with the SH3 domains of members of the Src family of protein tyrosine kinases. Additionally NS5A interacts with the SH3 domain of the p85 regulatory subunit of phosphatidylinositol 3-kinase (PI3K) via a novel SH3-binding motif. These interactions have also been investigated *in vivo*, both in transiently transfected cells and cells containing autonomously replicating HCV derived subgenomic RNA's encoding the non-structural proteins (replicons). NS5A modulates the activity of both Src kinases and PI3K. We are currently using phage display and *in vitro* and *in vivo* binding assays to more precisely understand these interactions. Current work is also focussed on identifying the functional consequences of these interactions. In particular we are using the BacMAM system - baculovirus vectors with mammalian specific promoters that are able to efficiently enter hepatic cells and drive expression of foreign genes – to express NS5A both alone and in the context of the complete HCV genome in hepatocyte derived cell lines.

Publications:

McCormick, C.J., Rowlands, D.J. and Harris, M. (2002) Efficient delivery and regulable expression of hepatitis C virus full length and mini-genome constructs in hepatocyte-derived cell lines using baculovirus vectors. *Journal of General Virology* **83**, 383-394.

Gao, R., McCormick, CJ., Arthur MJP., Ruddell, R., Oakley, F., Smart, DE., Murphy, FR., Harris, MPG., and Mann, DA. (2002) High efficiency gene transfer into cultured primary rat and human hepatic stellate cells using baculovirus vectors. *Liver*, **22**, 15-22.

Collaborators:

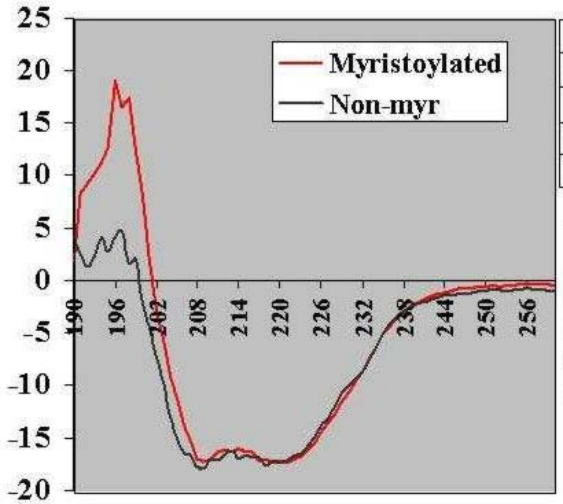
Derek Mann, University of Southampton
Kalle Saksela, University of Tampere, Finland

Funding: BBSRC, MRC, Wellcome Trust.

Structural and functional studies on the HIV-1 Nef protein.

Caitriona Dennis, Gemma Dixon, Matthew Bentham, Sabine Mazaleyrat,
Joachim Jäger and Mark Harris.

HIV-1 Nef is a 205 amino acid N-terminally myristoylated protein that plays a critical role in viral pathogenesis. Myristoylation is an eukaryotic specific co-translational modification that is catalysed by a ribosomal associated enzyme - N-myristoyltransferase (NMT). Several partial Nef structures have been determined by NMR and X-ray crystallography e.g. a C-terminal core domain of Nef (amino acids 71-206) lacking a disordered loop (residues 156-173) - but as yet there is no information about the structure of the full-length myristoylated form. To generate large amounts of myristoylated Nef for structural studies, we have co-expressed C-terminally 6-His tagged Nef in *E.coli* with human NMT. A range of biophysical and biochemical techniques have been used to determine the effects of myristoylation on the physical properties of Nef, by comparison with the corresponding non-myristoylated form (expressed in the absence of NMT). For example, circular dichroism analysis (Fig. 1) revealed that myristoylated Nef contained more α-helical content than the non-myristoylated form, consistent with a structural role of myristoylation. We are also establishing sensitive *in vivo* and *in vitro* membrane binding assays to analyse the mechanism of interaction between Nef and cellular membranes - by analogy to other cellular myristoyl-proteins this is likely to involve not only myristate but additional interactions, e.g. electrostatic interactions between basic amino acids and acid phospholipids.



	Myr-Nef	Non-myr-Nef
α-helix	69%	60%
β-sheet	15%	16%
Turns	5%	8%
Disordered	11%	19%

Fig. 1 Circular dichroism analysis of purified myristoylated and non-myristolated Nef. Samples were analysed at 298K on a Jasco J710 spectrometer. Secondary structure precentages were calculated using the programme CDSSTR.

One of the functions of Nef is to down-modulate the cell surface glycoprotein CD4 (which also functions as the viral receptor). This involves a direct physical interaction between myristoylated Nef and the cytoplasmic domain of CD4. We have developed novel *in vitro* assays to analyse the interaction between Nef and CD4. We have established a bank of cell lines expressing a panel of CD4 mutants that are currently being used to investigate the *in vivo* interaction between Nef and CD4. The long term goal of this project is to develop high throughput screening assays that can be used for drug screening.

Collaborators:

EU Framework Five 'Targeting Nef' (URL: http://www.uta.fi/~ltkasa/eu/index.html)

Funding

BBSRC, MRC, EU Framework Five

The transporter associated with antigen processing (TAP)

Eric Hewitt

Introduction

Major histocompatibility complex (MHC) class I molecules are expressed on the cell surface of all nucleated cells and present peptides derived from intracellular proteins to cytotoxic T lymphocytes (CTLs). CTLs monitor MHC class I molecules for peptides derived from viral proteins and kill infected cells. Peptides presented by MHC class I molecules are generated predominantly in the cytoplasm by the proteasome complex. The transporter associated with antigen processing (TAP) translocates these peptides into the endoplasmic reticulum (ER) lumen for loading onto nascent MHC class I molecules. Translocation of peptides into the ER by TAP is a key step in MHC class I assembly because only peptide loaded MHC class I molecules are transported to the cell surface.

TAP is a member of the ATP binding cassette (ABC) transporter family whose members translocate molecules across membranes utilising energy from ATP hydrolysis. ABC transporters have a modular structure comprising two nucleotide-binding domains (NBD) and two polytopic membrane domains. TAP is encoded in two halves. The two subunits TAP1 and TAP2 combine to form a heterodimer and each has an N-terminal membrane domain and a C-terminal cytoplasmic NBD (Fig. 1). The membrane domains of TAP form the peptide-binding site and presumably a pore through which peptides are translocated. The highly conserved NBDs of ABC transporters couple ATP hydrolysis to substrate translocation by the membrane domains. Indeed, the TAP1 NBD, shares significant structural homology with bacterial ABC transporter NBDs. They all have a characteristic two lobed fold. Lobe I contains the Walker A and B motifs that form the core ATP binding site, lobe II is perpendicular to lobe I and contains the signature motif (with the consensus sequence LSGGQ) within an α–helical subdomain.

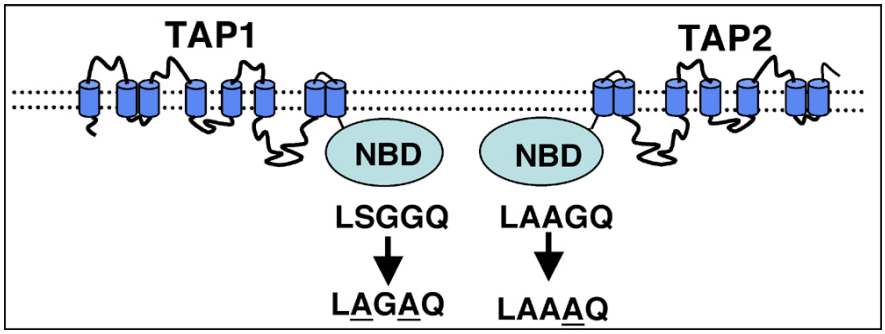


Fig. 1. Schematic representation of the mutations introduced into the TAP1 and TAP2 signature motifs.

Peptide binding by TAP stimulates ATP binding and hydrolysis, which presumably energises translocation. Yet for ABC transporters little is known about how substrate binding by the membrane domains is signalled to the NBDs, and likewise how ATP hydrolysis by the NBDs is coupled to substrate translocation. However, recent studies suggest a key role for the signature motif. This is illustrated by structure of the *Methanococcus jannaschii* ABC transporter NBD MJ0796. The MJ0796 NBD dimerises in the presence of ATP sandwiching two ATPs between the two subunits. Both signature motifs are located at the dimer interface with their serine and the second glycine hydrogen bonding with the γ-phosphate of ATP bound by the opposing subunit. ATP binding induces a displacement of the signature motif and rotation of the NBD α-helical subdomain. An attractive interpretation is that the signature motif functions as a γ-phosphate sensor that transmits conformational signals upon ATP binding and hydrolysis to the membrane domains.

Investigation of the role of the TAP signature motifs in peptide translocation

To understand how signature motifs participate in the transport cycle of ABC transporters, I have begun to examine their role in peptide translocation by TAP. TAP1 has the canonical signature motif sequence LSGGQ, whereas the TAP2 motif with the sequence LAAGQ is unusual because it has no serine. Based upon the MJ0796 NBD structure, mutations were designed that would be predicted to interfere with the interactions between the signature motif and the ATP γ-phosphate. To this end the serine and second glycine in the TAP1 signature motif and the glycine in the TAP2 signature motif were substituted with alanine (Fig. 1). These constructs were transfected into the TAP negative T2 cell line, either in combination, or with the complementary wild type subunit, and stable cell lines selected. The mutations had no gross effect upon TAP structure as both subunits combined to form the TAP heterodimer. Peptide and ATP binding were also unaffected, but peptide translocation was abolished when either TAP1 or TAP2 were mutated (Fig. 2). This not only demonstrates that the signature motifs are necessary for TAP activity, but that they are required after peptide binding to facilitate peptide translocation. Further investigation is required to determine their precise role. In particular it is of interest to determine if the signature motifs are involved after peptide binding to stimulate ATP binding and hydrolysis.

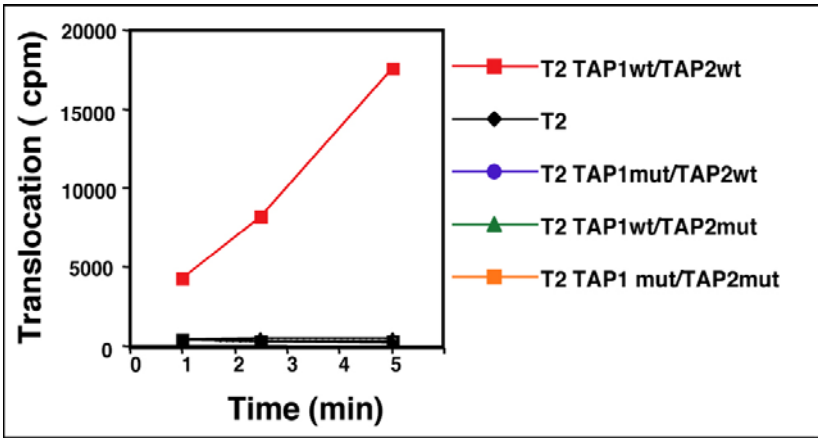


Fig. 2. Peptide translocation is impaired in the signature motif mutant TAP molecules. **S**treptolysin-O permeabilised T2 cells expressing the wild type and mutant TAP heterodimers were incubated at 37°C with a radioiodinated peptide whose sequence includes an N-linked glycosylation consensus sequence. Translocation into the ER was quantified by the recovery of glycosylated peptides on concanavalin-A sepharose and counting on a γamma-counter. Mutant (mut), wild type (wt).

Publications

Hewitt, E.W. and Lehner, P.J. (2003). The ABC transporter signature motif is required for peptide translocation but not binding by TAP. *Eur. J. Immunol*. **In press**.

van Endert, P.M., Saveanu, L, Hewitt, E.W. and Lehner, P.J. (2002). Powering the peptide pump: TAP crosstalk with energetic nucleotides. *Trends in Biochemical Sci*. **27**, 454-461.

Lehner, P.J. and Hewitt, E.W. (2002). Transporter associated with antigen presentation. Encyclopaedia of Molecular Medicine. Creighton, T.E. (ed) Wiley and Sons, New York.

Hewitt, E.W., Sen Gupta, S., and Lehner, P.J. (2001). The human cytomegalovirus gene product US6 inhibits ATP binding by TAP. *EMBO J*. **20**, 387-396.

Karttunen, J., Lehner, P.J., Sen Gupta, S. Hewitt, E.W., and Cresswell, P. (2001). Distinct functions of the nucleotide binding domains of TAP1 and TAP2. *Proc. Natl. Acad. Sci. USA*. **98**, 7431-7436.

Acknowledgements

This work was performed at Cambridge University in Paul Lehner's Laboratory.

Derivation of per-residue thermodynamic parameters for ligand-protein interactions

Antonio Hernández-Daranas, Arnout Kalverda, Hiroki Shimizu and Steven W. Homans.

Introduction

The Human Genome Project is providing a wealth of information on nucleic acid and protein sequences. However, if this information is to be of value for rational drug design, it is necessary to obtain a deeper understanding of the molecular basis of ligand-protein interactions. The key to understanding the affinity of a ligand for its receptor lies in the dynamics and thermodynamics of the association rather than a simple static picture.

With isothermal titration calorimetry (ITC), it is possible to obtain the global thermodynamic parameters governing a biomolecular association. However, from the point of view of ligand optimization, it would be of immeasurable benefit to obtain these thermodynamic parameters on a per-residue basis. Recently, new methodologies have been described by which these thermodynamic parameters can be derived on a per-residue basis from Nuclear Magnetic Resonance (NMR) relaxation data, thus offering a means by which the thermodynamics of the interactions can be characterized at a level of detail that has until now not been possible. Our aim is to use both techniques to rationalize differences in the thermodynamics of binding in structural terms.

NMR studies of ABP

The chosen system is the L-arabinose-binding protein (ABP), a 33.5kDa protein derived from *E. coli* that binds L-arabinose and D-galactose with similar affinities in the μM range. This system has been selected because the ligands are simple monosaccharides and excellent crystal structures are available for ABP in complex with some monossacharides.

Crucial to the success of this project is the availability of the resonance assignments of ^{1}H, ^{13}C and ^{15}N signals. We have finished the NMR assignment of the ABP-galactose complex using a combination of three independent 3D triple resonance experiments [HNCA/HN(CO)CA, HNCO/HN(CA)CO, HN(CA)CB/HN(COCA)CB] in their TROSY versions as well as the ^{15}N-HSQC-NOESY.

NMR relaxation experiments to measure T1, T2 and NOE for the backbone amides have been performed for ABP in its 'apo' form, and in complex with D-galactose, D-fucose and L-arabinose at two different magnetic fields. From the previous experiments the order parameter for each residue can be calculated and therefore the entropic contribution of each residue to the binding process can be estimated.

Finally, the thermodynamic parameters estimated by NMR will be compared with the global parameters obtained from the calorimetric measurements.

ITC studies of ABP

In addition to the 'natural' ligands (L-arabinose and D-galactose) we are using all the possible deoxy-analogues of D-galactose to delineate the precise thermodynamic contribution to binding of the individual hydroxyl groups. The binding affinities of this panel of deoxy analogues are

being determined by isothermal titration calorimetry, which additionally provides the ΔG^o, ΔH^o and ΔS^o of binding for each analogue. These parameters will be examined in order to decompose the 'global' thermodynamic parameters into contributions on a per-residue basis.

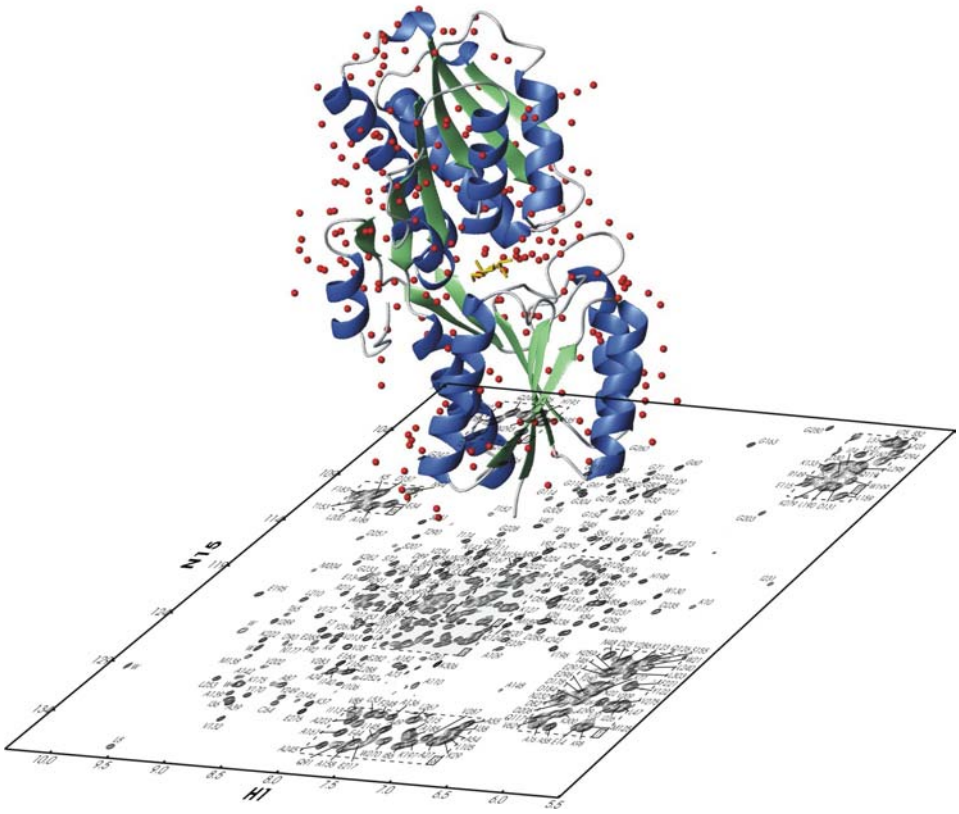


Fig. 1. Three dimensional structure of the ABP-D-galactose complex and the assigned ^{15}N-HSQC spectrum of it recorded at 600 MHz.

Funding
A. Hernández-Daranas is grateful to FP5 for a Marie Curie research fellowship.

Dissecting the cholera toxin-ganglioside GM1 interaction by isothermal titration calorimetry

Bruce Turnbull and Steve Homans.

Introduction

Cholera is a severe diarrhoeal disease that affects more than 130,000 people annually and is lethal in over 3% of cases. A further 6 million people per annum suffer from the less severe traveller's diarrhoea, largely on trips to southern Europe and developing countries. The causative agents of these two debilitating diseases are cholera toxin (CT) and heat-labile enterotoxin (LT) released by *Vibrio cholerae* and *Escherichia coli* bacteria, respectively. These two protein toxins have an AB_5-type multimeric structure, with essentially identical A-subunits and share 80% sequence identity in their B-subunits. The pentameric B-subunit (5 × 11.8 kDa) is a carbohydrate-binding protein that specifically recognises the oligosaccharide portion of a glycosphingolipid — ganglioside GM1 — which is present on the surface of cells forming the gut wall. On binding to five copies of this glycolipid, the A subunit (27 kDa) enters the cell through an, as yet, unkown mechanism, where it catalyzes ADP-ribosylation of the signal transduction protein G_s-α. This modification prevents deactivation of G_s-α, and consequently leads to high intracellular levels of cAMP, which, in the small intestine, results in fluid loss and severe diarrhoea. As B-subunit adhesion to the surface of a target cell is a prerequisite for entry by the A-subunit, this protein-carbohydrate recognition event is a potential target for developing drugs against the toxic effects of these bacteria.

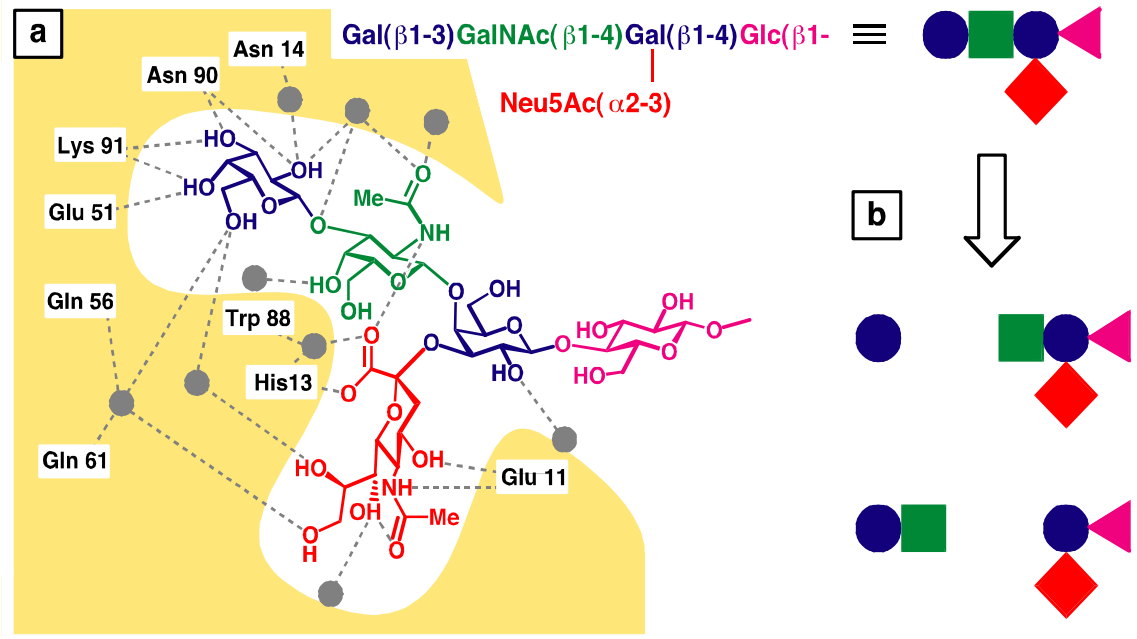


Fig. 1. a) Complex of GM1 oligosaccharide with the cholera toxin B-subunit (CTB) with key hydrogen bonds between the ligand, protein and bound water molecules (grey circles) indicated as broken lines; b) Cartoon representation of the fragments of the oligosaccharide ligand that are being used in binding studies with CTB.

Dissecting the "bivalent" interaction

Although very important in cell surface biology, protein-carbohydrate interactions are notoriously weak, often having dissociation constants in the millimolar range. Nature circumvents this problem by displaying multiple copies of both the carbohydrate ligands and their protein receptors in such a way that many weak interactions reinforce one-another to give a strong overall adhesion — not unlike molecular-scale velcro. In the case of CTB and

LTB, this so-called multivalent effect manifests itself on two levels, most obviously in the pentavalent binding of the B_5 ring to five copies of GM1. However additionally, X-ray crystallography of the complex has previously revealed that on the smaller scale of an individual subunit, the branched oligosaccharide grabs hold of the protein in a "two fingered grip" (see Fig. 1). These "bivalent" interactions rank among the highest intrinsic affinities in glycobiology and thus form a suitable model system for analysing the thermodynamics of interaction on a per-saccharide residue basis. Therefore, we have studied the binding abilities of fragments of the natural oligosaccharide ligand with the aim of dissecting the individual contributions from each monosaccharide to the overall interaction.

Synthesis and binding studies of GM1 fragments

Whereas smaller mono- and disaccharide fragments of ganglioside GM1 are either commercially available or could be accessed readily by chemical synthesis, larger fragments — including the full pentasaccharide — were most easily produced by stepwise enzymatic degradation of the natural ligand. Isothermal titration calorimetry (ITC) is a technique that exploits small stepwise changes in heat that is released during the course of a ligand-receptor titration to allow derivation of all thermodynamic parameters (free energy, enthalpy and entropy changes) in a single experiment. We have used displacement titrations to determine binding information for the low affinity fragments, whilst exploiting the high sensitivity of the high affinity GM1os interaction. Analysis of the intrinsic binding affinities for the ligands has shown that the terminal galactosyl and sialosyl residue each contribute just under 50% of the CTB-GM1os interaction, but paradoxically, sialic acid itself has no appreciable affinity for the receptor. Our results have thus also highlighted the origin of the high selectivity of CTB for GM1.

Acknowledgement

This work has been funded by the Wellcome Trust through the award of an International Prize Travelling Research Fellowship to BT.

Molecular dynamics in mouse urinary protein by NMR methods.

Kothandaraman Seshadri, Arnout Kalverda, Shih-Yang Hsieh and Steve Homans

Introduction

Protein function is dependent on dynamics. Molecular dynamics studied by NMR relaxation phenomena offers a great potential by which standard free energies and entropies can be derived on per-site basis - a level of detail crucial for the understanding of molecular basis of ligand-protein interactions. Despite the vast accumulation of X-ray crystallographic, static data of bio-molecular structures of proteins that are enzymes and receptors, relatively little is known about their dynamic behaviour. A wide range of time scales, extending from picoseconds to seconds, can be probed by spin relaxation. Such data obtained on residue-by-residue basis greatly enhances opportunities in protein engineering and ligand optimisation in rational drug-design.

MUP as a model system

The mouse major urinary proteins (MUPs) are a class of highly homologous, pheromone-binding proteins having molecular weights of about 19kDa. MUP-I is a member of group 1 gene (totally 30 genes encode the family) products that are expressed in the liver under hormonal control and are excreted in the urine of male mice at high levels. MUPs bind to several small hydrophobic pheromone molecules and the function is to carry them through aqueous environment, protecting from decomposition, regulating their release from urine. The high resolution X-ray crystal structure available indicates a conserved beta-barrel of eight beta-strands and an alpha helix. A preliminary isothermal calorimetric (ITC) study in our laboratory indicated that two of the ligands to MUP, 2-isopropyl, 3-methoxypyrazine (propyl) and 2-isobutyl, 3-methoxypyrazine (butyl) display an order of difference in binding affinity to MUP, despite similarity in their chemical structures. MUP and its small hydrophobic ligands serve as a suitable model system in which to conduct NMR relaxation studies and to rationalise differences in relaxation parameters between different protein-ligand complexes in terms of binding specificities.

Methyl dynamics from side chain ^{13}C and ^{2}H relaxation

The strategy to study aspects of protein dynamics and entropy employs methyl groups as probes since methyls are in general well dispersed in proteins, are often present at molecular interfaces, and have favourable spectroscopic properties. Fractionally deuterated methyl groups (CHD_2 isotopmer) offer advantage in studying side chain dynamics in that they avoid difficulties arising from interference effects between multiple ^{13}C-^{1}H bond vectors as ^{13}C is relaxed solely by a single proton. The experiments involve recording a series of high-resolution ^{13}C-^{1}H correlation maps with the intensity of each correlation attenuated by ^{1}H relaxation as a function of delays characterising appropriate relaxation parameter (T_1 or $T_{1\rho}$) and mono exponentially fitting the decay of intensities to obtain the relaxation time constants T1 and $T_{1\rho}$. On the other hand, study of deuterium relaxation on a CH_2D isotopmer has the benefits of avoiding complex exchange processes. Since deuterium is dominated by quadrupolar relaxation, the governing equations are more accurate.

The relaxation parameters (T1, $T_{1\rho}$ and nOe) sensitive to the picosecond-nanosecond time scale probe the amplitude of bond vector fluctuations. These parameters can be expressed in terms of an order parameter (S^2) that varies between 0 for an unrestricted internal motion to 1 for rigidity. S^2 can be related to the entropy of the corresponding molecular degrees of freedom under the 'model-free' formalism described by Lipari and Szabo. The data obtained both in the absence and in the presence of ligand can be interpreted in terms of free energies and free entropies of binding on a per-residue basis.

The analysis of S^2 values indicate indeed that the methyl groups exhibit a range of dynamic characteristics in both apo MUP and complexes. There are also specific differences between the complexes. An illustrative example is provided in the figure where the binding site methyl carbons are shown as spheres surrounding the butyl ligand (crystal structure courtesy, Prof. Simon Philips) colour coded as ranging from red (S^2=0) to blue (S^2=1). Comparative analysis of the data between apo, propyl and butyl cases is underway.

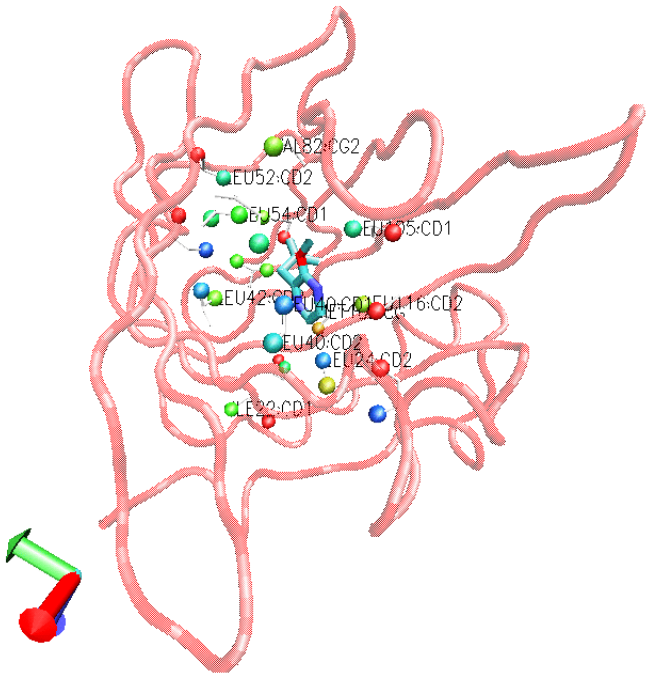


Fig. 1. MUP binding pocket displaying methyl carbons as spheres colour coded according to order parameter (S^2). For the labelled residues, colours range from red (S^2=0) to blue (S^2=1) via green (S^2=0.5). The bound ligand 2-isobutyl, 3-methoxypyrazine is also shown.

Collaborator:
Prof. J B C Findlay

Funding:
We acknowledge the support of the BBSRC.

Solution state NMR studies on a large α-helical membrane protein

Simon Patching, Peter Henderson, Richard Herbert and Steve Homans

Introduction

The structure elucidation of membrane proteins is notoriously very difficult; indeed, of the thousands of high-resolution protein X-ray crystallographic structures, relatively few are for membrane proteins and of these only three are for large α-helical transporters (see http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). Until recently, the application of solution state NMR techniques to elucidate protein structure was limited to soluble proteins with a molecular mass of less than 30 kDa. Larger proteins tumble more slowly in solution, such that NMR transverse relaxation is more efficient and this produces increased linewidths and a corresponding decrease in sensitivity in NMR spectra. These effects are even more pronounced for membrane proteins, which have to be solubilised in detergent micelles. The use of isotopic labelling strategies in combination with new NMR experiments and higher field NMR magnets have substantially increased the size limit of proteins amenable to structure-elucidation by solution state NMR and these techniques have been extended to membrane proteins. Notably, when perdeuterated proteins, which results in removal of transverse relaxation pathways for nuclei *via* directly attached protons, are used with transverse relaxation-optimised spectroscopy (TROSY) and related experiments, an impressive reduction in the transverse relaxation effects on solution state NMR spectra is produced. For example, TROSY has been used elsewhere to achieve the complete spectral assignment of the octameric DHNA aldolase (110kDa) from *S. aureus* and to determine the structure of the *E. coli* outer membrane protein OmpX (16.5kDa) in DHPC micelles. However, the TROSY technique has not yet been successfully applied to an α-helical membrane protein.

The *E. coli* galactose transporter, GalP

GalP is a H^+-gradient driven symporter found in the inner membrane of *E. coli* that catalyses the uptake of sugars, including D-galactose and D-glucose, into the organism. The protein is a member of the major facilitator superfamily and is closely related to the human glucose transporter GLUT1. Comprised of 464 amino acid residues, GalP has a molecular mass of 51kDa and is predicted to form 12 transmembrane-spanning α-helices. The system is an ideal candidate for NMR structural studies, since GalP can be overexpressed to 50-60% of the total protein in native inner membranes and has been characterised extensively in the Henderson laboratory using a range of biochemical and biophysical techniques and by site-directed mutagenesis. We are therefore attempting to perform solution state NMR TROSY measurements on GalP.

NMR studies on GalP

To enable ^{15}N-^{1}H TROSY NMR measurements, we have developed growth conditions for successfully incorporating uniform ^{2}H and ^{15}N labelling into GalP. The large isotope effect of deuterium reduces the cell yield and produces a decrease in the level of over-expression of the protein compared to that from unlabelled conditions; however, we are still able obtain 2 mg of purified [U-^{2}H, U-^{15}N]GalP per litre of culture medium. The stability of GalP to precipitation and its binding activity have been assessed with the protein solubilised in a range of types and concentrations of detergents. Using a fluorescence assay of protein activity through the binding of the GalP inhibitor, cytochalasin B, we have shown that the protein retains activity at 25 °C under the chosen solubilisation conditions for the duration of NMR experiments. We have obtained initial TROSY NMR data at 600 MHz for [U-^{2}H, U-^{15}N]GalP in which resonances representing important amino acid residues, *e.g.* tryptophan, glutamine, asparagine, may be distinguished (Fig. 1).

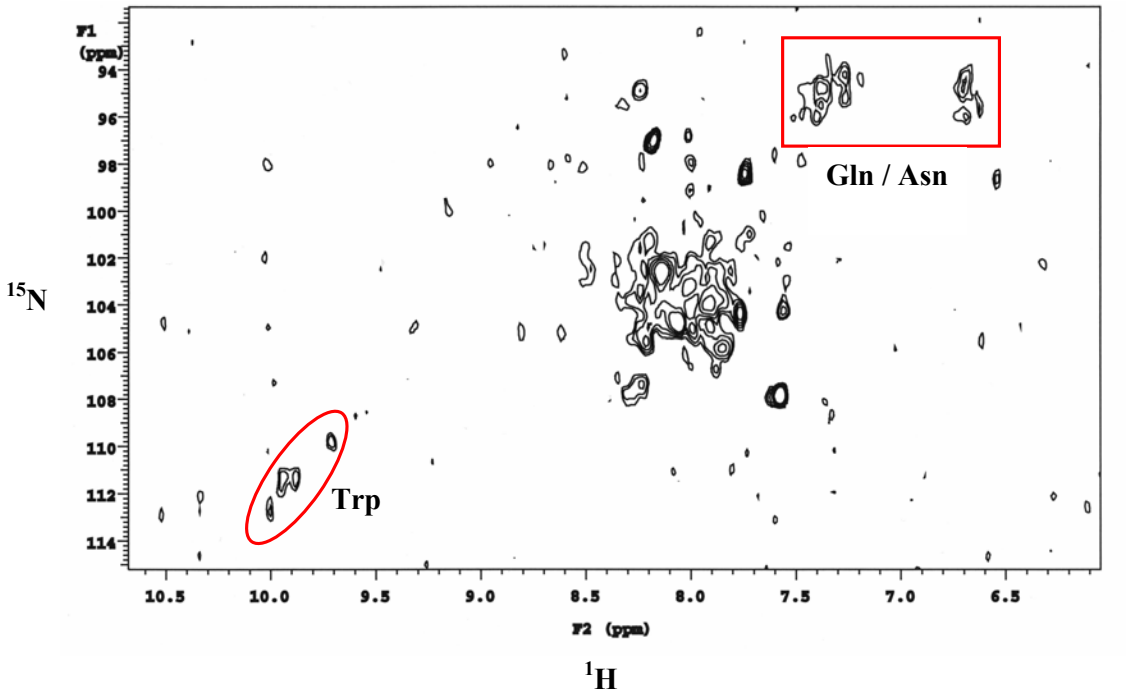


Fig. 1. 2D [^{15}N, ^{1}H]-correlation TROSY spectrum of [U-^{2}H, U-^{15}N]GalP. The spectrum was obtained at 600 MHz for ^{1}H at 25 °C over 12 h for 3 mg of [U-^{2}H, U-^{15}N]GalP (*ca.* 120 μM) in 10 % DDM (196 mM), 10 mM phosphate, 5% glycerol, 10% D_2O.

Analogous spectra of the protein were also recorded in the presence of cytochalasin B; these appeared to show changes in the position of certain resonances compared to those in the spectrum obtained in the absence of the inhibitor. To achieve spectra with improved sensitivity and resolution we intend to perform TROSY measurements at 750 MHz using larger quantities of protein.

We are also attempting to produce samples of labelled GalP in which the protein is weakly aligned within a polyacrylamide gel; under such conditions dipolar coupling information is recovered to NMR spectra, from which internuclear distances in the protein could be measured.

Funding
We thank the BBSRC for funding.

The NMR solution structure of the VMA-7 subunit of the vacuolar H^+-ATPase from *Saccharomyces cerevisiae*

Gary Thompson, Liz Barratt, Lyndsey Durose, Bruce Turnbull,
Mike Harrison, John Findlay and Steve Homans

Introduction

Vacuolar H^+-ATPase (VATPase) is a membrane spanning proton pump which is found in virtually all eukaryotic cells. The action of the pump results in translocation of protons across vacuolar membranes, allowing the pH of intracellular compartments to be regulated. Errors in the function of the pump have been implicated in the pathology of a number of diseases including osteoporosis, diabetes and several common cancers including cervical cancer. A structural understanding of this complex at the atomic level will open up paths to new therapeutics for these diseases by control of its activity. Though the structure of the VATPase enzyme complex has been shown to be related to that of the F_1F_0 ATPase, there are major differences in the subunits and the composition of the two enzymes. Therefore, a significant effort is underway at Leeds under the MASIF scheme to determine the structure of the enzyme subunits and the complete complex using modern structure determination methods: X-ray crystallography, NMR spectroscopy and electron microscopy. The VATPase complex contains at least ten distinct subunit types and four of these are of a size amenable to modern NMR techniques. The current target, VMA-7, is a 14 kDa subunit of the complex, which has unknown function and fold, and can be over expressed in *E. coli* and purified as a GST fusion protein. Preliminary structural studies using FT-IR, CD and NMR showed that it is folded in aqueous solution and contains ~ 30 % alpha helix.

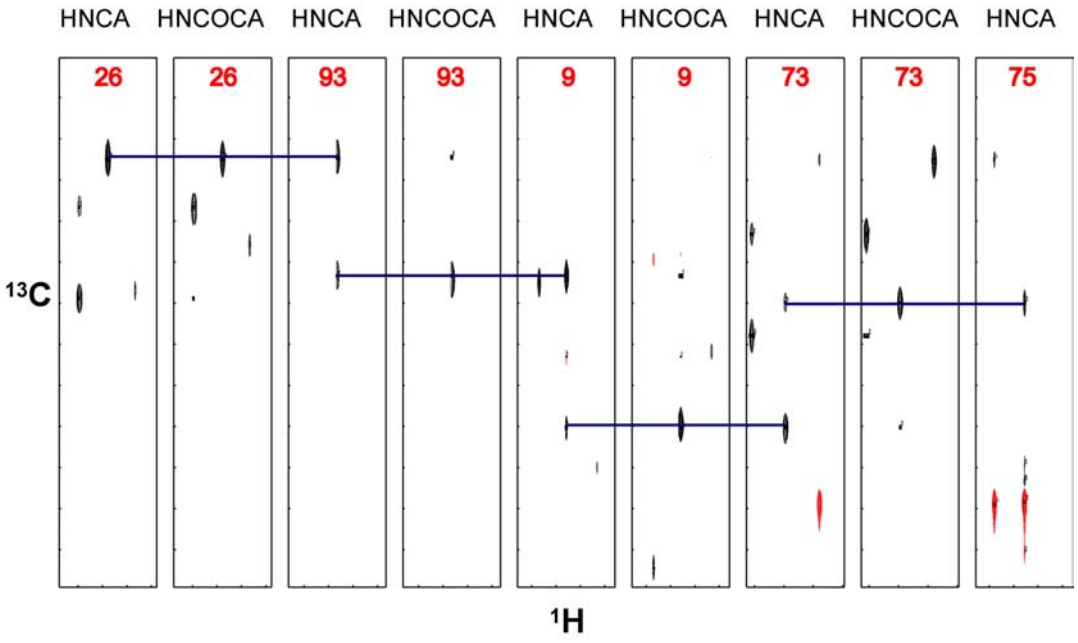


Fig. 1. Inter-residue correlations in strips taken from HNCA and HN(CO)CA spectra of VMA-7. Numbers assigned to clusters sharing common backbone amides are listed in red.

The VMA-7 cDNA has been transferred to a GST Precision Protease Fusion vector which has allowed us to produce protein free from low levels of contaminating proteases that are found in thrombin. The over-expression of the protein in M9 minimal media has been optimised to give 4 mg of fully purified ^{13}C and ^{15}N enriched protein per litre of culture. Triple resonance

HNCA, HN(CO)CA, HNCO and HNCACB experiments have been acquired and are being used to assign the backbone resonances of the protein. This process is being facilitated by the development of novel methods involving three way decomposition in order to resolve badly overlapped regions of the 3D spectra. The 3D structure of the complex will be calculated using a hybrid approach using both residual dipolar coupling and NOESY restraints.

Several other subunits of the VATPase complex have also been over-expressed in *E. coli* and will soon be taken forward into solution NMR studies.

Publication

Jones, R.P.O., Hunt, I.E., Jaeger, J., Ward, A., O'Reilly, J., Barratt, E.A., Findlay, J.B.C. & Harrison, M.A. (2001), Expression, purification and secondary structure analysis of *Saccharomyces cerevisiae* vacuolar H^+-ATPase subunit F (VMA7p), *Molecular Membrane Biology*, **18**, 283-290.

Funding

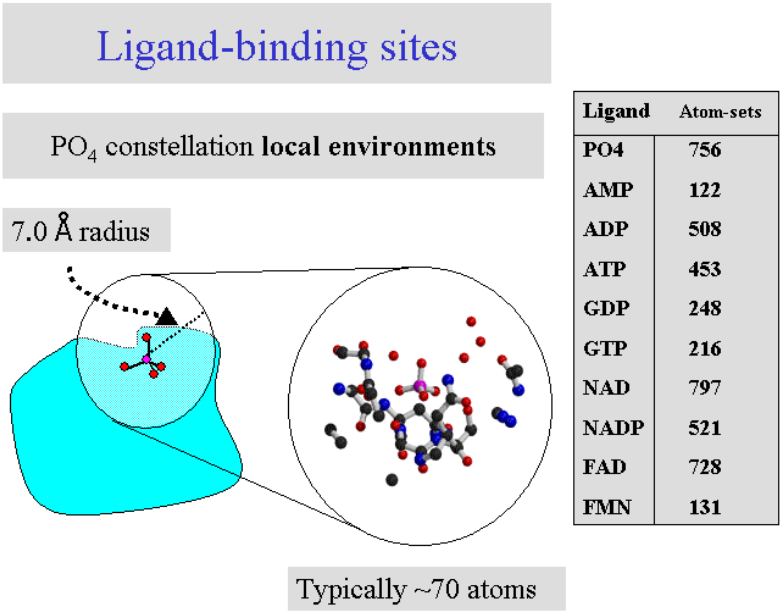
We wish to thank the BBSRC and EPSRC for funding.

A new method for comparing ligand binding sites in biomolecules

A. Brakoulias and R. M. Jackson

Analysis of completed genomes from a number of organisms reveal that about half of all gene products are classified as functionally unknown hypothetical proteins, and structures are already being solved prior to any knowledge of function. This will make the structure-based prediction of molecular function of increasing importance in the future. The ability to detect and classify local atomic level similarity will allow the development of rules for the prediction of function directly from structure (the premise of Structural Genomics).

Initially, proteins that bind ligands containing the phosphate moiety are being studied. In particular, the large class of nucleotide ligands (ATP/ADP, GTP/GDP, NAD, FAD etc.).



Ligand	Atom-sets
PO4	756
AMP	122
ADP	508
ATP	453
GDP	248
GTP	216
NAD	797
NADP	521
FAD	728
FMN	131

Fig. 1. The local protein-ligand atomic environments are extracted from the PDB for proteins that bind phosphate containing ligands.

We have developed a method that identifies the extent of the similarity between the structure of different protein-ligand binding sites allowing their structural classification.

The classification procedure allows the generation of structural binding site "templates". These could be used to identify similarity between the binding site of proteins in the database (for which an enzyme mechanism or binding site is well determined) and a newly determined protein of unknown function. Clearly, knowledge that a protein has a nucleotide binding site as well as information on the function of other proteins in the database that share this site will be valuable.

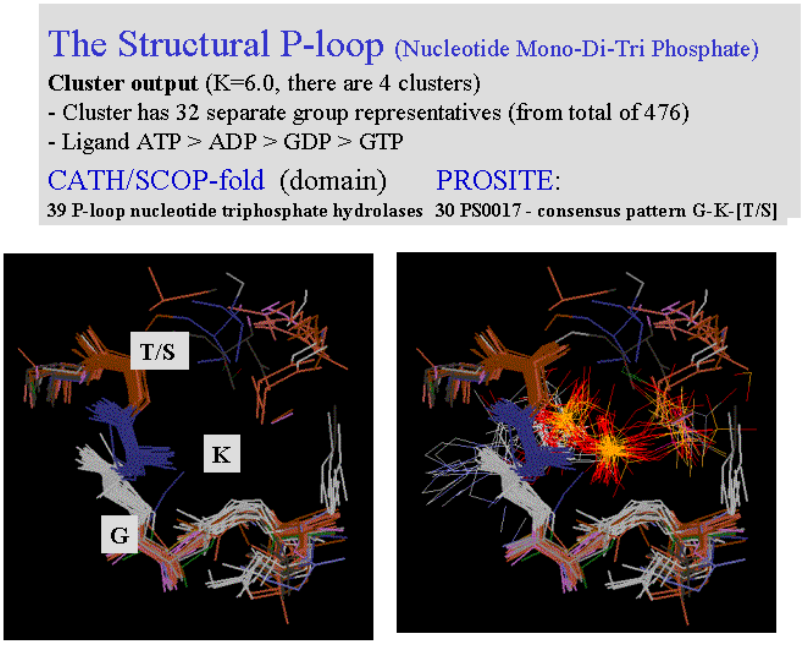


Fig. 2. Following an all-against-all comparison of nucleotide phosphate binding sites group representatives are clustered to identify similar sites. This figure shows one such cluster (main-chain atoms only) which corresponds to the structural P-loop. The residues are coloured according to residue type. It can be seen that members of the cluster contain an identifiable sequence pattern identified in the PROSITE database (GK-[T/S]). The nucleotide moieties are also shown in the right hand picture.

Collaborators

Prof. Janet Thornton (EBI)

Publication

Brakoulias A., Jackson R. M. (2003) A new method for comparing ligand binding sites in biomolecules. *Manuscript in preparation.*

Funding

We wish to acknowledge support of The Wellcome Trust, BBSRC and Onasis Public Benefit Foundation.

Analysis of SH2 domain phosphopeptide interactions

S. J. Campbell & R. M. Jackson

Src homology 2 (SH2) domains are highly homologous phosphotyrosine-binding motifs of approximately one hundred amino acids. They are found in a range of signal transduction proteins and are therefore important targets for therapeutic intervention. The domains are well-conserved, particularly at the two major binding regions known as the pTyr and pTyr+3 binding pockets.

We have applied a variety of computational & molecular modelling techniques to the system, including structural and sequence alignments, clustering of binding sites and residue conservation scoring. This has allowed us to investigate molecular recognition within and beyond the traditional pTyr to pTyr+3 linear motif. The clustering of binding sites is shown in Fig. 1, where the newly clustered domains are in qualitative agreement with the results of experimental screening, indicating that the method used to cluster is able to independently generate functional information about the domains.

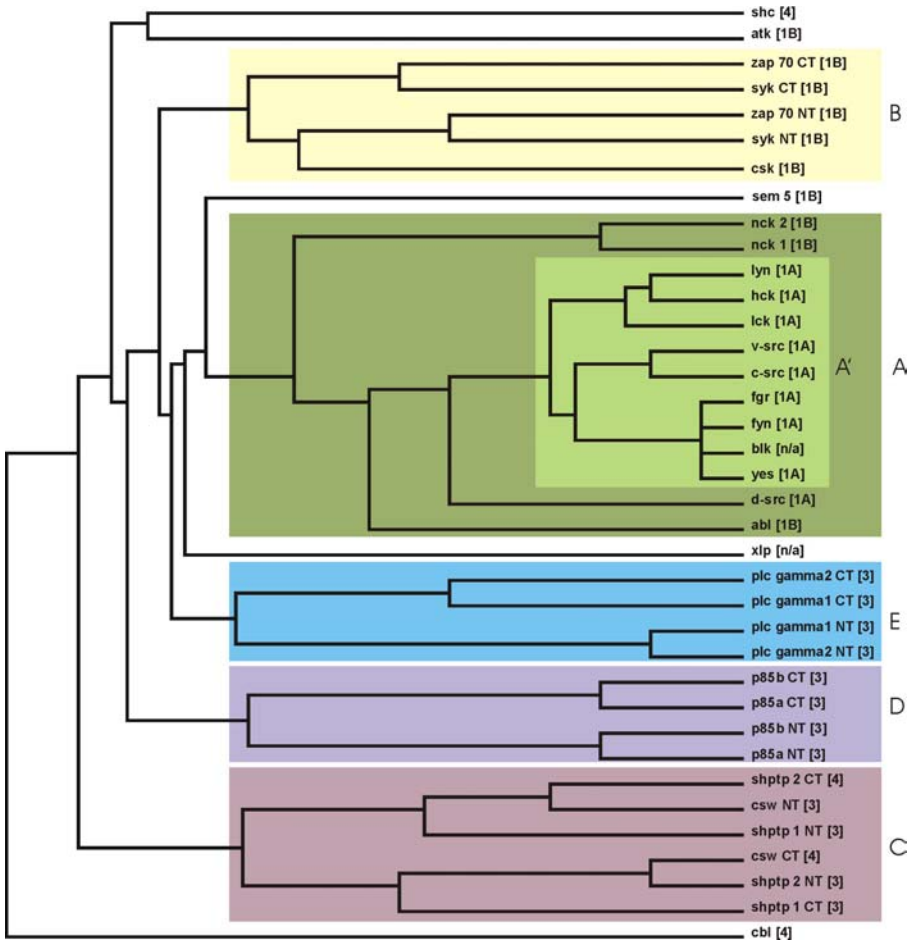


Fig. 1: Clustering of Binding Sites. We have used the experimentally defined protein-ligand contacts to define the protein binding site, allowing the clustering of the domains according to amino acid conservation in the binding site.

Fig. 2 shows residue conservation within a large experimentally determined group of SH2 domains, versus a group of domains determined using the clustering method. All members of the theoretical group are present in the larger experimental 'parent' group. The conservation

maps reveal regions beyond the traditional pTyr to pTyr+3 binding motif where the theoretical group shows residue conservation but the experimental group does not. Such regions could be involved in the discrimination and recognition of ligands.

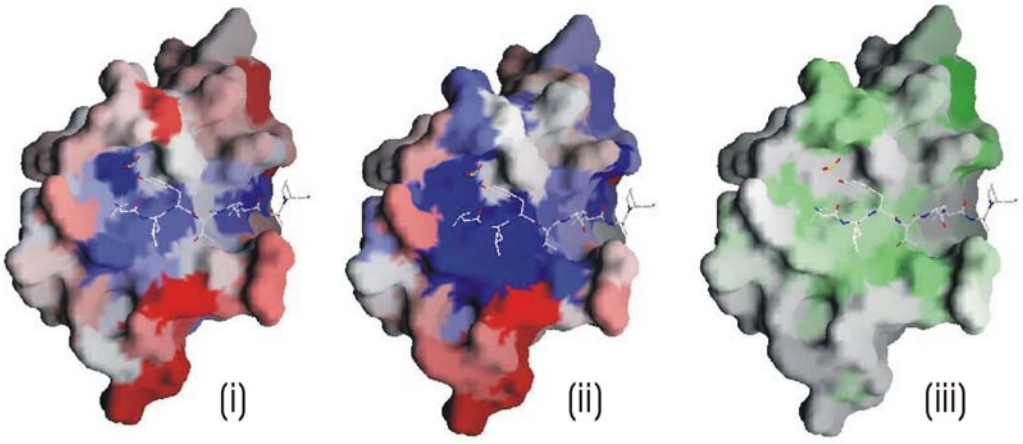


Fig. 2: Diversity within the phosphotyrosyl binding site. (i) 'Parent' group representing a group of SH2 domains based on experimental screening according to recognition within the pTyr to pTyr+3 motif. (ii) Group C SH2 domains from Fig. 1, determined computationally by comparison of binding site residues. (i) and (ii) are coloured by residue conservation score where blue indicates a high level of conservation and red indicates low conservation. (iii) Difference map showing the difference in residue conservation scores between the experimental group and the theoretical group mapped back to the same structure. Green regions represent a gain in conservation in the theoretical group relative to the experimental parent group. Grey regions represent a loss of conservation in the theoretical group relative to the experimental group. White regions represent no change in conservation score.

Publications

Campbell, S. J., & Jackson, R. M. (2003) Diversity in the SH2 domain family phosphotyrosyl peptide binding site. *Protein Engineering*, **16**, 217-227

Funding

This work was supported by the BBSRC.

Q-fit: A method for docking molecular fragments by sampling low energy conformational space

P. Oledzki, P. Lyon and R. M. Jackson

QFIT is a small molecule docking algorithm for predicting biomolecular-ligand interactions. It uses a new probabilistic sampling method and a conventional molecular mechanics force field for finding small molecular fragments in the binding site of a rigid protein. The aim is to facilitate the process of finding novel small molecule lead compounds in a fast and efficient way. Potentially these fragments could be connected using combinatorial principles to create *de novo* compounds, or used as the base fragment for searching existing chemical databases.

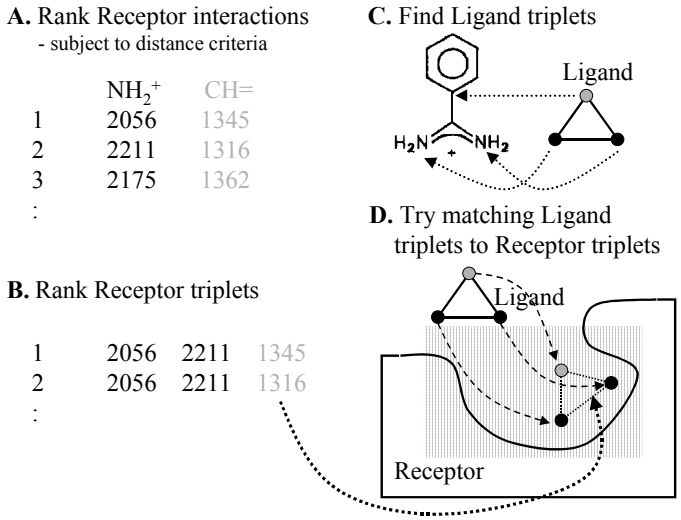


Fig. 1. Schematic representation of the steps involved in the docking algorithm. The numbers in A and B refer to indices of the grid points.

The method proves to be very successful in its ability to predict experimental binding modes when using the point interaction model and force field of GRID. The binding modes of several ligand fragments are predicted that have previously proved problematic. The binding modes of ligands used to bench mark several other docking algorithms are reproduced in excellent agreement with experiment. Furthermore, because the probabilistic docking methodology tends to generate low energy conformations first, the time to perform this docking is very fast. Docking of a single ligand fragment including energy minimization can be performed in 10-15 seconds on a desktop PC using conservative parameters for the search depth.

Recently, we have extended the method to include flexible ligand docking. The method proceeds by docking the individual rigid fragments contained within the larger ligand. The top scoring placements are then extended by incremental construction to produce a placement of the complete ligand. Currently, we are optimising parameters for the steps involved.

Publication

Jackson, R. M. (2002). Q-FIT: A Probabilistic Method for Ligand Docking by Sampling Low Energy Conformational Space. *Journal of Computer-Aided Molecular Design*. **16**, 43-57.

Funding

We wish to acknowledge support of The Wellcome Trust and BBSRC.

Crystallographic studies of (+) strand RNA-directed RNA polymerases

Amy Cherry, Caitriona Dennis, Russell Green, Rachel Trowbridge, Elena Curti, Mark Saw, Ben Linford, David Rowlands, and Joachim Jäger

Positive strand RNA viruses

Positive strand RNA viruses such as picornaviruses and flaviviruses are serious human and animal pathogens of significant medical, social and economical relevance throughout the world. For example, Hepatitis C virus, which affects around 3% of the world population, has been identified as the main causative agent of non A/non B hepatitis. Other members of the flaviviridae include Yellow Fever virus, Dengue Fever virus, West Nile virus, Japanese Encephalitis virus and many others. These viruses produce their own replication machinery, namely an RNA-dependent RNA polymerase that is responsible for the initial synthesis of a complementary (-) strand, which in turn serves as a template for the production of progeny (+) strand. In the search for new viral treatments, three essential enzymes of HCV, the NS5B polymerase and the NS3 protease-helicase, have been selected as potential targets for antiviral therapy. The successes in the development of effective drugs directed against HIV-1 reverse transcriptase highlights the importance of viral replicases as drug targets.

Crystallographic Studies

Various crystal forms of NS5b polymerase (consensus sequences HC-BK and HC-J4, genotype 1b) have been reported recently. In the Astbury Centre, we have established a low ionic strength crystal form of HC-J4 polymerase (300mM NaCl, 16% PEG4000 and 10% glycerol) in which we have been able to introduce various ligands into the active site of NS5b. A single binding site for ribonucleoside-triphosphates (rNTP) has been established for HC-J4 RNA polymerase. The rNTPs bind to the catalytic site of NS5b in a productive fashion, which is evident from the pattern of interactions with both catalytic metal ions, numerous interactions with strictly conserved basic residues and a water molecule that is found 3.9Å away from the α-phosphate. The water is presumably mimicking the attacking oxygen of the primer strand in the nucleotidyl (or phosphoryl) transfer reaction mediated by all polymerases. The rNTP complexes also suggest that the active site is 'pre-assembled' and that very little conformational change will be required to initiate RNA synthesis *de novo*. This is in contrast to DNA polymerase where large motions of the 'O-helix' and the thumb domain are required to form a catalytically competent and faithfully replicating active site cleft.

In HC-J4 RNA polymerase the template strand enters the active site through a funnel formed by structural elements of the finger tips near the N-terminus and the fingers domain. Four to five nucleotides have to be separated from the duplex in the third stem-loop (SL3) of the HCV 3'UTR to reach into the active site and to form a Watson-Crick base pair. In the 'template-only' crystal structure of HC-J4 NS5B, i.e. in the absence of an incoming

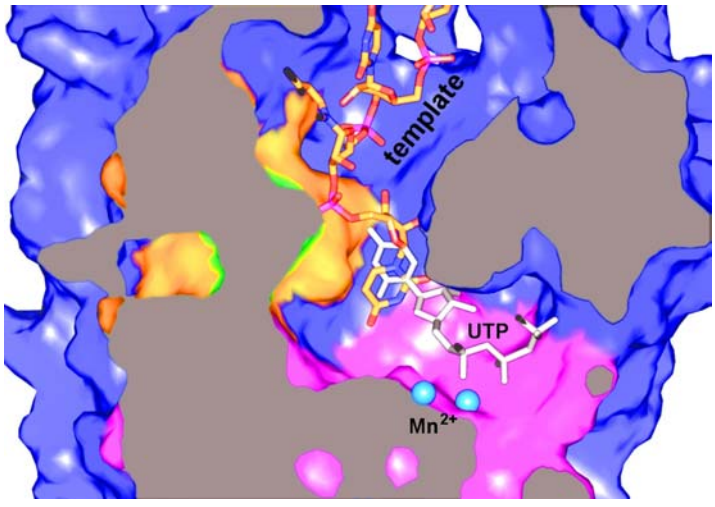


Fig. 1. The active site cleft of HCV RNA polymerase indicating the path of the viral template and the position of the incoming rNTP.

nucleotide, the template base overshoots the expected position near residue Ile160 by 5-7Å.

Functional studies of other RNA dependent RNA polymerases are currently underway. The design of site directed mutants is guided by the rNTP and ssRNA complexes of HC-J4 RNA polymerase and the co-crystal structures of phi6 RNA polymerase complexed with dsDNA.

Publications:

O'Farrell, D., Trowbridge, R., Stock, N., Rowlands, D. & Jaeger, J. (2003) Co-crystal structures of HC-J4 RNA polymerase: implications for nucleotide binding and *de novo* initiation. *J.Mol.Biol.* **323**, 1025-1035

Jaeger, J. & Pata, J.D. (2002). Viral RNA-directed polymerases. *Structure function Relationships of Human Pathogenic Viruses.* E.Bogner ed., Kluwer Academic/Plenum Publishers.

Collaborators

Pia Thoemmes, Virology, GlaxoSmithKline, Stevenage,UK .
Onkar Singh, Virology, GlaxoSmithKline, Stevenage,UK
Tobias Restle, MPI Dortmund, Germany
Karla Kirkegaard, Stanford, USA
Jens Bukh, NIH, Bethesda, USA
Berwyn Clarke, VisGen Ltd., Cambridge

Funding:

GSK, MRC, BBSRC and the Wellcome Trust

Functional analysis of nucleotide import in Hepatitis C virus RNA polymerase

Rachel Trowbridge, Amy Cherry, Russell Green, Ben Linford,
David Rowlands, and Joachim Jäger

Conserved sequence motifs in viral RNA polymerases

Hepatitis C virus RNA polymerase mediates the positive and negative strand synthesis of the viral RNA genome. Careful analysis of multiple alignments of polymerase sequences from positive strand RNA viruses reveals five conserved motifs (A to E). The crystal structures of poliovirus protein 3D, phi6 bacteriophage, rabbit hemorrhagic fever virus and HCV RNA polymerase have confirmed that these motifs map into the polymerase active site. Crucially, motifs A and C carry three strictly conserved carboxylate residues which in turn bind the catalytic magnesium ions. Motif E is involved in positioning the 3' end of the primer in the active site less than 4Å away from the α-phosphate of the incoming nucleosidetriphosphate. More recently, through the comparison of HIV RT and HCV polymerase an additional structural motif has been identified (motif F). This motif is thought to control the import and faithful incorporation of a cognate NTP.

Control of nucleotide import and formation of Watson-Crick base pairs

The functional role of residues in motif F have been probed by SDM. In HCV, Arg158 interacts with the α-phosphate and the base of the incoming nucleotide. In analogy to HIV-1 RT, the guanidinium group of this residue likely forms a stacking interaction with nascent Watson Crick base pair. Generally, the replacement of highly conserved Arg and Lys residues in this motif with Ala results in variant polymerases with reduced activity. Presumably, the rate and the accuracy of nucleotide incorporation is hampered by these mutations, however, a detailed analysis of the dsRNA product is required to shed support this hypothesis.

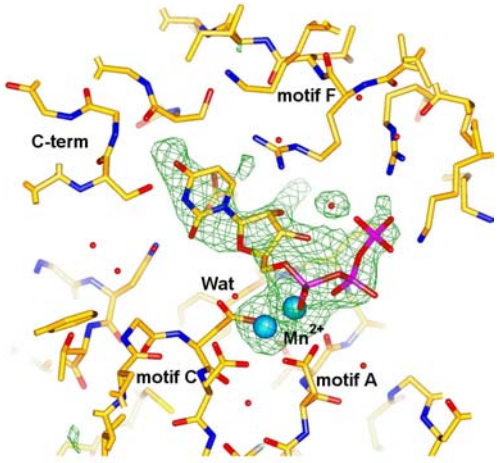


Fig. 1 2Fo-Fc omit map of the HCV NS5B/UTP/Mn complex

RdRps discriminate efficiently between rNTPs and dNTPs

Steady state kinetics are being used to investigate the overall rate of miscincorporation and discrimination between rNTPs and dNTPs. Residues which are thought to be involved in recognising the chemical and structural moeities in the (de)oxyribose are Asp225 and Asn287. Replacement of Asp225 by Ser and Glu results in a reduction of overall polymerase activity by no more than 3 fold. However, dNTPs are essentially not incorporated and polymerase activity for both rNTPs and dNTPs can only be fully restored by adding 2-3mM Mn to the reaction. This is in contrast to *E.coli* and T7 DNA polymerases, where changes in this particular position (in motif C, from Glu to Asp or Ser) had profound effects on the discrimination between dNTPs and rNTPs even in the presence of Mg. Clearly, (+) strand RNA viruses have evolved a more stringent and complex mechanism (than bacteria) in accepting rNTPs and rejecting dNTPs.

The transition from initiation to elongation mode

The structures of HCV and phi6 RNA polymerases have revealed that the thumb domain and the C-terminus (either a short peptide of 30 amino acids or small sub-domain) insert into the

polymerase active site. Interactions between the incoming nucleotide and residues from the β-flap and the C-terminal domain facilitate the primer-independent initation step of the replication cycle (*de novo* priming). Several site-directed mutants in the β-flap and the C-terminal arm of HCV polymerase display an increase in overall polymerase activity over the wild type enzyme. This increase in the incorporation is presumably due to facilitating the transition from the initiation mode (closed form) to elongation mode (open form).

Publications:

O'Farrell, D., Trowbridge, R., Stock, N., Rowlands, D. & Jaeger J. (2003) Co-crystal structures of HC-J4 RNA polymerase: implications for nucleotide binding and *de novo* initiation. *J..Mol.Biol.* **323**:4, 1025-1035

Collaborators:

Pia Thoemmes, Virology, GlaxoSmithKline, Stevenage,UK

Funding:

MRC, BBSRC and the Wellcome Trust

Structural determinants of polymerase fidelity and nucleotide discrimination in mammalian DNA pol ß

Mausumi Maitra*, Shu-Xia Li*, Elena Curti and Joachim Jäger

DNA polymerase ß mutator mutants

Pol ß is one of the smallest eukaryotic polymerases. It is a simple, well-characterised system to study the correlation between polymerase structure and fidelity of DNA synthesis. Polß plays a crucial role in base excision repair (BER) as it removes 5'-deoxyribose phosphate and fills in short lesions in DNA (short patch synthesis). Furthermore, polß is implicated in nucleotide excision repair. There is also evidence to suggest a meiotic function for polß.

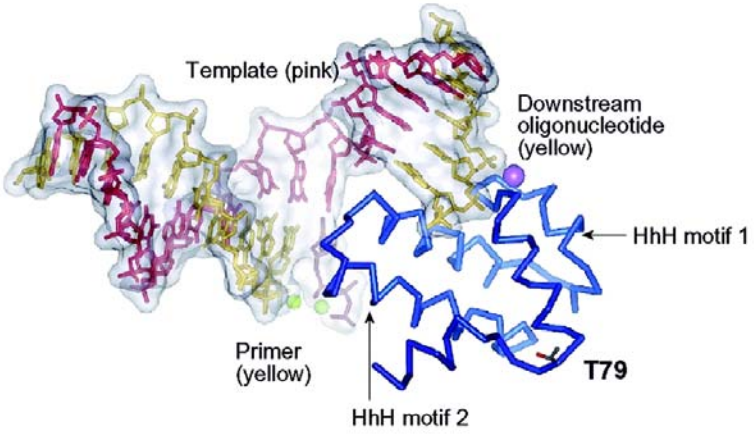
For these studies, polß mutants (for example F272L, M282L, and T79S) are being identified using an *in vivo* genetic screen based on the replacement of the DNA *pol*A gene in *E.coli* with the mammalian *polß* gene. Mutator phenotypes were isolated using a Trp^+ reversion assay from a library of random polß mutants. Polß variants that produce Trp^+ revertants in significant numbers and frequencies over wild type polß are isolated, transfected into the *E.coli* SC18-12 strain and characterised biochemically and structurally.

Polymerase fidelity

Transient-state kinetic methods were used to investigate the mechanism of the intrinsic mutator activity of F272L, M282L and T79S. In the case of F272L, the kinetic data show that the mutant produces error rates at a frequency 10-fold higher than that of wild type both *in vivo* and in *in vitro* assays. The mutant displays an increase in the frequency of frameshift mutations as well as in the rate of misincorporation. A G:A mismatch occurs 4 times more frequently in the F272L mutant than in wild-type Polβ.

For M282L, biochemical studies indicate an overall reduction in fidelity (11-fold decrease in dNTP substrate discrimination at the level of ground-state binding and 3-fold increase in nucleotide discrimination during the structural step and/or chemical step). Thermal and chemical equilibriuim unfolding studies were used to determine the overall stability of the mutant Polβ. The CD spectra indicate that the reduced fidelity appears to be related to a distinct increase in protein stability.

Thr-79 is located in the N-terminal 8kDa domain of polβ and has no contact with either the DNA template or the incoming dNTP substrate. The T79S mutant produced 8-fold more multiple mutations in the HSV1–TK assay than wild-type Polβ. Surprisingly, T79S is a misincorporation mutator only when using a 3'-recessed dsDNA substrate. In the presence of a single nucleotide-gapped DNA substrate, T79S displays an antimutator phenotype when catalyzing DNA synthesis opposite template C and has similar fidelity as wild type opposite templates A, G, or T. Thr79 forms a hydrogen bonding network within the helix-hairpin-helix motifs of the 8kDa domain that is important for positioning the DNA within the active site.



Interestingly, a change to Ser appears to disrupt this hydrogen bonding network and results in an enzyme that is unable to bend the DNA into the proper geometry for accurate DNA synthesis.

Structural studies
X-ray crystallographic methods are used to investigate the structural basis of the decreased DNA synthesis fidelity in these polß mutator mutants. Residue 272 is located on the DNA binding surface of polβ and a mutation at this position is not expected to result in a large conformational change of the entire molecule. The loss of the aromatic ring in the F272L mutant, however, appears to produce a geometrically less constrained dNTP-binding pocket. It also triggers a cascade of side chain movements indicating increased flexibility in the dNTP at the groundstate (prior to the structural or chemical step). This increase appears to be responsible for the decrease in dNTP affinity and fidelity.

In wild type polβ, Met-282 is located in the interior of the C-terminal domain and does not directly interact with the substrates. When mutated to Leu, residues 283, 293, and 295 shift to accommodate the reduction in side chain volume. In comparison with the wild type structure, the M282L mutant produces a collapsed, more densely packed hydrophobic core, which in turn results in a polymerase with enhanced protein stability. The increase in protein stability was corroborated by equilibrium unfolding studies in the presence of urea.

Publications:
Maitra, M., Gudzelak, A., Li, S.X., Matsumoto, Y., Eckert, K.A., Jager, J. & Sweasy JB. (2002) Thr79 is a residue that governs the fidelity of DNA polβ by helping to position the DNA within the active site. *J Biol Chem.* **277**, 35550-60

Shah, A.M., Conn, D.A., Li, S.X., Capaldi, A., Jaeger, J. & Sweasy, J. (2001) A DNA polymerase beta mutator mutant with reduced nucleotide discrimination and increased protein stability. *Biochemistry* **40**, 11372-81.

Collaborators:
Joanne Sweasy, Departments of Therapuetic Radiology and Genetics, Yale University School of Medicine.

Funding:
We thank the NIH for support

Proteomics and allied technologies

Jeff Keen

Introduction

The detailed molecular characterisation of proteins is a critical component of an active research environment in structural biology and requires appropriate instrumentation. In recent years Leeds has been very successful in acquiring the essential equipment and expertise to enhance protein analysis for biological and medical research in the Post-Genomic Era, including a Q-Tof™ tandem mass spectrometer (JREI 1998), an extensive protein chemistry facility (JREI 1999) and a high-throughput proteomics facility (JREI 2000).

Proteomics

Proteomics is widely used for the global analysis of protein expression patterns. The set of proteins expressed by a cell - the ***proteome*** (the ***prote***in complement of the gen***ome***) - is a dynamic entity, varying in response to the effects of environment on the genome. Proteomics enables a "snapshot" to be taken of the current state of protein expression, providing both qualitative and quantitative information on protein profiles and can be used to investigate changes in protein expression patterns in differing situations. For example, proteomics can be used to study the effects of genetic changes (*e.g.* mutation, gene knockout), the effects of environmental challenge (*e.g.* pollution, disease, drug intervention) or to investigate linked biochemical pathways.

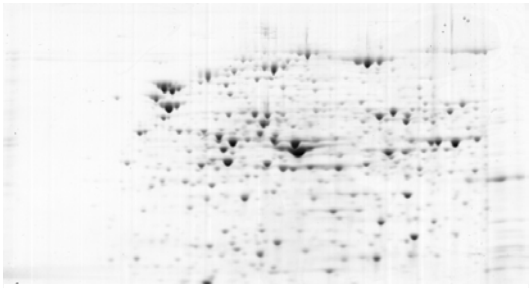


Fig. 1. 2Dimensional electrophoresis of *E.coli* extract.

The central technology typically involves 2-dimensional electrophoresis (2DE) for high-resolution separation of proteins, coupled to mass spectrometry for high-throughput protein identification. The aim is to separate highly complex protein mixtures into discrete components that can then be analysed independently to provide identification. This usually involves digestion of the protein and analysis of the peptide mixtures using mass spectrometry. The results are used to search databases to provide the identity of the parent protein. This knowledge can then be related to the biological situation, improving our understanding of how cells respond to differing situations.

There is considerable local interest in membrane proteins and membrane-localised biochemical events. Study at the proteomics level is hampered by protein hydrophobicity and low abundance. We are using our expertise in the handling of membrane proteins to improve the investigation of membranes using proteomics, such that numerous studies will benefit.

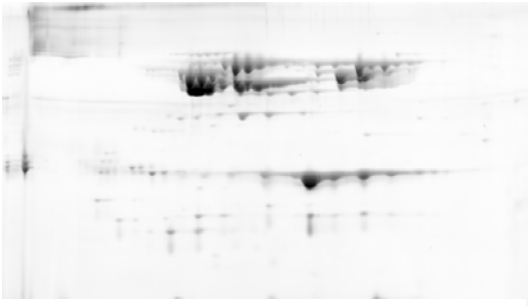


Fig.2 2Dimensional electrophoresis of *E.coli* outer membrane proteins.

Following an award in 2000 from HEFCE (through the JREI scheme), a comprehensive Proteomics Facility is being established, based on the Bio-Rad/Micromass ProteomeWorks suite of equipment, to enable more extensive and wider-ranging investigations. Typically, in any particular investigation, protein samples from different situations (*e.g.* control *v.* disease

tissue) will be separated by high-resolution 2DE, the protein patterns visualised through staining of the gels and digital images acquired. Sophisticated software will then be used to compare the images in order to establish standard and altered protein expression patterns. Proteins showing significantly different levels of expression in different circumstances will be selected for identification. Robotic equipment will be used to excise protein spots from gels, treat them individually with protease to generate peptide mixtures and prepare these mixtures for analysis by mass spectrometry. MALDI-MS will be used to screen these digests, producing lists of peptide masses that will be used to search databases automatically to identify the parent protein. Digests that do not provide a protein identity will be put forward for Q-Tof™-MS/MS to generate sequence tags to aid identification. A single experiment could generate the identities of many proteins exhibiting varied expression levels, providing numerous clues to what is happening at the cellular level and identifying targets for detailed structural and functional characterisation.

We have been working with a number of collaborators within Leeds to establish parameters for successful proteomic investigation of their diverse systems. These studies include the identification of bacterial enzymes involved in bioremediation (with Prof. Simon Baumberg and Dr. Jerry Knapp), the dissection of the Reg signalling response in *Rhodobacter* (with Dr. Mary Phillips-Jones), the identification of proteins involved in microbial biofilm formation (with Dr. Jon Sandoe) and the identification of *Schistosoma* tegumental membrane proteins (with Dr. Alison Agnew)

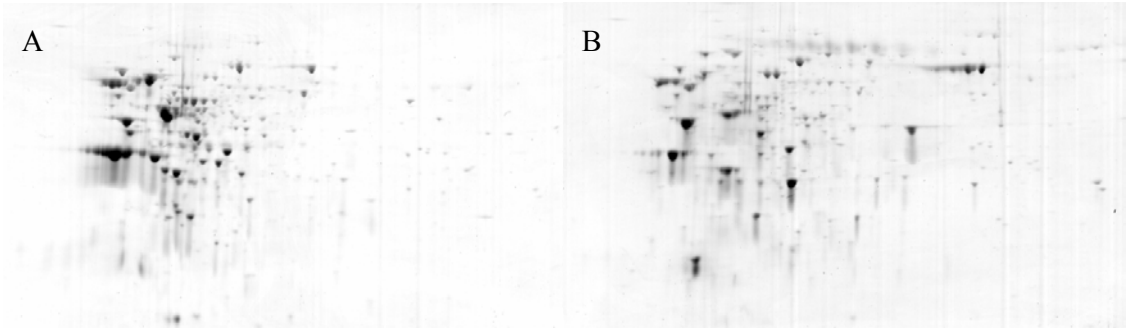


Fig. 3. 2 dimensional electrophoresis of membrane fractions of bacteria grown in standard culture conditions (A) or in serum (B).

Protein sequencing

The N- and C-terminal sequencing facility (funded through JREI 1999) has continued to attract samples for analysis from various national and international academic institutions and commercial enterprises over the last year. The N-terminal instrument has been used in internal and external projects for the confirmation of recombinant protein fidelity, the identification of unknown proteins and the provision of *de novo* sequences for cloning projects. It has on occasions provided such information at sub-picomole sensitivity, but typically requires 10-50 pmol starting material. The C-terminal facility is almost unique within the UK, with relatively few instruments worldwide. Interest in the characterisation of protein C-termini has remained steady, but there are limitations in its use currently. The C-terminal instrument requires considerably more material (nmol level) than N-terminal sequencing and generally can provide only a few residues of information. Thus far, it has been used to investigate the C-terminal integrity of a number of recombinant proteins and over-expressed sub-domains in structural investigations. There is a need now to improve the sensitivity of analysis such that lower amounts of material are required.

Antimicrobial peptides
A collaborative project with Dr. Deirdre Devine and colleagues (Oral Biology) is investigating the response of the opportunistic pathogen *Burkholderia cepacia* to challenge with antimicrobial peptides. The project has involved the synthesis of a variety of peptides with antimicrobial activity and the testing of their biological activity against a range of strains of the organism. Molecular biological and proteomic approaches are now being used to characterise the response of the organism.

Additional work in this area involves the investigation of the interaction of antimicrobial peptides with biological membranes. This involves collaboration with Dr. David Gidalevitz (Chemical Engineering)

Collaborators
John Findlay, School of Biochemistry and Molecular Biology, Leeds.
Deirdre Devine, Oral Biology, Leeds.
Jon Sandoe, School of Biochemistry and Molecular Biology (Division of Microbiology), Leeds.

Publications
Beaumont, N.J., Skinner, V.O., Tan, T.M-M., Ramesh, B.S., Byrne, D.J., MacColl, G.S., Keen, J.N., Bouloux, P.M., Mikhailidis, D., Bruckdorfer, K.R., Srai, S.K.S. & Vanderpump, M.P. (2003) Ghrelin can bind to a species of high-density lipoprotein containing paraoxonase. *J. Biol. Chem.* In press.

Isaac, R.E., Parkin, E.T., Keen, J.N., Nässel, D.R., Siviter, R.J. & Shirras, A.D. (2002) Inactivation of a tachykinin-related peptide: identification of four neuropeptide-degrading enzymes in neuronal membranes of insects from four different orders. *Peptides* **23,** 725-733.

Siviter, R.J., Nachman, R.J., Dani, P.M., Keen, J.N., Shirras, A.D. & Isaac, R.E. (2002) Peptidyl dipeptidases (Ance and Acer) of *Drosophila melanogaster*: major differences in the substrate specificity of two homologues of human angiotensin I-converting enzyme. *Peptides* **23,** 2025-2034.

Venter, H., Ashcroft, A.E., Keen, J.N., Henderson, P.J.F. & Herbert, R.B. (2002) Molecular dissection of membrane transport proteins: mass spectrometry and sequence determination of the galactose-H^+ symport protein, GalP, of *Escherichia coli* and quantitative assay of the incorporation of [ring-2-^{13}C]histidine and $^{15}NH_3$. *Biochem. J.,* **363,** 243-252.

Funding
We acknowledge the support of MRC/DERA, Applied Biosystems, Dionex, BioRad, Micromass and the University of Leeds.

Dynein structure and power stroke

Stan Burgess, Matt Walker and Peter Knight

Introduction

The beating of cilia and flagella depends on the action of members of the protein motor family, the dyneins. These bridge between the members of the outer ring of 9 doublet microtubules that form a major part of the well-known 9+2 arrangement of microtubules in the axoneme. It has been difficult to decipher the structure of the intact axoneme because its components are diverse, densely packed, and superposed. On the other hand purified dyneins have proven very unstable, and therefore unsatisfactory for crystallisation or even single molecule electron microscopy. Moreover dynein molecules are biochemically complex comprising one to three ~500 kDa heavy chains complexed with intermediate and light chains. For all these reasons the understanding of the structure and mode of action of dyneins is poor compared to that of myosins and kinesins, which are the living cell's other linear motors. A synthesis of biochemical and structural data indicated that monomeric dyneins comprise a central head domain, built in part from AAA domains, from which extends a stem and a stalk.

Our collaborators in Japan recently succeeded in isolating an unusually stable monomeric dynein (dynein c) from the alga *Chlamydomonas reinhardtii*, and we have used this preparation to obtain detailed images of the molecules after adsorption on a thin carbon film and negative staining. Extensive use of single-particle image processing methods has revealed both consistent details of the structure and the range of flexibility within it.

Dyneins use ATP hydrolysis to power movement during which the tip of the stalk is thought to undergo cyclical interactions with microtubules. The power stroke of the motor is thought to accompany loss of ADP and Pi from the enzyme. To investigate this structural transformation we have therefore compared images of the stable dynein.ADP.vanadate complex with those of the apo-molecule.

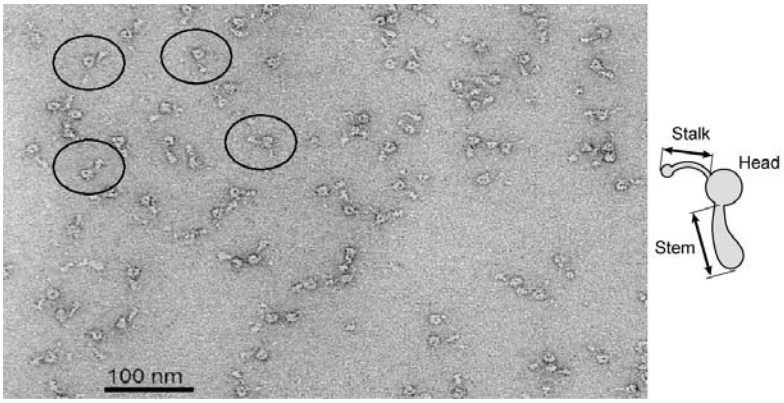


Fig. 1. A field of negative stained dynein c molecules.

Results

Micrographs show unprecedented detail, with both the stem and stalk visible in raw data (Fig. 1). Image processing (Fig. 2) shows that most appearances derive from a roughly toroidal shaped head that lands on the carbon film on one or other face, and the substructure of the two faces is not the same. The stem and stalk emerge much closer together than expected from sequence analysis, but we found evidence for a linker structure that explains the discrepancy.

Both stem and stalk display a range of shapes, indicating that considerable flexibility is possible.

There are many changes in the structure when products of ATP hydrolysis are released. The most pronounced is a movement of the stem relative to the head, to bring it closer to the stalk, and this is probably caused by movement of the linker across the surface of the head. This has the effect of shortening the molecule, so that a ~15 nm translation of the microtubules would occur under zero load. The coiled-coil stalk also stiffens, which suggests a means of transmitting information between the tip of the stalk and the ATPase site in the head.

Our observations lead to a speculative hypothesis for the mechanism of this motor (Fig. 3) in which we envisage conformational changes in the AAA domains couple to changing interactions with the linker and the stalk to produce cycles of movement and detachment.

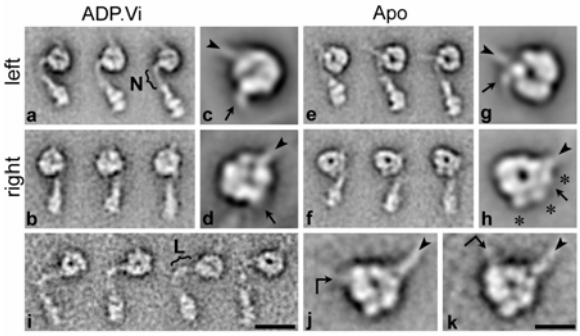


Fig. 2. Characteristic views of dynein c revealed by single-particle analysis. Selected class averages of the most common views. **a** - **d**, ADP.Vi-dynein c. **e** - **h**, Apo-dynein c. Selected class averages of rare molecules showing **i**, the linker (L), and **j** & **k**, head substructure after head alignment. Location of stalk (arrowheads) and apparent stem attachment (arrows) on the head are indicated. N denotes the neck. Asterisks show three subdomains. Scale bar for **a**, **b**, **e**, **f** & **i** 25 nm; **c**, **d**, **g**, **h**, **j** & **k** 10 nm.

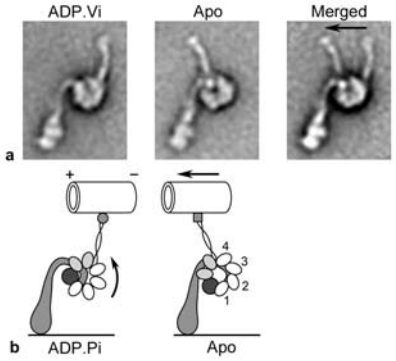


Fig. 3. Structure and power stroke of dynein c. **a**, Left-views of dynein c showing the mean conformation of the stem and stalk relative to the head. Arrow indicates 15 nm. **b**, Speculative structure of the molecule suggesting an origin of the power stroke and how this might be converted into microtubule movement. The first four AAA modules (of 6) and C-terminal sequence (black) are indicated.

Collaborators

Hitoshi Sakakibara & Kazuhiro Oiwa, Kansai Advanced Research Centre, Communications Research Laboratory, Kobe, Japan.

Publication

Burgess, S.A., Walker, M.L., Sakakibara, H., Knight, P.J. and Oiwa, K. (2003). Dynein structure and power stroke. *Nature*, **421**, 715-718.

Funding

We acknowledge support from NIH (to J. Trinick and H. White) and BBSRC.

An essential activity at the centre of a macromolecular machine

Yulia Redko, Lily Tong, Chris Adams and Kenneth McDowall

Introduction

The decay of RNA is as important as transcription in determining levels of gene expression. In the last decade, our understanding of the process of RNA decay has increased to the point where many of the key players and steps have been identified. Major challenges now are to characterise individual components and determine the mechanisms by which their activities are coordinated and controlled. In *E. coli,* the decay of many, if not most, transcripts is initiated by RNase E, an essential ribonuclease that, at 1061 residues, is one of the largest polypeptides encoded by this organism. The N-terminal half of RNase E is sufficient for basal catalytic activity that is able to support cell viability. The C-terminal half contains sites to which other enzymes, including a 3'-exonuclease, attach in the assembly of the RNA degradosome. This macromolecular machine is arguably a major site for RNA metabolism in *E. coli*. An understanding of the workings of the RNA degradosome will undoubtedly require quantitation of the interactions between components of the complex and their substrates and the determination of catalytic parameters.

Integrating structure and function

Our lab is at the forefront of using oligonucleotide substrates synthesised using phosphoramidite chemistry to study RNase E cleavage. In the last year, we have switched to using fluorescently labelled substrates, and ion-pair reverse phase chromatography to separate reaction products. Coupled with automatic sample injection this allows relatively high throughput analysis of multiple reaction samples. As a result, we have been able for the first time to systematically optimise conditions for a polypeptide of RNase E with regard to pH and the concentrations of magnesium and salt. We chose to start with the N-terminal catalytic half of RNase E, as longer polypeptides are sensitive to proteolysis *in vivo* and as a result are more difficult to obtain as homogeneous species. Moreover, collaborators (Ben Luisi and Anastasia Callaghan, Cambridge) have been successful in obtaining crystals of this part of RNase E. These are providing a long awaited opportunity to integrate structural and functional studies of RNase E. The multimeric form of this enzyme that is active has been identified.

The conditions we have optimised for RNase E differ significantly from those used previously by us and others: We estimate that we have been able to increase the rate of RNA cleavage by at least 20 fold. This means that previous biochemical studies that investigated the relative roles of ribonucleases in various steps in RNA decay and processing, for example poly (A) removal, may have underestimated the contribution of RNase E. Consequently, further investigation may be required. We have also found that the dependence of the rate of RNA cleavage on magnesium is sigmoidal indicating that at least two metal ions are required for RNA cleavage and have determined the pK_as of unprotonated and protonated amino-acid side chains that are required for catalysis. Determining the mechanism used by RNase E will, however, require the mutagenesis of groups that are possibly involved in RNA cleavage. We decided to investigate the RNase E-catalysed

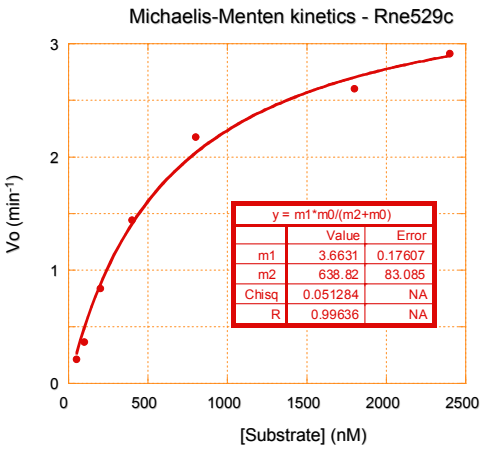
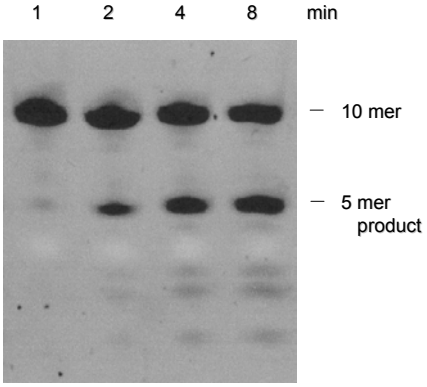


Fig. 1. Kinetics of RNA cleavage by recombinant RNase E

reaction by first modifying functional groups within RNA substrates because, unlike the active site, the targets of catalysis, susceptible bonds, have been identified. We have found that at least three functional groups on a key guanosine sequence determinant are required for efficient RNase E cleavage of RNA and can now explain why the rates of cleavage of oligonucleotides containing G to A or G to U substitutions are reduced. We have also analysed the requirement for a 2'OH on the ribose sugar directly 5' to the phosphodiester bond that is normally cleaved by RNase E and found that a ribose-like sugar pucker as well as proton sharing is required for efficient RNase E cleavage. The above biochemical analysis, although incomplete, should provide a foundation for interpreting structural models in terms of function. We are now in the process of quantifying the effects of additional domains of RNase E and structural elements within substrates on the rate of catalysis.

The regulation of RNaseE activity in response to changing growth conditions



Assay of RNase E in crude extracts

Although there have been a number of reports indicating that cleavages by RNaseE are regulated *in vivo*, there has never been a detailed analysis of the mechanisms by which RNaseE-catalysed cleavages are regulated in response to growth conditions. We have found that the positioning of fluorescein at the 3' end of our oligonucleotide substrates blocks exonucleolytic attack, and we are able to quantitatively measure RNaseE activity in crude extracts from small amounts of *E. coli* cells. Combined with Western blotting and the assay of RNaseE *in vivo* using a gene reporter system we are in the process of determining whether regulation alters the levels of RNaseE, the specific activity of the catalytic domain, or additional factors that control cleavage by the catalytic domain. Although this work is very much in progress, the initial results suggest that regulation is most likely to be at the level of controlling cleavage by the catalytic domain. More complex substrates are being synthesised in an effort to identify structural elements in RNA that may mediate this form of control.

Collaborators

Jane Grasby, Chemistry Sheffield
Vladimir Kaberdin, Vienna Biocenter, Austria
Ben Luisi, Cambridge

Funding

We thank the BBSRC and the Royal Society for funding

Cofactor processing in galactose oxidase

R. Hurtado Guerrero, Susan J. Firbank, Peter F. Knowles, Simon E.V.Phillips,
Malcolm A. Halcrow and Mike J. McPherson.

There is increasing realisation that enzymes adopt a variety of strategies to increase their catalytic potential. In some cases exogenous cofactors are recruited including NAD(P), FMN, FAD and pyridoxal phosphate. An increasing number of enzymes are being discovered that contain cofactors derived from endogenous amino acids that are post-translationally modifed either by oxidation or by formation of cross-links. The copper-containing enzyme galactose oxidase has a post-translationally derived thioether bond between the active site residues Cys228 and Tyr272. This feature is critical for enzyme activity as the site of a radical with substantially reduced redox potential.

Associated with the formation of this thioether bond is an autocatalytic cleavage of an N-terminal prosequence. These processing steps are dependent upon the presence of copper. Initially to study these processing events we have focused on a form of galactose oxidase lacking the prosequence. This protein has been purified under coopper-free conditions from *Pichia pastoris.* Formation of the cross-link is conveniently monitored by an SDS-PAGE assay. Radical formation is measured spectrophotometrically. Our studies have revealed that the rate of cross-link formation is the limiting step in generating the radical-containing enzyme, which only forms in the presence of oxygen. Under anaerobic conditions the cross-link can still form, but there is no radical produced, although this form of GO is competent for one electron oxidation by other inorganic or organic oxidants. Crystallographic studies have provided structures of precursor and processed forms of galactose oxidase, that provide a basis for describing processes involved in thioether bond formation. In particular there is evidence from structural and spectroscopic studies for a copper(II)-Cys228 species that may represent an intermediate in processing. Our attention is now focused on prosequence cleavage and the relationship between this event and thioether bond formation.

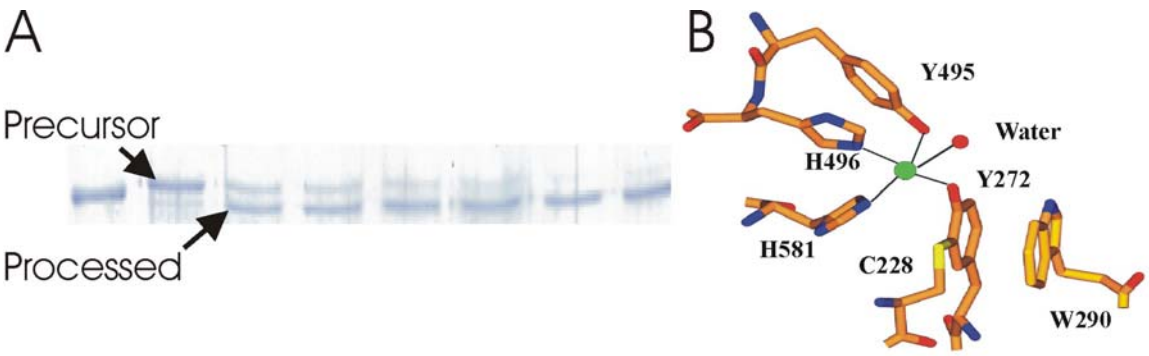


Fig. 1 Processing of galactose oxidase following addition of copper to the precursor copper-free enzyme. **(A)** SDS-PAGE assay for thioether bond formation **(B)** X-ray structure of active site of processed galactose oxidase

Collaborators

This project represents a collaboration between Leeds and the group of Professor D M Dooley in Montana State University, Bozeman, USA.

Funding

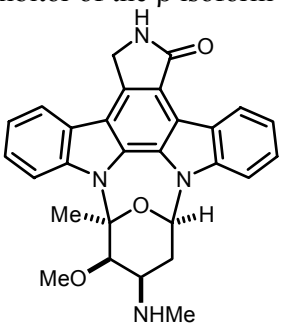
This work was funded by the BBSRC

Macrocyclic bisindoylmaleimides: Synthesis, conformation and potential as tools for studying protein kinases

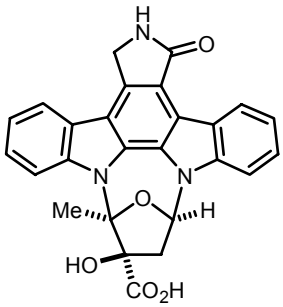
Stephen Bartlett and Adam Nelson

Introduction

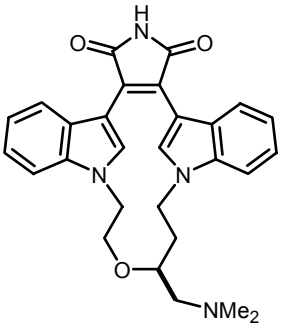
Protein kinases constitute approximately 1.7% of all human genes, and the indolocarbazole alkaloids, which target the ATP binding site, are inhibitors of many of these enzymes. For example, staurosporine (**1**) and K252a (**2**) are potent inhibitors of protein kinase C. Intringingly, by breaking the planarity of the indolocarbazole core, as in **3**, it is possible to discover compounds which are selective for particular kinases: **3**, for example, is a selective inhibitor of the β isoform of protein kinase C.



Staurosporine, 1
IC_{50} 2.7 nM



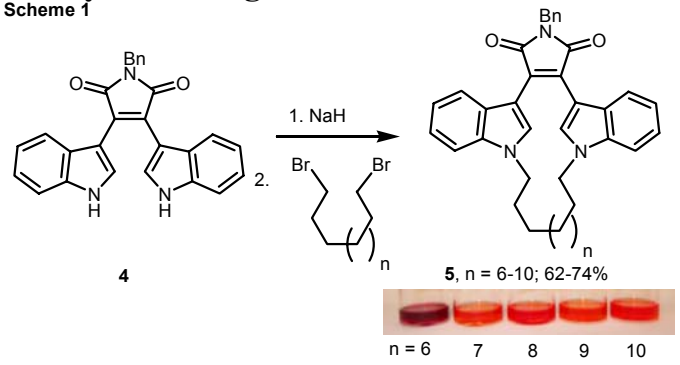
K252a, 2
IC_{50} 32 nM

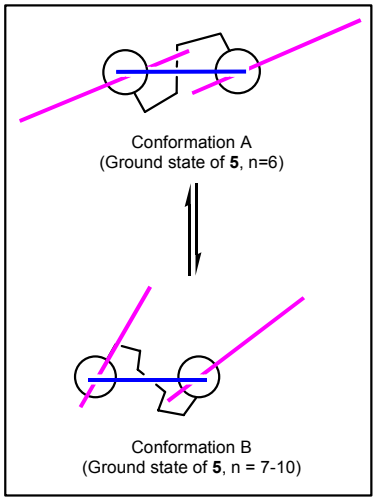


3
IC_{50} 5 nM (PKC β)

Macrocyclic analogues of the indolocarboazole alkaloids

Scheme 1





We have prepared a series of macrocyclic bisindoylmaleimides (Scheme 1), either by alkylation (→ **5**; n = 6-10) or by ring-closing metathesis (→ **7**). The u.v. spectra of these compounds are remarkably different [λ_{max} for **5** (n=6) is at 35 nm longer wavelength than **5** (n=7-10)], and these changes are reflected in the different conformational preferences of the individual compounds. X-ray crystallographic analysis has revealed that **5** (n=6) adopts an approximately C_2-symmetric conformation (Conformation A) and **5** (n=7-10) adopt an alternative conformation (Conformation B). We have deduced the regions of conformational

space which may be populated by these compounds (θ and ø are dihedral angles for the two aryl-aryl bonds), and these ideas may be summarised in a Ramachandran-type plot (Fig. 1).

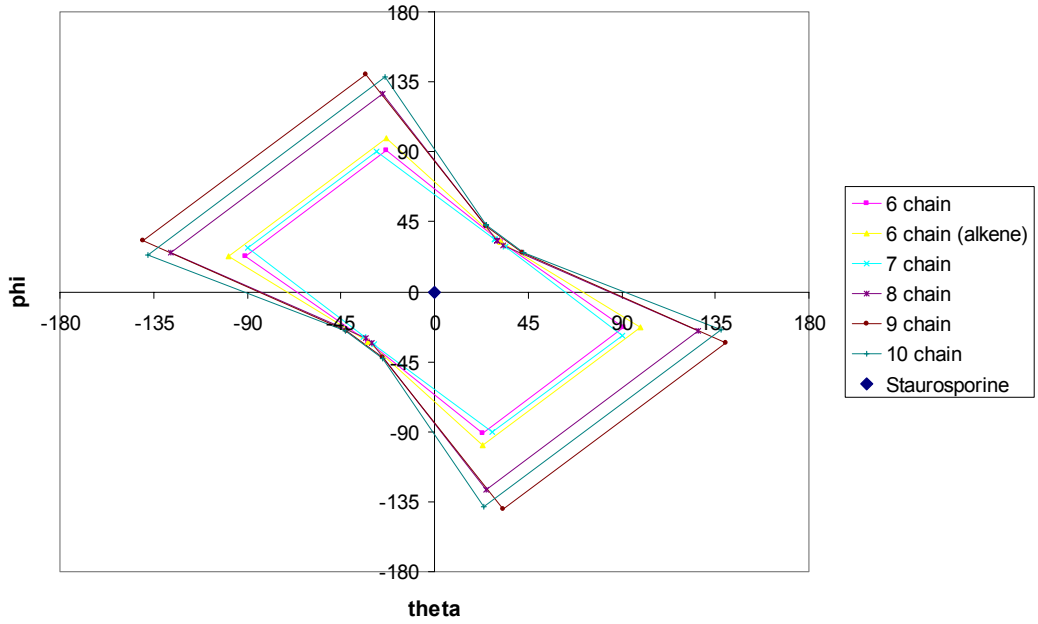


Fig. 1: Ramachandran plot describing the populated conformers of bisindoylmaleimide cyclophanes and the pathways for their interconversion

Summary

We have prepared a series of potential fluorescent ligands for protein kinases which access regions of conformational space which cannot be populated by the indolocarbazole alkaloids. It is expected that these flexible compounds may be used (a) to identify the preferred bound conformation of different kinase enzymes, and (b) as tools to monitor conformational changes taking place within the active site.

Acknowledgements

We thank EPSRC for a project studentship, and Pfizer and AstraZeneca for strategic research funding.

Crystal structure of a catalytically impaired mutant of T7 endonuclease I (K67A) that retains the ability to bind metal ions

Jonathan M. Hadden and Simon E.V. Phillips

Homologous genetic recombination is important in DNA repair, resolving stalled replication forks, and the creation of genetic diversity. The four-way (Holliday) DNA junction is the central intermediate in this process, and is the substrate for the junction-resolving enzymes. These enzymes are basic, dimeric nucleases that are selective for the structure of the four-way junction, and cleave two strands to convert the branchpoint into nicked duplex species. They have been found in a wide variety of organisms, from bacteriophage-infected *Escherichia coli* to mammals and their viruses.

Bacteriophage T7 DNA undergoes genetic recombination during infection. The phage-encoded junction-resolving enzyme is endonuclease I. Mutants in the gene encoding this enzyme are deficient in recombination.

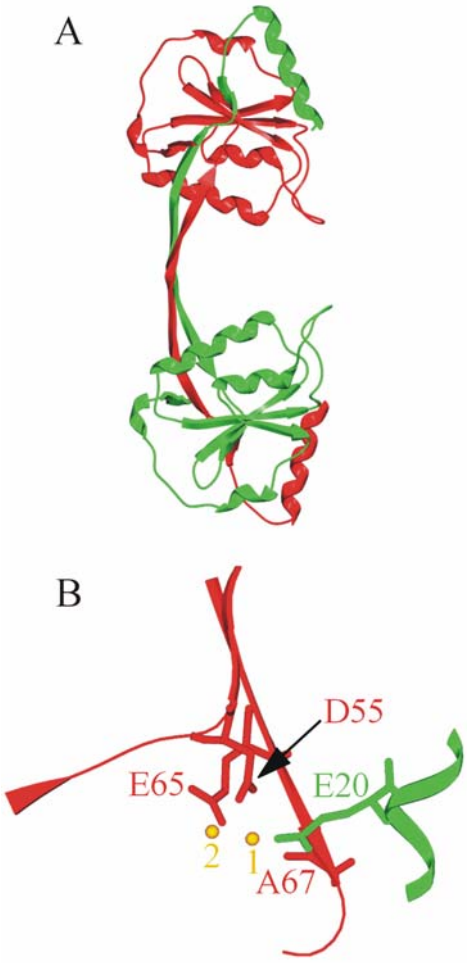


Fig. 1: The structure of T7 endonuclease I. Residues from one subunit are shown in green and from the other in red.
A The overall topology of the E65K mutant.
B Active site residues in the K67A mutant. Bound manganese ions are depicted in yellow.

We recently presented the crystal structures of a catalytically impaired mutant of endonuclease I (E65K) without metals bound (Fig. 1A) and the wild-type protein with metals bound. The active site of the protein contains three acidic side chains (Glu-20, Asp-55 and Glu-65) and a lysine (Lys-67). These correspond closely to the active site residues of a number of type II restriction enzymes, particularly *Bgl*I and are involved in the ligation of two metal ions.

In the present work we have studied the structure of another mutant, K67A. This mutant is significantly different from E65K in that it retains the ability to bind metal ions but is still significantly catalytically impaired. By studying the structure of this mutant we hope to be able to study the mechanism of DNA cleavage in more detail.

The overall structure of K67A is very similar to that of both the wild-type protein and the E65K mutant. However, electron density maps clearly showed that that lysine 67 had been substituted by alanine. Diffusion of manganese ions into the crystals of K67A revealed two peaks of electron density per active site, defining two metal ion binding sites (Fig. 1B). These peaks were situated in similar positions to the calcium ions found in the *Bgl*I structure.

Refinement of the structure is still ongoing and at this stage it is not possible to make any further detailed comparisons between the three available structures. Nevertheless the K67A mutation would appear to have made no significant global changes to

the structure of the enzyme and the protein is undoubtedly still able to bind two metal ions. It would appear that local electrostatic changes in the vicinity of the active site are responsible for the observed catalytic impairment.

Ultimately the structure of an endonuclease I - DNA complex will probably be required before we can conclusively identify the full mechanism by which endonuclease I hydrolyses specific phosphodiester bonds within the four-way DNA junction. In the absence of such information the structure of the K67A mutant will undoubtedly shed more light on the mechanism of action of T7 endonuclease I.

Publications

Hadden, J.M., Convery, M.A., Déclais, A-C., Lilley, D.M.J. & Phillips, S.E.V. (2001). Crystal Structure of Bacteriophage T7 Endonuclease I: A Holliday Junction Resolving Enzyme. *Nature Structural Biology*, **8**, 62-67.

Hadden, J.M., Déclais, A.-C., Phillips, S.E.V. and Lilley, D.M.J. (2002). Metal ions bound at the active site of the junction-resolving enzyme T7 endonuclease I. *EMBO J.*, **21**, 3505-3515.

Déclais, A.-C., Hadden, J.M., Phillips, S.E.V. and Lilley, D.M.J. (2001) The active site of the junction-resolving enzyme T7 endonuclease I. *J. Molec. Biol.*, 307, 1145-1158.

Déclais, A.-C, Fogg, J.M., Freeman, A., Coste, F., Hadden, J.M., Phillips, S.E.V. and Lilley, D.M.J. (2002) *EMBO J.*, The complex between a four-way DNA junction and T7 endonuclease I. *In the press*.

Collaborators

A.-C. Déclais, D.M.J. Lilley Cancer Research UK, Nucleic Acid Structure Research Group, Department of Biochemistry, University of Dundee, DD1 4HN, UK.

Funding

We are grateful to the Wellcome Trust and the Cancer Research Campaign for financial support.

Crystal structures of MS2 and Qβ RNA stemloop operators complexed with a bacteriophage MS2 coat protein mutant.

Wilf T. Horn, Nicola J. Stonehouse, Peter G. Stockley and Simon E.V. Phillips

Introduction

MS2 and Qβ are evolutionarily related *T=3* icosahedral bacteriophages with single stranded RNA genomes that infect *E. coli*. Although the protein subunits of the MS2 and Qβ capsids share less than 25% sequence identity, the structures of the Qβ subunits are very similar to those of MS2. Subunits of both the MS2 and Qβ capsid shells exist as three distinct conformers (A, B and C) that associate to form AB and CC dimers which comprise the basic building blocks of both capsid shells. The structures of the MS2 and Qβ capsids have been determined via X-ray crystallography by our collaborators in Uppsala, Sweden.

The two bacteriophages both utilise a similar mechanism of translational repression. *In vivo,* a small RNA stemloop within the viral genomes binds to a specific site on a coat protein dimer, acting to inhibit viral replicase gene translation. The translational complex of MS2 (Fig. 1) has, for many years, been the paradigm for studying RNA/protein interactions at the atomic level.

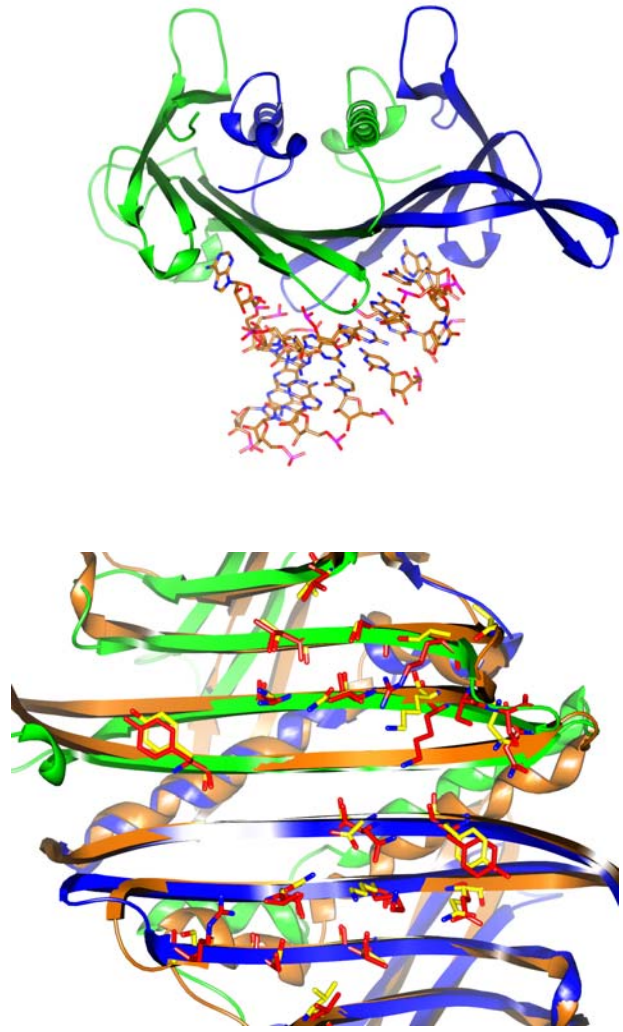


Fig. 1
(Top) Structure of the WT MS2 RNA stemloop-coat protein dimer complex, protein subunit A (blue) and subunit B (red). (Bottom) Ribbon diagram of the ten-stranded β sheet of an AB coat protein dimer of the Qβ phage dimer (in orange) superimposed on an AB dimer from the MS2 capsid (subunit A in blue, B in green). Amino acid residues that are involved in protein RNA interactions are shown in stick format, with MS2 residues in red and their Qβ equivalents in yellow.

The RNA stemloop operator binding site is located on a 10 stranded β sheet formed by AB and CC dimers within the capsid shells of both MS2 and Qβ. Many of the amino acid residues that have previously been shown to be important for high affinity binding of MS2 and Qβ stemloops are conserved between the two bacteriophages (Fig. 1). Although the protein surfaces of the stemloop binding site of the two bacteriophages display considerable similarity, profound differences exist in the sequence and secondary structures of the two stemloop operators (Fig. 2).

In vivo, each bacteriophage preferentially discriminates against binding the stemloop operator of the other. Affinity binding studies have, however, identified specific coat protein mutants of MS2 that overcome this discrimination mechanism, some of the mutations allowing the binding of the Qβ RNA operator to MS2 mutant capsids with an affinity comparable to that of the wild type MS2 operator. In order to gain new insight into this discrimination mechanism, Qβ RNA stemloop operators were soaked into pre-crystallised MS2 mutant capsids and the structure of the capsid/RNA complexes determined via X-ray crystallography.

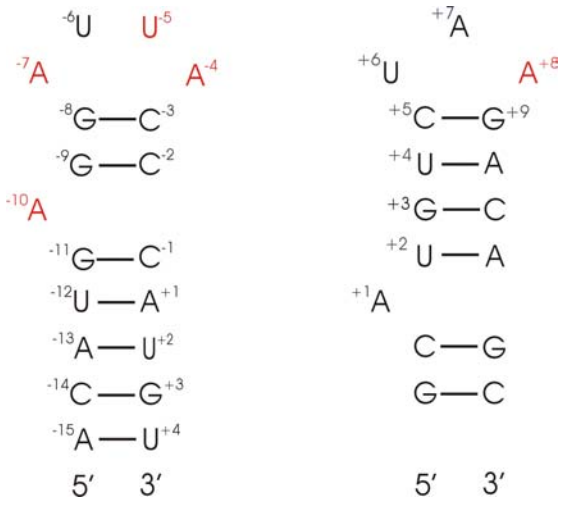


Fig. 2.
Diagram of the secondary structures of the MS2 (left) and Qβ (right) stemloop operators. Residues that have been shown to be important for high affinity binding to capsid protein are highlighted in red.

Results

Thus far, crystallographic studies have concentrated on a single mutation (N87S) within the MS2 coat protein, a mutation that decreases the capsids affinity for the MS2 stemloop operator whilst increasing affinity for the Qβ operator. Data were collected for the RNA free MS2 mutant capsid and MS2 mutant capsids soaked with wild type MS2 and Qβ stemloop operators at the SRS, Daresbury, UK.

Electron density maps demonstrated that there was little change in the mode of binding of the MS2 operator with the exception of two novel water mediated hydrogen bonds between the N87S mutation and the RNA operator. Electron density for the Qβ operator was weak in the N87S mutant/Qβ RNA complex, however, a tentative model for Qβ binding has been proposed. Further structural determinations using other MS2 mutant capsids are underway to test the validity of this hypothesis.

Collaborator

Lars Liljas, Department of Cell and Molecular Biology, Uppsala University, Sweden

Funding

We thank the BBSRC for funding.

Regulation of membrane protein function in human blood vessel formation

Vas Ponnambalam, Lorna Ewan, Gareth Howell, Shweta Mittar,
Mudassir Mohammed and Jane Murphy

Background

The major interest in the laboratory is how membrane proteins interact with different factors to programme cell and organ development in human health and disease. Our current model system is the human endothelial cell that lines the inside of all blood vessels and regulates angiogenesis i.e. the sprouting of new blood vessels from pre-existing blood vessels. Understanding this process is critical in treating diseases such as atherosclerosis and cancer. Importantly, many diabetic patients suffer from atherosclerosis and eye diseases due to dysregulation of endothelial cell function.

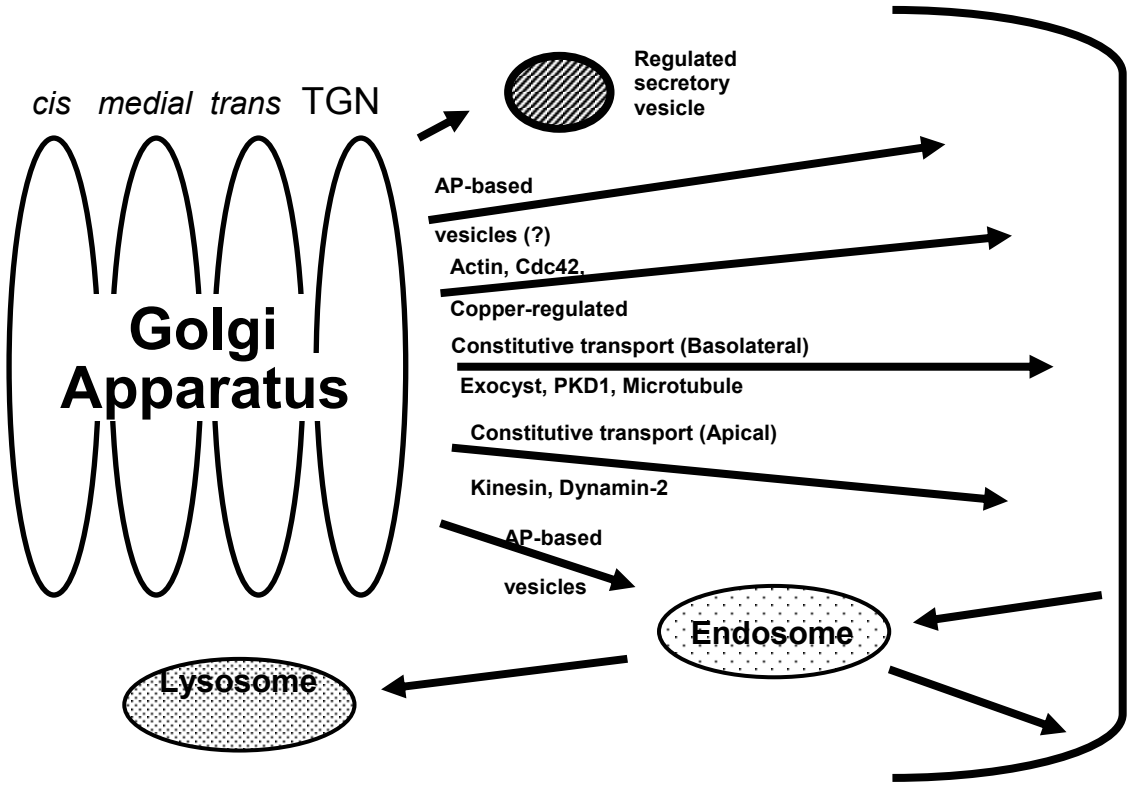


Fig. 1 Steps in membrane traffic to and from the Golgi apparatus, plasma membrane and endosome-lysosome system are under study using different model systems with a particular emphasis on human endothelial cells.

Structure of membrane proteins

Receptors on the surface of human endothelial cells bind a variety of substances including nutrients, lipoproteins and growth factors. Depending on the substance that is bound, a series of intracellular processes are triggered including sorting of receptor-ligand complexes, activation of signalling pathways, internalisation of receptor-ligand complexes, processing, sorting and receptor recycling. The goal of the laboratory is to study how such biochemical reactions can be linked to membrane remodelling in the endothelial cell in terms of blood vessel formation. Protein domains of such receptors expressed as single modules or linked to each other in bacterial and eukaryote systems will be tested using the excellent crystallisation facilities and technical support within the Astbury Centre. Large protein assemblies purified from mammalian

cells and tissues can be subjected to electron microscopy analyses, sophisticated and computerised image reconstruction to obtain a 3-D image. The information obtained from such studies will help us to better test our hypotheses for blood vessel development using site-directed mutagenesis and expression in human endothelial cells.

Analysis of conformational changes in membrane proteins

Detecting and understanding conformational changes in endothelial membrane receptors is an important at many levels including potential clinical assays for human disease situations. Increasingly, conformational changes in membrane proteins can be detected using sophisticated biophysical techniques such as NMR and fluorescence spectroscopy. NMR is a powerful and sensitive tool for detecting conformational changes of protein domains in solution. This is especially important where many membrane proteins are difficult to crystallise and solution-based interactions are critical in programming such interactions. The Astbury Centre has a state-of-art NMR facility including 500 MHz and 600 MHz NMR machines and provide excellent opportunities for such studies. The recently upgraded state-of-art spectroscopic and spectrometry facilities include Q-ToF, SELDI/MALDI/Protein Chip Reader, FT-IR, circular dichroism and fluorescence machines is also available for the study of macromolecular assemblies. In addition, atomic force microscope (AFM) systems are also being developed for detection of single molecule events in biological systems.

Collaborators

Prof. Steve Baldwin (BMB Leeds)
Dr. Francis Barr (Max-Planck Institute for Biochemistry, Munich, Germany)
Prof. Grahame Hardie (Univ. of Dundee, UK)
Dr. John Walker (BMB Leeds)
Prof. Vivek Malhotra (UC San Diego, USA)
Prof. Tony Monaco (Wellcome Trust Centre for Human Genetics, Oxford)
Dr. Ewan Morrison (Leeds Cancer Research UK Clinical Centre)
Dr. Michelle Peckham (BMS Leeds)
Dr. Asipu Sivaprasadarao (BMS Leeds)
Prof. Liz Smythe (University of Sheffield, UK)
Prof. Tony Turner (BMB Leeds)
Dr. Ian Zachary (Centre for Cardiovascular Biology and Medicine, University College London)

Publications

Cobbold, C., Coventry, J., Ponnambalam, S. & Monaco AP (2003).The Menkes disease ATPase (ATP7A) is internalized via a Rac1-regulated, clathrin- and caveolae-independent pathway. *Hum. Mol. Genet.* In Press.

Cobbold, C., Ponnambalam, S., Francis, M.J. & Monaco, A.P. (2002) Novel membrane traffic steps regulate the exocytosis of the Menkes disease ATPase. *Hum. Mo.l Genet.,* **11**, 2855-2866.

Grewal, S., Morrrison, E.E., Ponnambalam, S. & Walker, J.H. (2002) Nuclear localisation of cytosolic phospholipase $A_2\alpha$ in the EA.hy.926 human endothelial cell line is proliferation dependent and modulated by phosphorylation**.** *J. Cell Sci.,* **115**, 4533-4543.

Grewal, S., Ponnambalam, S. & Walker, J.H. (2003) Golgi localisation of cytosolic phospholipase A2α in the human A549 epithelial cell line. *J. Cell Sci*., **116**, 2303-2110.

Ponnambalam, S., Baldwin, S.A. (2003) Constitutive protein secretion from the *trans*-Golgi network to the plasma membrane. *Membr. Mol. Biol.,* In press.

Funding
We acknowledge funding by the British Heart Foundation, The Wellcome Trust and the White Rose Consortium.

Biophysical studies of β_2-microglobulin amyloid formation.

Victoria M^{c}Parland, Tali Gidalevitz, Susan Jones, Arnout Kalverda, David Smith, Clemens Stilling, Chi Trinh, Simon Phillips, Steve Homans and Sheena Radford.

Introduction

A number of proteins, that are unrelated in their sequence and native structure, undergo specific aggregation into amyloid fibrils *in vivo*, leading to the pathological disorder known as amyloidosis. These diseases are characterised by the deposition of normally soluble proteins into insoluble fibrils with a cross-β structure. Further understanding of the mechanism(s) of fibril formation from these globular proteins is vital for developing inhibition strategies.

The focus of our research is to gain new information on the mechanism of fibril formation from human beta-2-microglobulin (β_2m). Fibrils of wild-type β_2m are responsible for the severe and relatively common disease, haemodialysis related amyloidosis, which affects all patients with renal failure.

Non-native states as precursors to β_2m amyloid fibril formation

In previous studies, we have shown that disruption of the native beta-sandwich structure is a critical first step in amyloid formation for wild-type β_2m, as for other amyloidogenic proteins. This implicates the requirement for non-native states in the initiation of fibril formation, at least *in vitro*. We have used a variety of biophysical techniques, including circular dichroism, infrared spectroscopy, fluorescence, NMR and X-ray crystallography to identify and structurally characterise monomeric non-native conformations that represent candidate precursor conformations in fibril formation.

Structural properties of an amyloidogenic precursor revealed by 2D-NMR

At pH 3.6 β_2m is partially unfolded to a conformation that rapidly and spontaneously forms fibrils. By slowing the rate of fibril formation at this pH we can maintain the partially-folded protein in a monomeric state over a timescale synonymous with measurement of 2D-NMR spectra. However, at pH 3.6 the NMR signals are broadened, consistent with the formation of a dynamic conformation that precludes direct structure determination. Nevertheless, information regarding the structure and cooperativity of the partially-folded precursor state at pH 3.6 was determined indirectly using urea-titration experiments. By monitoring the non-cooperative unfolding of partially folded

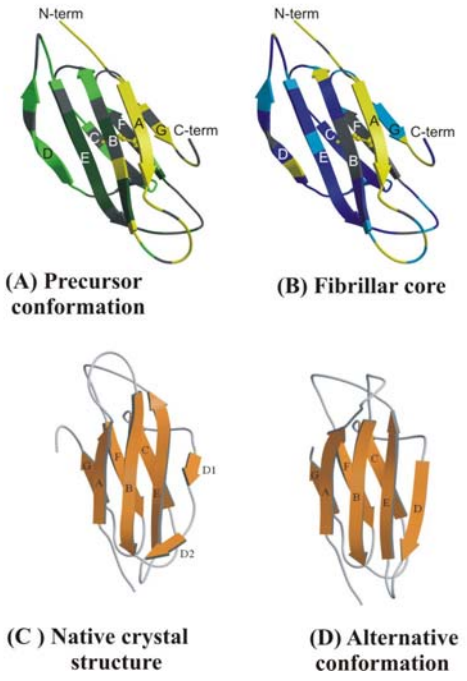


Fig. 1 Schematic representation of non-native states of β_2m. Ribbon diagram of native β_2m coloured according to the regional stability in a monomeric precursor at pH 3.6 (A) and according to the core structure in the assembled fibrils (B) (light to dark shades in (A) and (B) indicate more structured/more protected, respectively). The crystal structure of native β_2m with a beta bulge in strand D (C) compared to the alternative conformation without the beta-bulge in strand D (D).

β_2m residue-specifically by 2D-NMR we have shown that this state retains structure in six out of the seven native β-strands and that there are distinct regions of the protein with different stabilities. The most stable region of this conformation maps to β-strands B, C, E and F and the N-terminal strand is the most denatured. The data therefore suggest that the structured elements in partially folded β_2m are packed around a hydrophobic core that contains at least remnants of the native structure. Interestingly, the structural features identified in this monomeric precursor conformation are mirrored in fibrils formed from wild-type β_2m, as described in Hoshino *et al*. Furthermore, a peptide that corresponds to the core structure of partially-folded β_2m has been synthesised and shown to form fibrils in isolation and also to effect fibril formation of the full length protein when co-incubated.

Identification of an alternative state by crystallography.
An alternative state has been identified by crystallisation of β_2m at pH 5.6. The structure was solved to 1.8Å and some structural differences to the native structure at pH 7.0 were observed. The most significant alterations occur in β-strand D that forms an edge strand of the native β sandwich structure. In the native state in solution, there is a β bulge within β-strand D, whereas in the crystal structure the β-bulge is lost and β-strand D now forms a contiguous β-strand on the edge of the β-sandwich. The presence of β-bulges in edge strands of beta sandwich proteins has been assigned a protective role in prevention of aggregation, by disrupting potential intermolecular hydrogen bonding surfaces. Therefore we speculate that this alternative conformation at pH 5.6 that is devoid of the β bulge in the edge β-strand D may represent a more aggregation prone state, and therefore another candidate amyloidogenic conformation.

Publications
McParland, V. J., Kalverda, A. P., Homans, S. W. and Radford, S. E. (2002) Structural properties of an amyloid precursor of beta(2)-microglobulin. *Nat Struct Biol* **9**, 326-31.

Jones, S., Manning, J., Kad, N. M. & Radford, S. E. (2003) Amyloid-forming peptides from beta2-microglobulin:-Insights into the mechanism of fibril formation *in vitro*. *J. Mol. Biol*. **325**, 249-257

Trinh, C. H., Smith, D. P., Kalverda, A. P., Phillips, S. E. and Radford, S. E. (2002) Crystal structure of monomeric human beta-2-microglobulin reveals clues to its amyloidogenic properties. *Proc Natl Acad Sci U S A* **99**, 9771-6

Funding

We gratefully acknowledge the University of Leeds, BBSRC, EPSRC and The Wellcome Trust for financial support. SER is a BBSRC Professorial Research Fellow. NHT is an EPSRC Advanced Research Fellow.

Characterisation of amyloid fibrils by atomic force microscopy

Walraj Gosal, Sarah Myers, Sheena Radford, Alastair Smith and Neil Thomson.

Self-assembly processes in biological systems are prevalent and important for normal physiology. However, the self-assembly of a number of proteins forming amyloid fibrils (β-sheeted linear aggregates) has been implicated in various pathological conditions and collectively termed amyloidosis. In addition, a few proteins have been identified in yeast where amyloid-like fibre formation is associated with protein-only ('prion') inheritance. Atomic force microscopy (AFM), which probes samples bound to very flat surfaces with a fine probe, allows this association process to be followed *in vitro*, on the nanometer scale. The ability of AFM to function under fluid (as well as air), allows experiments to be carried out where the environment is carefully controlled (*i.e.* pH, ionic strength and the addition of biological co-factors) and enables processes to be followed in real time. Through control of environment and surface chemistry we hope to be able to mimic aspects of growth conditions *in vivo* to assess the kinetics of assembly, the resultant fibrillar structures, and also to relate these to the diseased state.

Self assembly of β_2-microglobulin into amyloid fibrils

The self-association of β_2-microglobulin (β_2m), a 99 residue all β-sheeted protein, is involved in a condition known as haemodialysis-related amyloidosis, which affects patients with renal failure. Fibrils can be formed *ab initio in vitro* at low pH, through the population of a partially denatured state. These amyloid fibrils are mostly constructed from filaments or protofibrils, which then laterally associate and twist to form 'higher-order' fibrillar structures such as twisted ropes and ribbons. We have recently attempted to gain an insight into the hierarchy of β_2m fibril assembly, through imaging of samples at various times during the aggregation process, and coupling these data to a thioflavin-T fluorescence assay. Thioflavin T binds to amyloid fibrils increasing fluorescence and is commonly used to follow the kinetics of growth *in vitro*. These data indicate that large amorphous aggregates are present in the nucleation phase, and the formation of fibrils coincides with a sharp increase in thioflavin-T fluorescence. Furthermore, mature fibrils at the end of growth can have at least three different morphologies. For example, a type III fibril (Fig. 1A & B) is composed of four protofibrils laterally associated (a twisted ribbon).

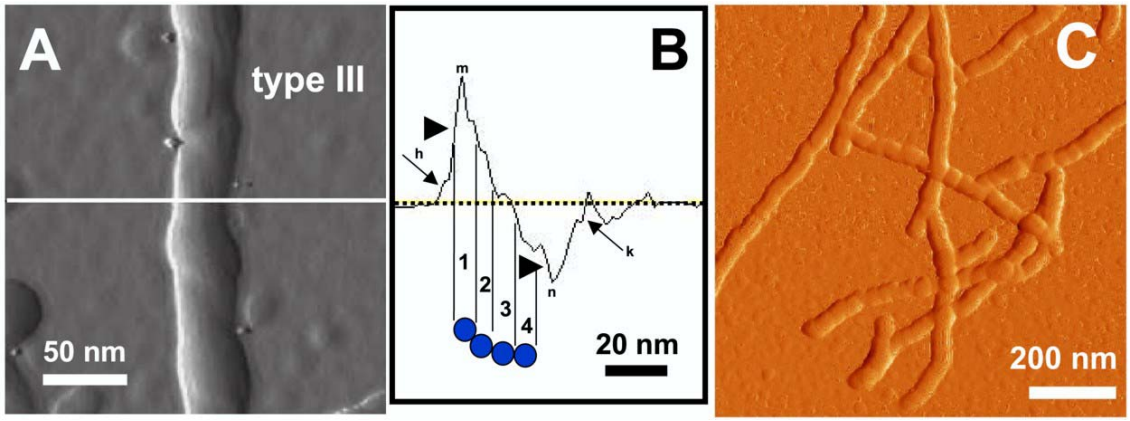


Fig. 1. (A & B) The protofilaments within a type III β_2-microglobulin fibril are resolved in the AFM amplitude signal. The model of the amyloid fibril in blue, shows four protofibrils, with diameter ~7 nm. Markers 'h' and 'k' indicate where sideways interactions of the AFM tip with the fibril have modulated the cantilever amplitude; points 'm' and 'n' indicate where the feedback loop regains control after the tip encounters a high feature on the sample, and the large horizontal arrowheads indicate where the amplitude is changing rapidly on initial contact with and leaving contact with the fibril. (C) Fibrils formed by the yeast protein Ure2p, at the end-point of a thioflavin-T fluorescence assay.

Fibrils formed from the yeast prion protein Ure2p

Conformation change, and subsequent fibril formation, of the yeast protein Ure2p is involved in the non-Mendelian inheritance of the [URE3] phenotype (the ability of cells to take up ureidosuccinate). This unusual characteristic (*i.e.* the ability of certain conformations of Ure2p to convert normal protein into an inactive state), underlies it's 'prion' functionality. Recently, we have revealed that this largely helical protein spontaneously forms long fibrils at neutral pH *in vitro* (Fig. 1C). Remarkably, our data suggest that full-length Ure2p retains its helical structure (as indicated by deconvoluting FTIR absorbance spectra), and ability to bond substrate analogues, upon fibre formation. We present a model whereby Ure2p fibril formation is based on the self-assembly of native-like monomers, driven by interaction between the N-terminal glutamine and asparagine-rich regions and the C-terminal functional domain. Studies from other groups show that when the N-terminal region is cleaved from Ure2p it rapidly forms fibres with a cross-beta sheet structure, typical of amyloid. This contrasts with our experiments, which show that a cross-beta structure is not formed when the full-length protein assembles into fibres.

Future studies

Future work will concentrate on the imaging of the assembly of proteins under aqueous fluid conditions by AFM, to assess nucleation mechanisms, polarity of fibril growth, and relative rates of assembly under the influence of various biologically relevant co-factors. AFM will also be able to probe the mechanical integrity of fibrils to investigate relationships between morphology, kinetics and stability.

Collaborators

Luc Bousset and Ronald Melki, CNRS, France.

Publications

Kad, N.M., Smith, D.P., Myers, S.L., Smith, D.A., Radford, S.E. & Thomson, N.H. (2003). Hierarchical assembly of β_2-microglobulin amyloid *in vitro* revealed by atomic force microscopy. *Journal of Molecular Biology*, *submitted*.

Bousset, L., Thomson, N.H., Radford, S.E. & Melki, R. (2002). The yeast prion Ure2p retains its native alpha-helical conformation upon assembly into protein fibrils *in vitro*. *EMBO Journal*, **21**, 2903-2911.

Acknowledgements

We gratefully acknowledge the University of Leeds, BBSRC and EPSRC for financial support. SER is a BBSRC Professorial Research Fellow. NHT is an EPSRC Advanced Research Fellow.

Characterising the factors that describe the mechanical resistance of proteins

Anthony Blake, David Brockwell, Peter Olmsted, Sheena Radford, Alastair Smith and Rebecca Zinober.

Introduction

Many proteins are designed to withstand, or react to, an applied force as part of their function. For example, proteins with an immunoglobulin (Ig) or Ig-like fold form part of the giant muscle protein titin to prevent over extension of the sarcomere. Domains with similar structures are found to act as force sensors in the extracellular matrix protein fibronectin, and in tenascin, to support and guide cell-rolling in migrating cells. Furthermore, it is becoming clear that a protein's resistance to mechanical stress is of fundamental importance in protein translocation and degradation. The development of Atomic Force Microscopy (AFM) and laser tweezer technologies to routinely allow the nanomanipulation of biomolecules has allowed the mechanical properties of these proteins to be directly measured on simplified systems at a single molecule level. Despite the wealth of data obtained by AFM studies on model Ig systems (most commonly the 27th Ig domain from titin's I-band) the generic features that endow mechanical resistance onto any protein fold are still unclear.

What factors affect the mechanical stability?

We have previously shown that mutation of the amino-acid sequence of I27 has different effects upon the barrier to unfolding by chemical denaturation or by extension of its termini. Furthermore the mechanical properties were only affected by mutations in a localised region previously thought to confer mechanical resistance onto a protein. We have investigated this observation further by altering the solvent in which the experiment is performed and comparing the effects in each type of unfolding experiment. Changing the pH and introduction of 8% (v/v) TFE was found to affect both processes in a similar manner. Addition of 0.4M sodium sulphate, however, stabilised I27 to chemical unfolding but had no effect on its mechanical stability. These data suggest that mechanical unfolding, unlike that induced by denaturant, is a local rather than global co-operative effect.

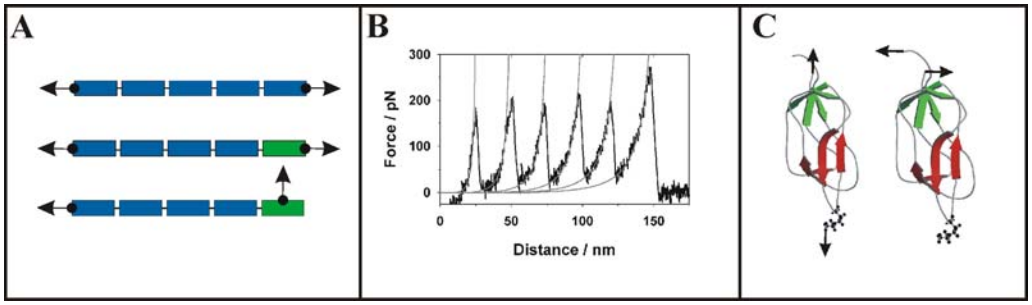


Fig. 1. (A) Concatamers used for mechanical unfolding are modular. Our experimental system $(I27)^*_5$ contains five copies of I27 (blue rectangles) constructed by ligating five cassettes of the gene together. Each cassette is defined by a unique pair of restriction sites 5' and 3' to each gene. The properties of this system have been examined. **(B) A sample force-extension profile for $(I27)^*_5$.** Each monomer of I27 unfolds at ~200pN (the last peak is due to the fully extended polypeptide chain breaking away from either the tip or substrate). The rising edge of each peak is well described by a function that describes the stretching of an unstructured polymer (the worm-like chain model, grey line). **(C) E2lip3 can pulled in different directions.** When present as the last cassette in an $(I27)_4$ scaffold, E2lip3 (green rectangle, centre (A)) is subjected to a force that peels apart the N and C-terminal strands (right image, (C)). Under this extension geometry E2lip3 unfolds at forces within the noise of the experiment (<15 pN). However when extended via its N-teminus and the lipoylated sidechain of K41 (bottom concatamer (A), left image (C), E2lip3 unfolds at a force similar to that of I27 (~170 pN at 700 nms^{-1}).

We have started to address this question by pulling a protein domain novel to the mechanical unfolding field. We have pulled the inner lipoyl domain of the dihydrolipoyl acetyltransferase subunit (E2p) of the pyruvate dehydrogenase (PDH) multienzyme complex from *Escherichia coli* (E2lip3) in two different directions (see Fig. 1). Initial results show a striking anisotropy in mechanical response. When extended *via* its termini E2lip3 unfolds at a force within the noise of the experiment (<15 pN). However, when extended by the N-terminus and the side-chain of K41, E2lip3 shows a similar resistance to I27. Mechanical resistance is thus unrelated to a proteins thermodynamic or kinetic stability but is intimately related to the geometry of the applied force vector relative to the secondary structural elements of the protein.

The parameters which describe a protein's mechanical properties are not obtainable directly from the experiments. Instead experimental data are fitted iteratively to a Monte Carlo simulation of the process. This work revealed an unexpected result: the mechanical properties of each monomer in the polyprotein are affected by the size and compliance of the rest of the domains in the polymer. Mechanical unfolding occurs one domain at a time. Therefore, when mechanically unfolding all five domains in our experimental system, the number of folded domains decreases with time (causing the unfolding force to increase) but the rate at which force is applied onto the remaining domains decreases (causing the unfolding force to decrease). The net result of these effects is to cause a non-linear event number dependence on the mechanical resitance on otherwise identical domains. This effect can also be seen experimentally.

A protein's mechanical resistance is thus controlled by the geometry of its termini relative to secondary structural elements, and is in turn modified by the sequence of this motif. However, each domain in turn is also modulated by subtle factors including the number and length of domains present in the polymer, the mechanical stability of these domains and the presence of unstructured regions. Variation of these parameters has thus been used to generate the wide variety of mechanical responses seen in Nature.

Collaborators
Godfrey Beddard, School of Chemistry, University of Leeds.
Peter Olmsted, Department of Physics, University of Leeds.
Emanuele Paci, University of Zürich, Switzerland.
Richard Perham, Department of Biochemistry, University of Cambridge.

Publications
Brockwell, D.J., Beddard, G.S., Clarkson, J., Zinober, R.C. , Blake, A.W., Trinick, J., Olmsted, P.D., Smith, D.A., and Radford, S.E. (2002). The effect of core destabilisation on the mechanical resistance of I27. *Biophys. J.* **83**, 458-472.

Zinober, R.C., Brockwell, D.J., Beddard, G.S., Blake, A.W., Olmsted, P.D., Radford, S.E. and Smith, D.A. (2002). Mechanically unfolding proteins: The effect of unfolding history and the supramolecular scaffold. *Prot. Sci.* **11**, 2759-2765.

Acknowledgements
We thank Keith Ainley for technical support and the Wellcome Trust, the University of Leeds, the BBSRC and EPSRC for funding. SER is a BBSRC Professorial Research fellow.

Transition states and intermediates in the folding of the four-helix bundle proteins Im7 and Im9

Andrew Capaldi, Claire Friel, Graham Spence, Stuart Knowling, Susanne Cranz, Bill Ashraf and Sheena Radford

Introduction

Most small, single domain proteins fold spontaneously from a highly disordered unfolded ensemble to their well-defined native states in a matter of seconds. Understanding how a protein correctly folds is of major importance as it will help to reveal how the folding process is directed by the amino acid sequence, and will hopefully shed light on the misfolding processes which underpin many diseases. To fully understand how a protein folds, we need to know the structures of each state on the folding pathway, i.e. the native and unfolded ensembles, high-energy transition states, and transiently populated intermediate species, and how these species interconvert. We are approaching this problem by studying two highly homologous four-helix bundle proteins, Im7 and Im9, which fold on similar time-scales. However, Im7 and Im9 appear to fold *via* pathways of different complexities. Im7 is known to fold through a compact, rapidly populated, on-pathway intermediate, whereas Im9, under identical conditions, appears to fold without populating an intermediate species.

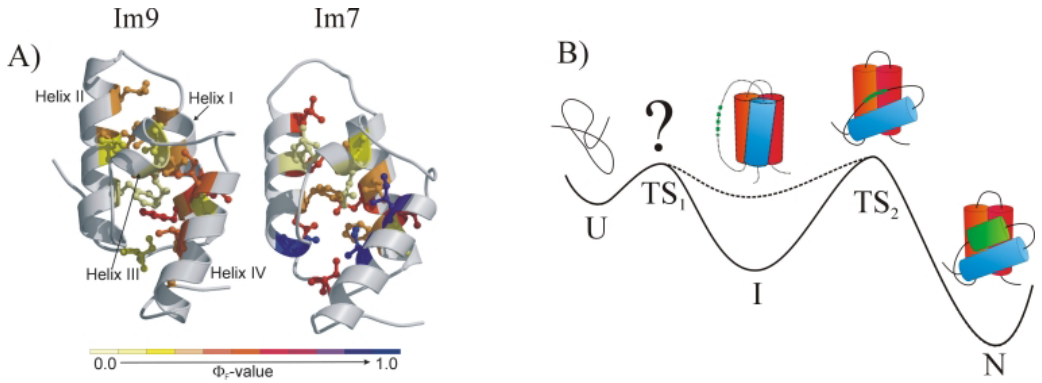


Fig. 1. (A) shows Φ-values for the rate-limiting transition states of Im9 and Im7 mapped onto the native structures. Data are shown only for mutations which remove atoms buried/partially buried in the native state. (B) shows a schematic diagram of the folding pathway of Im7 and Im9, showing the unfolded state (U), kinetic intermediate (I), native state (N), and the two transitions states, (TS1 and TS2), of which TS2 is rate-limiting.

Comparing the rate-limiting transition states of Im9 and Im7

The structures of the rate-limiting transition state ensembles for the folding of both Im7 and Im9 have been mapped using Φ-value analysis. This method uses a combination of the powerful techniques of site-directed mutagenesis and kinetic analysis to determine the structure of species that are transiently populated during folding. Despite the fact that Im7 and Im9 fold by different kinetic mechanisms, their rate-limiting transition state ensembles are found to be highly homologous: the pattern of Φ-values determined for both proteins is qualitatively very similar (Fig. 1A). For both Im7 and Im9 the transition state is a species in which helices I, II and IV are formed and specifically docked but helix III is absent. Differences between the transition state ensembles of these proteins are uncovered by making a quantitative comparison of the Φ-values determined for this species. Φ-values for the transition state of Im7 are generally higher than those for Im9 (Fig. 1A). This implies a transition state ensemble for Im7 that is restricted relative to that of Im9. Molecular dynamics simulations (in collaboration with Karplus, Paci, and Vendruscolo) using the experimentally determined Φ-values indeed result in a transition state ensemble for Im7 that is narrower than that of Im9. This restriction of the Im7 transition state ensemble may be a

consequence of folding through an on-pathway populated intermediate. Further studies are currently underway to determine the effect that increasing the stability of an intermediate has on the breadth of the subsequent transition state ensemble.

Kinetic and equilibrium folding intermediates

Structural information about the kinetic intermediate that is transiently populated during Im7 folding (Fig. 1B) has been obtained using Φ-value analysis. This species is suggested to be significantly misfolded, and comprises the three long, stable helices I, II, and IV docked in a non-native manner. At TS2, the rate-limiting transition state, helices I and II have rotated outwards, and helix IV is lying across the open end of this hairpin. Once the transition state is traversed, the binding site for helix III is revealed, allowing it to dock on to the rest of the structure, locking the native structure in place (Fig. 1B). Further insights into the structure of the Im7 folding intermediate are now being determined using equilibrium experiments, where a wide range of biophysical techniques can be employed. Disrupting the interaction of helix III with the rest of the structure proved key to trapping the on-pathway folding intermediate at equilibrium. A number of mutants have been created, all predicted to be trapped at the intermediate, by removing one or more residues involved in binding of helix III to the rest of the structure. A number of thermodynamic and spectroscopic parameters are known, or have been predicted, for the kinetic intermediate, based on stopped-flow fluorescence data. Of the four mutants tested against these parameters, two compared almost exactly, having the same stability, compactness, fluorescence, and helical content, as the kinetic intermediate. The remaining two mutants were the same in all but their fluorescence. This series of mutants, which populate an equilibrium state that is the same as the kinetic intermediate in every way tested, provide a unique opportunity to determine the structure of a transiently populated protein folding intermediate at high resolution.

Other ongoing studies on Im7 and Im9 include investigating the role of helix stability in determining the folding pathway of Im9. Based on predictions of helical propensity, a number of Im9 mutants have been made in which individual helices have been made more stable. The effects on the folding kinetics and equilibrium stability are being measured. In Im7, the same process is being applied to helix III. Increasing the stability of this helix will allow us to determine if the intermediate observed during Im7 folding is obligate.

Collaborators

Martin Karplus, Harvard University.
Colin Kleanthous, University of York.
Geoff Moore, University of East Anglia, Norwich.
Emanuele Paci, University of Zurich.
Michele Vendruscolo, Cambridge University.

Publications

Capaldi, A.P., Kleanthous, C. & Radford, S.E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nature Structural Biology*, **9**, 209-216.

Friel, C.T., Capaldi, A.P. & Radford, S.E. (2003). Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. *Journal of Molecular Biology*, **326**, 293-305.

Acknowledgements

We thank the BBSRC and the Wellcome Trust for financial support and Keith Ainley for technical assistance. SER is a BBSRC Professorial Research Fellow.

Targeting the Hepatitis C virus ion channel p7 for anti-viral therapy

Stephen Griffin, Dean Clarke, Steve Evans, Alastair Smith, Joachim Jäger,
Mark Harris and David Rowlands

Introduction

Hepatitis C virus (HCV) currently infects over 3 % of the world population and is the major indicator for liver transplant surgery in the west. Acute infection is usually asymptomatic but leads to persistence in the majority of cases causing chronic liver disease. Treatment of the virus is currently limited to the use of type 1 interferon either alone, or in combination with the guanosine analogue ribavirin, and this therapeutic regime is expensive, poorly tolerated, and effective in only 40 % of cases world-wide. Furthermore, resistance to this treatment is common in the viral genotypes found in the west. The search for a vaccine and alternative therapies has been hampered by the current inability to successfully culture the virus *in vitro*, making the identification of new anti-viral drug targets paramount.

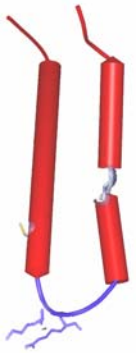


Fig. 1. Computer model of a p7 monomer

The HCV p7 protein is a small hydrophobic protein of 63 amino acids comprised of two *trans*-membrane alpha helices separated by a short positively charged cytoplasmic loop (Fig. 1). It is predicted to belong to a family of proteins known as viroporins, which homo-oligomerise to form aqueous pores in cellular membranes. Perhaps the best characterised of these proteins is the M2 channel of Influenza A virus: the target of the first anti-viral drug, Amantadine.

Biophysical studies of p7 oligomers

High level expression of a p7 was achieved by fusion with glutathione-S-transferase in *E. coli* separated by a 6-histidine linker: GST-HIS-p7. Oligomeric complexes were visualised by Transmission Electron Microscopy (TEM) in collaboration with Dr Lucy Beales (University of Leeds). Both GST-HIS-p7 and cleaved near-native HIS-p7 (Fig. 2) were shown to form ring-like structures with dimensions similar to those predicted by computer modelling performed by Dr Joachim Jaeger (University of Leeds). We are currently collaborating with the laboratories of Dr Alastair Smith and Professor Jennifer Kirkham (University of Leeds) in order to visualise these complexes by Atomic Force Microscopy (AFM) in lipid membranes under aqueous conditions.

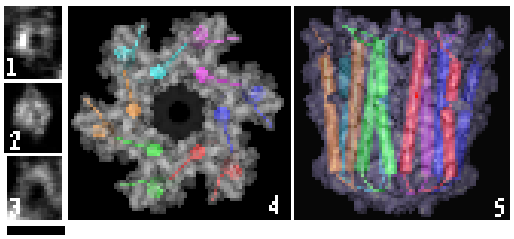


Fig. 2. 1-3; TEM of HISp7 oligomers. 4 & 5; computer model of p7 ion channel. Scale bar 5 nm

In vitro ion channel formation by p7 is inhibited by Amantadine

In collaboration with the laboratory of Professor Stephen Evans (University of Leeds), purified HIS-p7 was shown to form ion channels in artificial lipid bilayers using a Black Lipid Membrane (BLM) system. p7 ion channels showed a preference for calcium ions over potassium, consistent with the proteins' reticular localisation in living cells. Furthermore, this activity was abrogated by the addition of Amantadine (Fig. 3). Recent clinical trials that

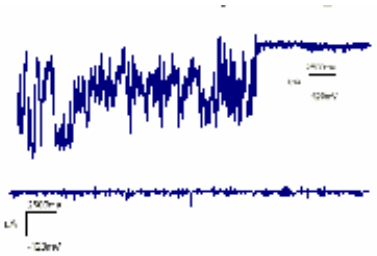


Fig. 3. p7 ion channel activity before (top) and after (bottom) adding 1 μM Amantadine

include this drug with existing therapy show efficacy in many cases, particularly in those patients who were previously non-responsive. We propose that p7 is the target of Amantadine and are working with GlaxoSmithKline to develop new anti-HCV compounds based on this finding.

p7 modulates cellular membrane permeability

Recently, we have gone on to assess the effect of expressing p7 in the context of living cells. Inducible expression of native p7 in *E. coli* was shown to cause a growth-inhibitory effect due to membrane leakage, which was reversed by the addition of Amantadine. Furthermore, mutation of the charged loop also negated this effect consistent with studies in the related Pestivirus: Bovine Viral Diarrhoea Virus.

We have also demonstrated p7-induced membrane permeability in mammalian cells causing enhanced susceptibility to Hygromycin B. In addition, we have shown in collaboration with Dr Wendy Barclay (University of Reading) that p7 can functionally replace Influenza A M2 protein in a bioassay and is again inhibited by the addition of Amantadine.

Further studies of p7 and other viroporins

We are currently looking at chimeric viral systems in order to assess the function of p7 in the context of virus replication and it's ability to functionally complement similar proteins from other viruses. The p7 of the closely related GB Virus B, BVDV p7, Picornavirus 2B (Fig. 4), and Flavivirus NS1 proteins are all candidate systems for such studies. Work has already begun on characterising these proteins and construction of viral chimeras. We are also attempting to crystallise the p7 of HCV and of BVDV, as well as pursuing other structural investigations such as NMR and mass spectroscopy in order to determine the precise composition of the ion channel structure for rational drug design.

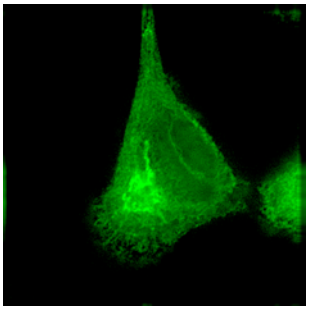


Fig. 4. GFP-tagged Picornavirus 2B protein

Publication

Griffin, S.D., Beales, L.P., Clarke, D.S., Worsfold, O., Evans, S.D., Jaeger, J., Harris, M.P. & Rowlands, D.J. (2003) The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. (2003) *FEBS Lett* **535**, 34-8

Funding

This work was funded by the MRC "Enabling Technologies" Co-operative, University of Leeds.

Single molecule spectroscopy to probe folding of individual proteins

Chris Gell, David Brockwell, Sara Pugh, Anil Nagalingam, Stuart Warriner, Sheena Radford and Alastair Smith.

Introduction

Spectroscopic measurements of bulk sample give only an ensemble-averaged measure of a molecular property. The signal obtained includes a contribution from a variety of different conformers and local environments within the sample. In order to study the heterogeneity of systems it is necessary to remove ensemble effects by examining such systems at the single molecule level. We have been implementing new and emerging single molecule techniques, in particular single molecule fluorescence resonance energy transfer (FRET). In addition, we have now turned our attentions to molecular biology and chemical synthesis to provide exciting new samples to study the protein folding problem.

Instrument Development

Previous reports described our efforts to use fluorescence correlation spectrscopy to probe protein folding, recently we have extended our capability to the measurment of single molecule FRET on both diffusive and immobilised samples.

Fig. 1A shows a fluorescence *versus* time trace for a solution of dye labelled DNA molecules measured on two detection channels simultaneously. Each detection channel monitors exclusively either the donor (green) or acceptor (red) dye fluorescence. Bursts are due to the transit of single molecules through the small detection volume (~0.1 fl). Coincident events (fluorescence from both the donor and acceptor monitored at the same point in time) can only occur when there is FRET between the dye pair. The ratio of the coincident burst heights can then be used to cvalulate the FRET efficiency for that particular single molecule.

Histograms of many single molecule events reveal peaks at the average FRET efficiency. Fig. 1B and C show calculated histograms for DNA samples with differing separations of the dye pair, thus with expected differing FRET efficiency.

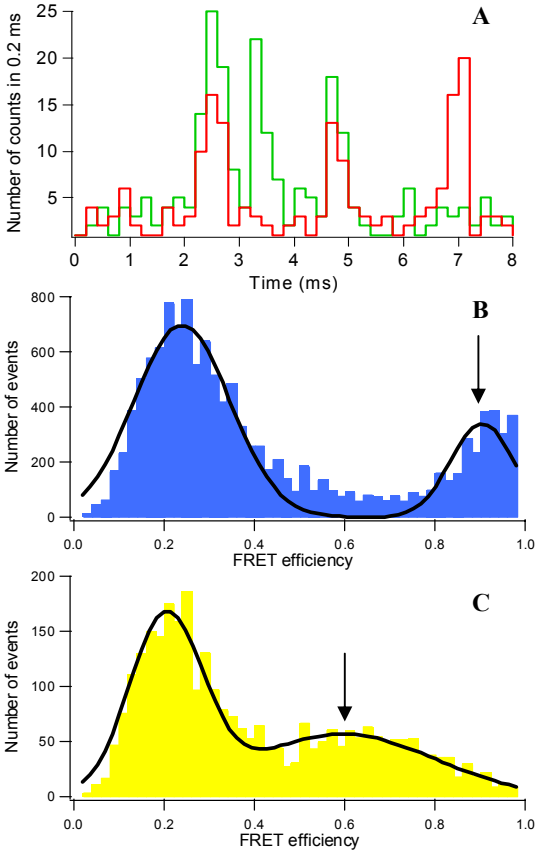


Fig. 1. Single Molecule FRET data from DNA labelled with a dye pair (green – red). Coincident single molecule events can be seen in the green (donor) and red (acceptor) channels **(A)**. Histograms calculated from DNA samples with the dyes 5 b.p. **(B)** and 17 b.p. **(C)** apart. The lower relative FRET efficiency can clearly be seen for the larger separation (arrows).

Fig. 1B shows a histogram for a separation of the dyes of 5 b.p. and Fig. 1C for 17 b.p. A peak at higher transfer efficiency is clearly seen for the closer separation (E ~ 0.9), compared to the longer separation (E ~ 0.6).

The strong peaks (E ~ 0.2) are a common feature of this type of experiment and demonstrate the ability of the technique to resolve heterogeneity in sample and are due to an excess of donor-only labeled DNA in the solution.

The potential of FRET studies on single immobilised molecules, particularly single protein molecules immobilised in the aqueous pores of agarose or acrylamide gels is great.

We have developed our instrumentation to study such immobilised systems by incorporating a scanning stage into our microscope. This enables the location and subsequent continuous monitoring of single molecules in gel films. Fig. 2 shows data demonstrating this capability.

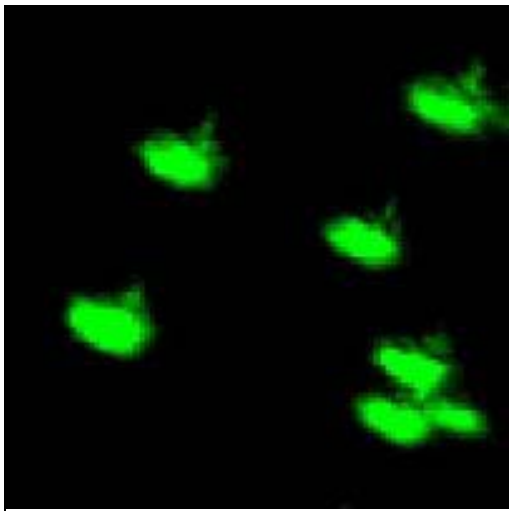


Figure 2. Image showing individual (20 nm diameter) dye-labelled polystyrene spheres immobilised in an agarose gel (contrast adjusted).

Proteins for single molecule FRET studies

A limiting factor in the study of single molecules, and in particular single protein molecules, has been the ability to homogenously label with a dye pair suitable for FRET. We are currently pursuing two routes to enable us to study single immunity proteins with the techniques and equipment described above.

Firstly, we are attempting a chemical labelling and molecular biology based approach. This includes engineering a protease recognition site followed by a cysteine into the N-terminus of one immunity protein. Cleavage of the purified protein with the relevant protease should yield an N-terminal cysteine. Using a novel thiol-ester acceptor dye it is possible to specifically label a cysteine at the N-terminus of the protein with the acceptor dye (carboxytetramethylrhodamine benzylthio ester), which was synthesised from the commercially available carboxytetramethylrhodamine succinimidyl ester. Site directed mutagenesis can then be used to introduce a second cysteine as an attachment site for the thiol reactive donor fluorophore *N*-(4,4-difluoro-5,7-dimethyl-4-bora- 3a,4a-diaza-*s*-indacene-3-propionyl) -*N*'-iodoacetylethylenediamine. Alternatively the dyes can be added to two cysteines engineered at different locations using mutagenesis. The position of the donor can be varied by altering the location of the second cysteine, making it possible to create a library of mutants with the donor dye at different positions.

Secondly, we plan to develop native chemical ligation methods to facilitate the total chemical synthesis of immunity proteins. Stepwise solid phase synthesis will be used to synthesise sections of the proteins, which will be linked by chemical ligation to yield the complete structure. Chemical synthesis allows us to readily generate protein molecules containing various modifications such as fluorescence labels at any specific position.

Publications

Gell C., Brockwell D.J., Beddard G.S., Radford S.E., Kalverda A.P. & Smith D.A. (2001) Accurate use of single molecule fluorescence correlation spectroscopy to determine molecular diffusion times. *Single Molecules* **2,** 177-181.

Acknowledgements

We thank the Wellcome Trust and University of Leeds for funding.

In vitro studies of MetJ structure-function relationships.

Tim Knowles, Steve Homans, Simon E.V.Phillips and Peter G. Stockley.

Control of DNA binding in the MetJ system appears not to be driven by structural change(s), but more likely by a novel electrostatic effect due to the positive charge on the sulphur of the co-repressor AdoMet although this is ~18 Å away from the phosphate backbone. Understanding the dynamics of this system in solution is therefore required to fully understand what is happening during DNA-binding. Does any conformational change occur? What part does the electric genetic switch play, and precisely what interactions occur between the protein and DNA? Nuclear magnetic resonance has therefore been implemented to answer these questions.

Double isotopically labelled ($^{15}N/^{13}C$) MetJ has been prepared and shown to be active and stable at 1 mM concentrations used for NMR. Using multiple triple resonance techniques (HNCA, HN(CO)CA, CBCACONH, HNCACB) 82% assignment of the backbone has been achieved (Fig.1), the remaining 18% unassigned is either not seen or cannot be assigned to a particular residue. Recent work has started comparing the solution and apo-repressor crystal structures using residual dipolar coupling data provided from decoupled HSQC experiments in isotropic and Pf1 Phage aligned media (Fig.2) to provide distance and angle constraints for structure comparison.

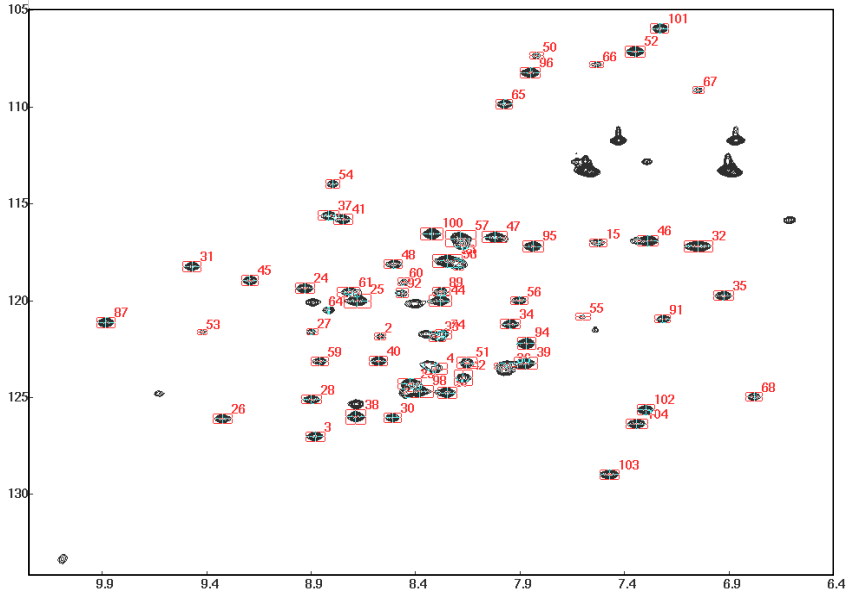
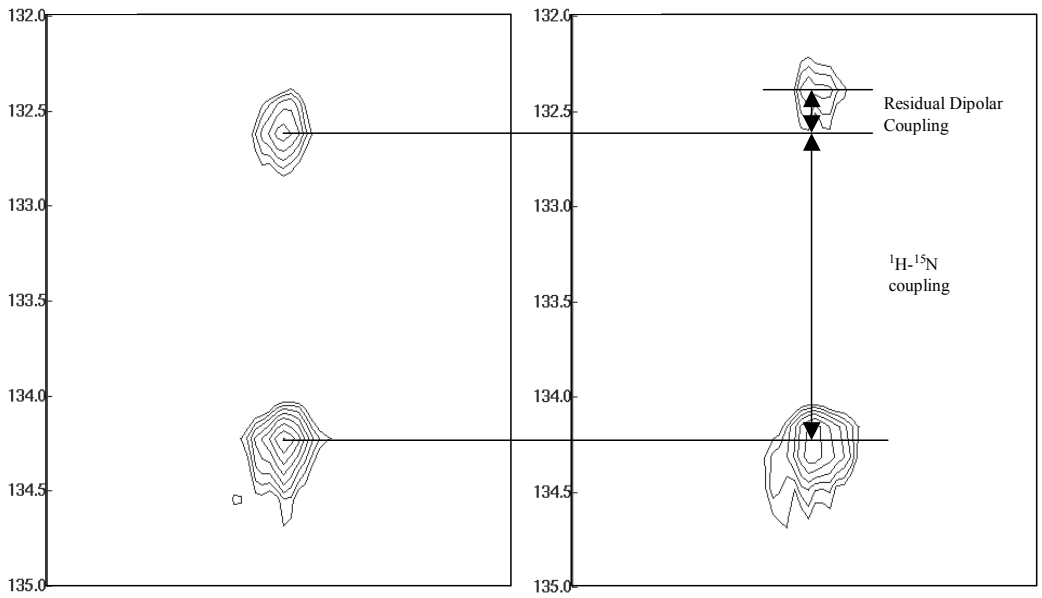


Fig. 1 – 500MHz ^{1}H-^{15}N HSQC spectrum of MetJ. Assigned peaks are labelled in red. Each peak corresponds to a ^{1}H attached to a ^{15}N backbone amide.

A.



B.

$$D_{ij} = -\left[\frac{\gamma_i \gamma_j h}{2\pi^2 r^3}\right] I_{zi} I_{zj} \left\langle \frac{(3\cos^2\theta - 1)}{2} \right\rangle$$

Fig. 2. Residual dipolar couplings. A: 600MHz HSQC spectra in isotropic and Pf1 phage aligned media. In phage aligned media a small proportion of the protein also becomes aligned providing the observed increase in coupling corresponding to the presence of residual dipoles. B: Residual dipolar coupling equation showing the dependence on distance (r^3) and angle ($3\cos^2\theta$-1).

Over the years we have studied a number of MetJ mutant proteins with altered DNA-binding properties. One of the most interesting of these is Q44K, a protein with a single amino acid substitution outside the β-ribbon DNA-binding motif. Q44K has the unusual property, compared to wild-type protein, of being able to bind stably to DNA targets encompassing only single met-box sites; the wild-type protein requiring two tandem such sites before it will bind stably. The wild-type behaviour arises due to the protein-protein co-operativity generated by an extensive protein dimer-dimer contact on the DNA. Q44K can still make this higher order species, as well as the single dimer complex, implying that each diner's intrinsic DNA affinity has been increased whilst its protein-protein co-operativity has been lowered. Crystal structures of both types of complex offer an explanation for this behaviour. The side chain of Lys-44 makes an additional intermolecular contact to the DNA phosphate backbone, but, in the single dimer complex, the protein is orientated differently from in the higher order complex, requiring conformational rearrangement in order to make the protein-protein contacts responsible for the higher order species. We believe that this species is a good model for an intermediate on the pathway to operator site saturation, the first such species for which an X-ray structure is available.

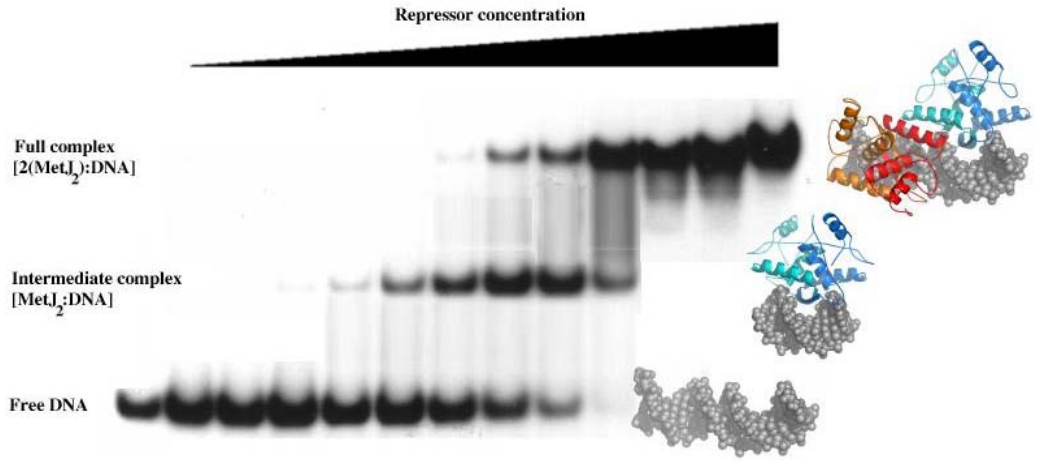


Fig. 3. Electrophoretic mobility shift assay of Q44K binding to a tandem Met box site. As repressor concentration increases the intermediate complex is observed, composed of a single repressor dimmer bound to DNA. With increasing repressor concentration this then forms the stable double repressor bound complex observed in the wild-type.

Publications

Structural and functional studies of an intermediate on the pathway to operator binding by *E.coli* MetJ. (2002) *J.Mol.Biol* **320**, 39-53.
Yi-Yuan He, Colin W.Garvie, Stone Elworthy, Iain W. Manfield, Teresa McNally, Isobel D Lawrenson, Simon E V Phillips and Peter G Stockley.

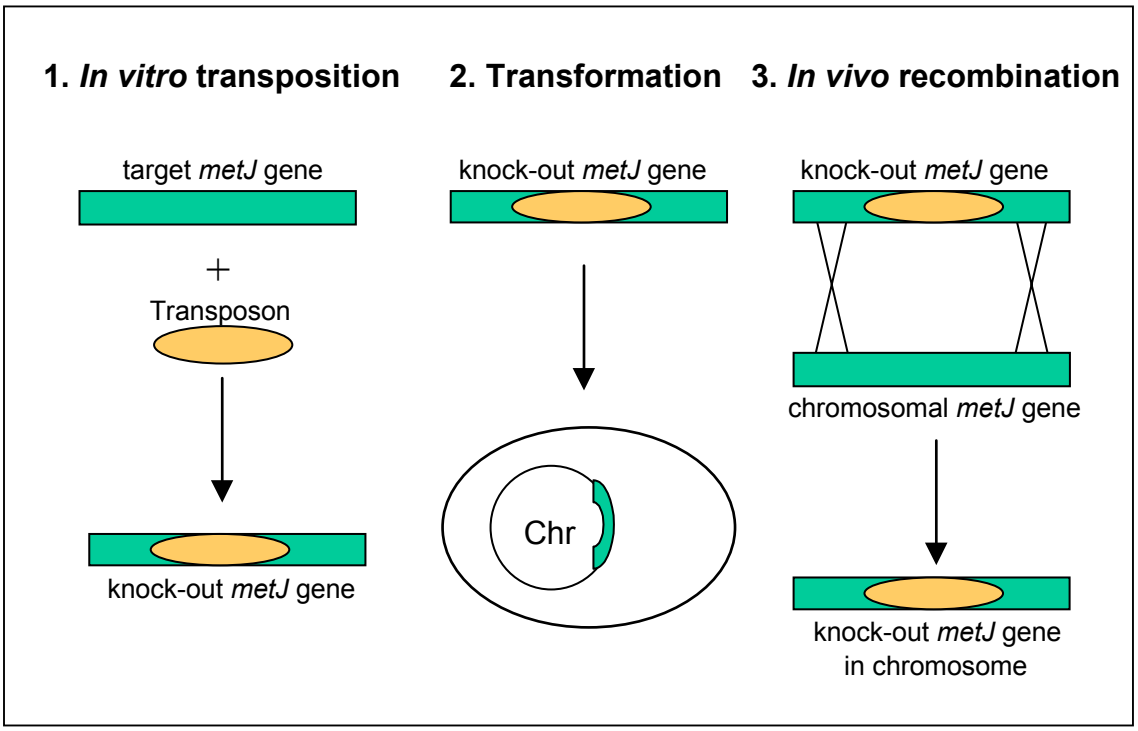
Funding

Financial support from the University of Leeds, the Wellcome Trust and the BBSRC is gratefully acknowledged.

Microarray studies of gene expression in the methionine biosynthetic pathway of *Escherichia coli*

Ferenc Marincs, Iain W. Manfield, Kenneth J. McDowall and Peter G. Stockley.

The *met*J gene of *Escherichia coli* encodes the apo-repressor protein MetJ, which, together with the co-repressor S-adenosine-methionine (AdoMet), represses expression of the methionine biosynthetic genes if methionine is present in the environment. MetJ accomplishes its repression function by binding to met-box operator sequences located in the 5'-untranslated regions of the regulated genes. In theory, destroying the function of a repressor protein should result in derepression of the regulated genes. An obvious way to generate a non-functional repressor is to make a mutation in its gene. To achieve this with MetJ, the *met*J gene was mutagenised using a combined *in vitro* transposon mutagenesis - *in vivo* recombination approach as outlined in the scheme below:



Transcriptomics employs micro-array technology as its technical platform to compare gene expression in two different samples, allowing large number of transcripts to be compared and analysed simultaneously. This technology was used to study expression of the methionine biosynthetic genes of the *met*J knock-out strain compared to its wild-type counterpart. A microarray containing 76 *E. coli* genes was manufactured in our laboratory by spotting PCR-made DNA fragments onto poly-lysine coated glass microscope slides with an Affymetrix microarray spotter. mRNA of the wild-type and knockout strains was isolated, transcribed into fluorescently labelled cDNA that was mixed and hybridised to the microarray. Hybridisation signals were obtained by scanning the slides with an Affymetrix laser scanner, and raw data were analysed using ArrayPro software.

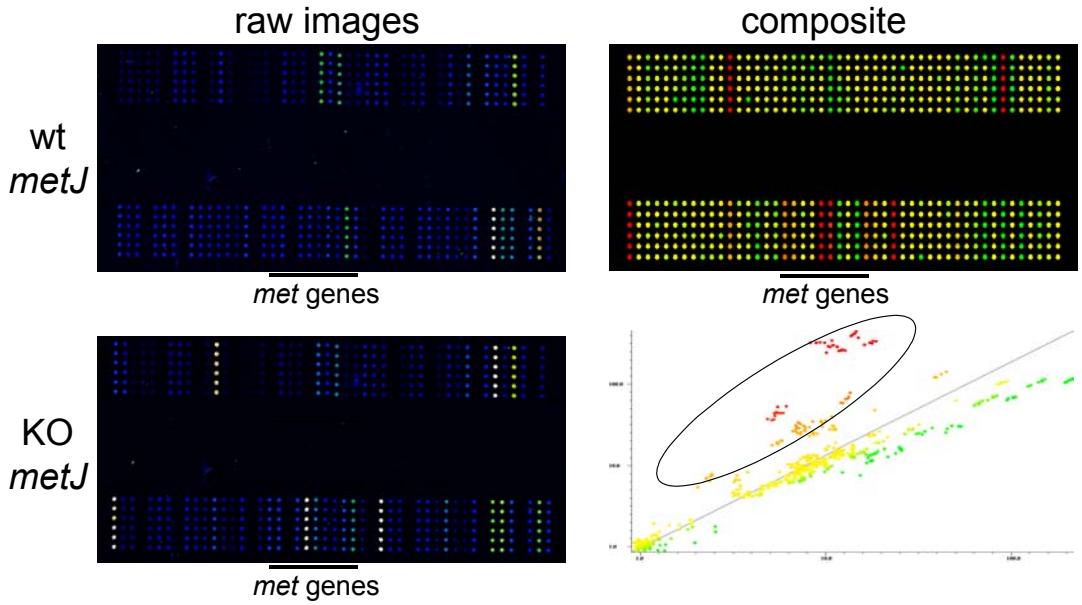


Fig. 1 As indicated by orange and red spots in the scatter plot and the microarray composite image, data analyis revealed that the *met* genes are derepressed in the absence of a functional MetJ repressor protein. These data allow us now to compare the functions of mutant repressor proteins *in vivo* to complement *in vitro* studies.

Funding

Financial support from the University of Leeds, the Wellcome Trust and the BBSRC is gratefully acknowledged.

Protein-RNA interactions and virus assembly studied by mass spectrometry

Alison E. Ashcroft, Simona Francese, Hugo Lago, Chris J. Adams, Andrew M. Smith, Nicola J. Stonehouse, Wilf Horn, Simon E.V.Phillips and Peter G. Stockley.

RNA bacteriophages have been studied for many decades as models in which it is experimentally feasible to decipher the molecular details of regulatory RNA-protein interactions. MS2 is a bacteriophage that infects *E. coli*. The *T*=3 capsid enclosing the MS2 single stranded RNA genome consists of 180 coat protein (CP) subunits (MW 13,728.6 Da). Self-assembly of the capsid is catalysed by a sequence specific RNA-protein interaction between a dimer of the coat protein (CP_2) and an operator stem-loop of the RNA (the TR sequence) of just 19 nucleotides (MW 6045.7Da) in length (Fig. 1). The TR/CP_2 molecular recognition event is now reasonably well understood, especially from X-ray crystal structures of complexes formed by soaking RNA variants into crystals of RNA-free recombinant capsids. Indeed, we can manipulate the capsid in a variety of ways, for instance to generate a nanoscale cell-specific drug delivery vehicle. Although this sequence-specific interaction shows remarkable plasticity in its details, as demonstrated by the structures of a series of purine RNA variants, understanding how it gives rise to the precise assembly pathway from this initial interaction remains a critical unsolved problem.

Fig. 1. Ribbon diagram of an AB coat protein dimer bound to operator RNA. The capsid sub-units A and B are coloured blue and green respectively; the RNA is yellow.

We are using electrospray ionisation-mass spectrometry (ESI-MS) as one of a number of bio-analytical methods to determine the assembly pathway of the MS2 virus and other, related RNA bacteriophages. In our preliminary ESI-MS experiments we have shown that the CP can be analysed under denaturing conditions where it appears as a monomer, or under native conditions where it is a non-covalently bound dimer which can be dissociated to monomer by the use of collisionally induced decomposition (CID) (Fig. 2).

Analysis of CP mixed with TR under native conditions showed a [CP_2+TR] species as the major component (which is significantly more resistant to CID than the CP alone) accompanied by an intricate mixture of several higher order species (data not shown). We have also monitored the interaction between CP and TR variants known to be unable to trigger self-assembly. In both cases there was evidence for formation of a [CP_2+RNA] complex but without the higher order structures observed with the TR sequence. The analysis of recombinant phage capsids also did not produce similar species to those seen in assembly reactions, suggesting that the higher order species are genuinely associated with virus assembly rather than simply MS artefacts or assembly by-products. We are at present undertaking an extensive ESI-MS study of CP binding (wild-type and mutants) to TR

sequences (and variants) to characterise fully the higher order species and their rôle in the assembly pathway.

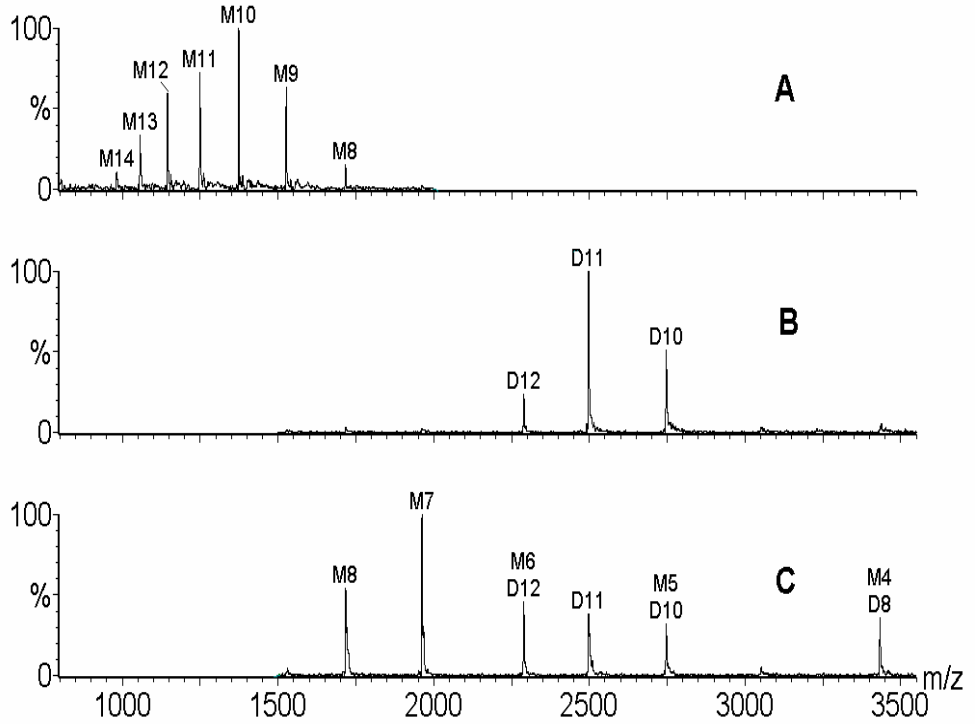


Fig. 2. Mass spectrum of MS2 coat protein (CP) obtained under (A) denaturing conditions with the protein dissolved in 1:1 *(v/v)* acetonitrile : 0.1% aq. formic acid using a cone voltage of 40V; (B) native conditions with the coat protein dissolved in 40 mM aq. ammonium acetate using a cone voltage of 40V; (C) as in (B) with the voltage raised to 80V. Spectra were obtained on a Q-Tof mass spectrometer (Micromass UK Ltd) using nanoflow ESI with a capillary voltage of 900-1200V and MCP 2400-2850 V. The peaks are labeled as CP monomer (M) or dimer (D) followed by the charge state of the ions.

Publications

Brown, W.L., Mastico, R.A., Wu, M., Heal, K.G., Adams, C.J., Murray, J.B., Simpson, J.C., Lord, J.M., Taylor-Robinson, A.W. and Stockley, P.G. (2002) RNA bacteriophage capsid-mediated drug delivery and epitope presentation. *Intervirology* **45,** 371-380.

Helgstrand, C., Grahn, E., Moss, T., Stonehouse, N.J., Tars, K., Stockley, P.G. and Liljas, L. (2002) Investigating the structural basis of purine specificity in the structures of MS2 coat protein RNA translational operator hairpins. *Nuc. Acids Res.*, 30, 2678-2685.

Ashcroft,A.E., Lago, H., Stonehouse, N.J. and Stockley, P.G. (2000) *Virus assembly and disassembly investigated by electrospray ionisation mass spectrometry, 15th International Mass Spectrometry Conference*, Barcelona, Spain.

Parrott, A.M., Lago, H., Adams, C.J., Ashcroft, A.E., Stonehouse, N.J. and Stockley, P.G. (2000) RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour. *Nuc. Acids Res.*, **28,** 489.

Collaborators

Lars Liljas and colleagues, Uppsala University

Funding

Financial support from the University of Leeds, the Wellcome Trust, BBSRC, Micromass UK Ltd and AstraZeneca is gratefully acknowledged.

Small molecular weight mimics of MetJ.

Sipra Deb, Iain W. Manfield, and Peter G. Stockley.

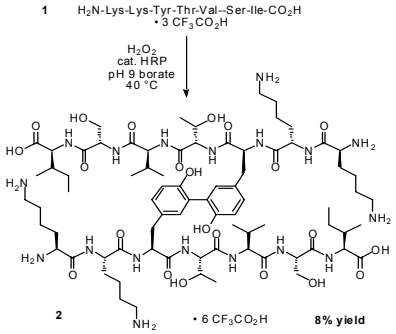


Fig.1. Synthetic scheme for the production of the dityrosine-linked MetJ beta-ribbon DNA-binding motif.

A unique dityrosine-linked synthetic peptide encompassing the β-ribbon DNA-binding sequence of MetJ has been synthesised (Fig. 1) and tested for its ability to bind DNA using surface plasmon resonance. The monomer peptide binds only very weakly, if at all, to a series of DNA fragments encompassing the natural *met* operator sites. In contrast, the dimer binds strongly and apparently co-operatively to these targets, but binding is non-sequence specific and eventually saturates the DNA with a stoichiometry of roughly one dimeric peptide per two base pairs of DNA. Despite its modest size, the peptide is able to ablate *in vitro* transcription by RNA polymerase T7 (Fig. 2). Further work is in hand to understand the molecular basis of DNA binding by this peptide and in order to introduce sequence-specificity to its interaction.

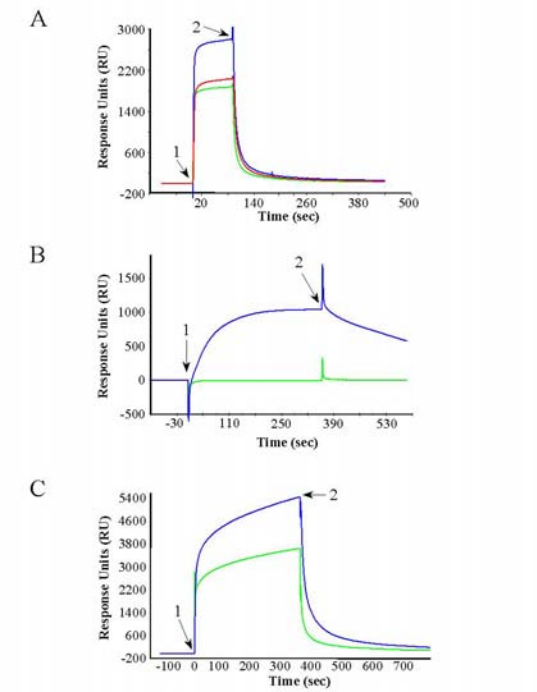


Fig. 2: Dityrosine peptide binding to three differing templates (A); T7 RNA polymerase transcribing two of these, one with and one without a promoter (B), and (C) polymerase and peptide injected together across an SPR sensorchip. The last experiment shows the characteristic features of peptide binding, i.e. transcription was inhibited.

Publication

Yoburn, J.C., Deb, S., Manfield, I.W., Stockley, P.G. & Van Vranken, D.L. (2003) A biaryl peptide crosslink in a MetJ peptide model confers co-operative, nonspecific binding to DNA that ablates both repressor binding and *in vitro* transcription *Bioorganic and Medicinal Chemistry*. 11, 811-816.

Collaborators
Dave Van Vranken (UC, Irvine)

Funding
Financial support from the Wellcome Trust is gratefully acknowledged.

Molecular interactions in the bacteriophage ϕ29 DNA packaging motor

Jonathan Wood and Nicola J. Stonehouse

Introduction

During the packaging of genomic DNA by the bacteriophage ϕ29, components of the virus form one of the most powerful molecular motors known, with DNA packaged to almost crystalline density. The motor is an RNA-protein complex and consists of a prohead incorporating a head-tail connector particle, self-assembling RNA molecules (pRNA), together with viral ATPase (Fig. 1). Both the connector and pRNA components are macromolecular assemblies, each forming ring-like structures with a central pore, through which the viral DNA is translocated. During viral assembly, rotary action of the connector is thought to be converted into linear translocation of the DNA.

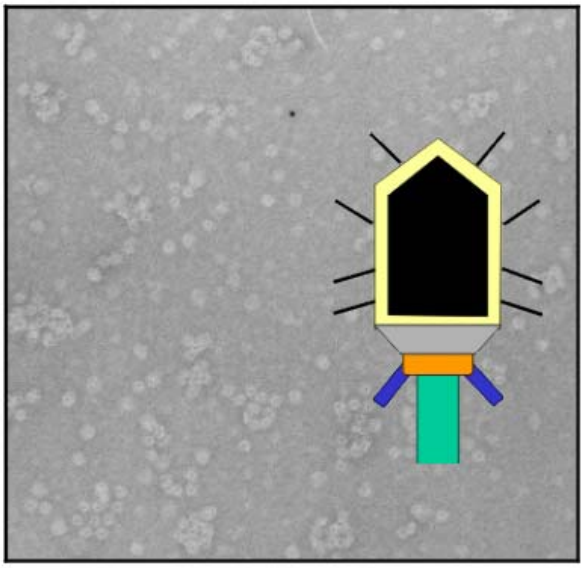


Fig. 1
Transmission electron micrograph, stained with uranyl acetate, of ϕ29 connector particles overlaid with a cartoon representation of the phage particle.

The involvement in the action of a rotary motor is a new role for RNA molecules and thus an exciting area of research. In addition, an understanding of the key interactions between nucleic acids and viral proteins is key to the development of anti-viral therapies. Many aspects of the assembly of ϕ29 are mirrored in Herpes and Adenoviruses, major human pathogens. Such fine dissections of molecular interactions have proved to be complex in these viruses, however, data from a model phage systems has underpinned our knowledge of virology in the past and is likely to continue to do so in the future.

Little is known about the affinity of the molecular interactions involved in assembly of the virus, especially the pRNA multimerisation event. Knowledge of the stoichiometry and affinity of interactions between molecules is vital in the understanding of how these interactions occur.

The self assembly of pRNA

Intermolecular base pairs between two loop regions in the secondary structure of 120 nucleotide (nt) pRNA form a multimeric RNA assembly in the presence of magnesium ions. This ability to form multimers is essential for DNA packaging, although different studies have suggested both

active pentameric and hexameric pRNA rings. We have performed light scattering and equilibrium sedimentation experiments in the analytical ultracentrifuge (AUC) with pRNA in order to investigate the stochiometry of the self-association. This was facilitated by our previous efforts into calculation of an accurate partial specific volume for pRNA as, in contrast to proteins, there is very little literature regarding such values for RNA molecules.

Dynamic light scattering shows that the mean size of the pRNA population increases on addition of magnesium ions, and that this effect is reversible on addition of EDTA. Experiments with a pRNA molecule that contains mutations so that it cannot self-assemble suggest that the size changes are mainly due to multimerisation, but that conformational changes in the presence of magnesium ions may also occur.

Using a size exclusion column in conjunction with the light scattering apparatus has made it possible to separate the species in the population of RNA molecules. In this way, we identified a novel pRNA species – a trimer. Studies are underway to characterise this molecule.

Collaborators
Neil Thomson, Department of Physics
Peixuan Guo, Purdue University, USA

Funding
We thank the MRC and the University of Leeds for funding.

New crystals of replication initiator proteins

Steve Carr, Simon Phillips and Chris D. Thomas

Background

Staphylococcal plasmids of the pT181 family use a "rolling circle" mode of replication. This requires a plasmid-encoded initiator protein, Rep. Although very similar in amino acid sequence, Rep proteins of these plasmids are specific for their cognate plasmids *in vivo*: thus RepD protein initiates replication for pC221, RepC for pT181, and RepN for pCW7. Although these proteins have sequence motifs in common with the gene II initiator protein of bacteriophage M13, to date there are no known structures of this family of proteins.

Recent findings

Crystallisation trials have been performed with the protein RepDN - a chimeric protein consisting of the N-terminal domain of RepD fused to the C-terminal domain of RepN. RepD and RepN are share greater than 90% sequence identity and so although RepDN is a non-natural protein its fold is likely to be representative of either of the native proteins. RepDN yielded large, cubic crystals from a condition containing 100 mM sodium citrate pH 5.5 and 2.5M 1,6-hexanediol. The crystals diffract to a resolution of at least 3 Å and are of the cubic space group I432. Seleno-methionine labelled protein has been produced and been shown to crystallise from similar conditions to native RepDN. The selenium labelled protein is currently being used to generate MAD phases with a view to solving the structure of RepDN.

Fig. 1. Crystals of RepDN.

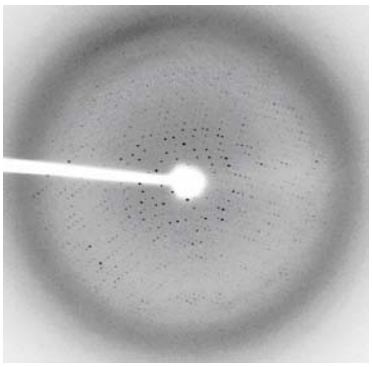


Fig. 2. Typical diffraction image from a RepDN crystal - the edge of the diffraction pattern corresponds to a resolution of 3.0Å.

Acknowledgements

We thank Val Sergeant for technical support. This work is funded by the Wellcome Trust.

Molecular mechanism of Staphylococcal plasmid transfer

Jamie A. Caryl and Chris D. Thomas

Background

Horizontal gene transfer in bacteria results in genetic diversity with important medical consequences. Small non-self transmissible, mobilisable, Staphylococcal plasmids such as pC221 offer a simple system that embodies the initial events in plasmid mobilisation.

pC221 is a 4.6kb chloramphenicol resistance plasmid of *Staphylococcus aureus*. It is a non-self transmissible plasmid that can be mobilised by a co-resident self-transmissible plasmid such as pGO1. Being a small plasmid, pC221 contains only those genes required for its own DNA processing and contains four such loci: an origin of transfer (*oriT*); a DNA relaxase, MobA; and the putative accessory proteins MobB and MobC.

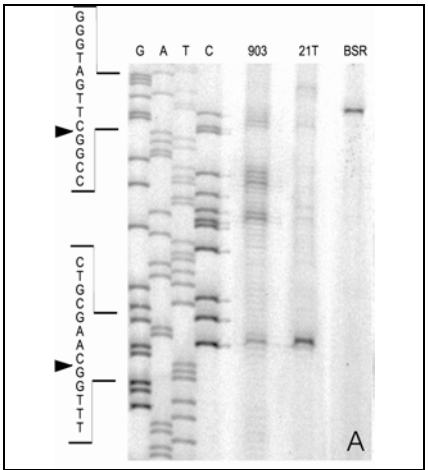


Fig. 1. Mapping of the *nic* site and DNA binding of MobC. Upper arrow indicates *Bsr*FI control. Lower arrow indicates MobA *nic* site.

Recent findings

We have demonstrated that mutations in either the *mob*A or *mobC* ORFs of pC221 result in a loss of observable nicking *in vivo*. In addition, previous work has failed to demonstrate nicking with purified MobA protein alone. The MobC protein has been purified and shown to be a 33.4 kDa dimeric, basic protein that binds *ori*T DNA specifically and cooperatively. The nicking reaction has been reconstituted *in vitro* (Fig. 1) and has demonstrated the requirement for MobA and MobC proteins, in the presence of Mg^{2+} or Mn^{2+}, for site- and substrate-specific nicking.

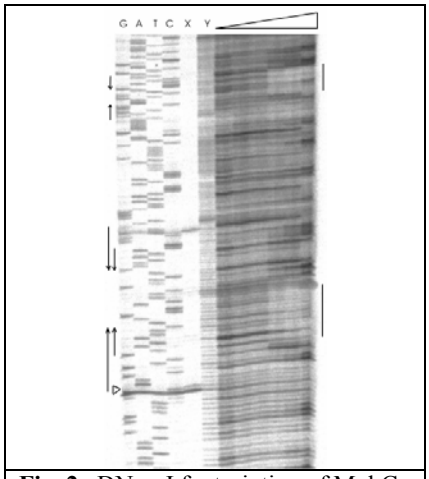


Fig. 2. DNaseI footprinting of MobC at *oriT*. Titration of MobC. Vertical bars indicate protection. Arrows indicate dyad symmetries.

Footprinting studies, probing with $KMnO_4^-$ and DNaseI, have been performed with MobC and MobA on supercoiled pC221*cop903 oriT* DNA and analysed by primer extension. These indicate that structural perturbations are introduced within a sequence of dyad symmetry upstream of the *nic* site by the binding of MobC, the binding site(s) of which have been determined (Fig. 2). Further footprinting is planned to investigate the binding of MobA in the presence and absence of MobC. Double-stranded oligonucleotides containing binding regions will be used to resolve the minimum binding sequence using electrophoretic mobility shift assays and fluorescence anisotropy.

Publications

Caryl, J.A. and Thomas, C.D. (2002) Initial events in small staphylococcal plasmid transfer. *Plasmid Biology 2002: International Symposium on Plasmid Molecular Biology. Plasmid* **48**, p.238.

This work has been presented at the **UK Mobile Genetic Elements Workshop 2002**, Birmingham and **Plasmid Biology 2002: International symposium on the molecular biology of bacterial plasmids**, Pittsburgh, USA.

Acknowledgements

We thank Val Sergeant for technical support. This work is funded by the Wellcome Trust.

Dimerisation determinants of replication initiator proteins

Chris D. Thomas

Background

Replication of staphylococcal plasmids such as pT181 and pC221 requires the plasmid-encoded Rep proteins RepC and RepD. These proteins are over 80% identical in primary sequence, yet retain DNA binding specificity for their cognate plasmid replication origins due to a small divergent region of six amino acids within the carboxyl terminus.

A previous fusion between RepD and RepC consisted of residues 35-216 of the former and 217-314 of the latter, a 34kDa derivative designated RepDC. As expected, this fusion protein displays DNA sequence specificity for pT181, a property inherited from the wild-type RepC protein. However, it was also found that heterodimers could not be formed between RepDC and RepD, suggesting that some form of dimerisation determinant also resided in the C-terminal portion of the protein.

Recent findings

Inspection of the amino acid sequence of the carboxyl terminus of the Rep protein family suggested four distinct regions (I - IV) of predicted alpha-helical content, bounded by helix-breaking residues (Fig. 1). Interestingly, the second such region (region II) displayed only half the average amino acid sequence identity between RepC and RepD and contained a potentially buried region (the filled area in the figure). The DNA sequence specificity determinants (boxed) were contained within the adjacent region III.

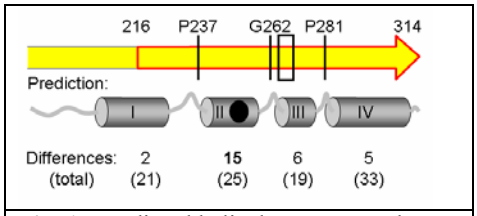


Fig. 1. Predicted helical structure at the carboxyl terminus of the Rep proteins.

A series of crossover mutations was constructed, exchanging regions I-IV in various combinations between RepD and the previous construct, RepDC. Each of these was tested for the ability to form heterodimers with full-length RepC or RepD, and also for DNA binding specificity as the pure homodimer. In addition, a novel fusion between RepD and the carboxyl terminus of RepN was also tested. A selection of the results is presented in Fig. 2.

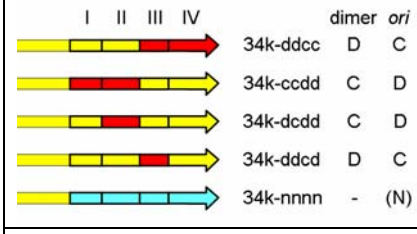


Fig. 2. Dimerisation (*dimer*) and DNA binding (*ori*) specificities are indicated.

From these and other data it is clear that region II alone is sufficient to define the dimerisation specificity, and region III defines the DNA binding specificity of the Rep proteins. Not surprisingly, no evidence of heterodimer formation was observed with the derivative of RepN, a third distinct member of the family.

Close proximity of these two regions may indicate the need to preserve the association between plasmid initiator specificity and dimerisation partner: in a cell where pT181 and pC221 are both replicating, the Rep proteins must maintain their specificity from one another - heterodimer formation could result in a reduction in cellular concentration of homodimeric initiator, followed by reduction in copy number and even plasmid loss. In addition, copy regulation is maintained as the

replication-inactive RepD* protein shows no evidence of monomer exchange, even against native RepD.

Using the approaches outlined above, further work is directed to understanding the parameters of the dimerisation interaction, with specific mutants designed to yield novel asymmetric DNA sequence preferences.

Publications

Thomas, C.D. (2002) Dimerisation of the RepD initiator protein is a specific and stage-regulated process. *Plasmid Biology 2002: International Symposium on Plasmid Molecular Biology. Plasmid* **48**, pp.300-1.

This work has been presented at the **UK Plasmids Workshop 2001**, Birmingham and **Plasmid Biology 2002: International symposium on the molecular biology of bacterial plasmids**, Pittsburgh, USA.

Acknowledgements

Thanks to Val Sergeant for technical support. This work is funded by the Wellcome Trust.

Atomic force microscopy of DNA-protein interactions

Neal Crampton and Neil Thomson

Atomic force microscopy (AFM) can be used to distinguish proteins bound to nucleic acid templates. It has the capability of imaging under aqueous fluid conditions at nanometre resolution allowing the imaging of interactions and transactions between molecules in real-time. Current acquisition times for the raster-scanned images of the sharp tip across a surface are on the order of one minute for a micron squared area. This means that AFM is extremely adept at following slower processes, i.e. diffusion controlled events, such as facilitated searches for promoter sequences on DNA templates by transcription factors. Improvements in imaging times are on-going which will allow "movies" of biological processes to be acquired with image times close to video rate, a goal that has already been achieved by one or two research groups. As this technology becomes more widely available, AFM techniques will play a larger role in understanding of mechanisms of interaction between proteins and DNA.

Currently, we are studying two systems with AFM : DNA gyrase and a nested-gene transcription model.

DNA gyrase

DNA gyrase is a bacterial motor protein in a class known as topoisomerases, which are responsible for controlling the topological properties of DNA (i.e. amount of supercoiling or catenation). Most topoisomerases can relax supercoiled DNA, which is an energetically favourable process. DNA gyrase is unique amongst this class, because it can introduce supercoils as well as remove them. To wind or unwind DNA it must break both strands of DNA, capture another segment of the same DNA molecule and pass this through the double-strand break before resealing. We have been using AFM to study the length of DNA wrapped around the gyrase in the presence and absence of ADPNP to understand in more detail how gyrase achieves the strand passage of one segment of DNA through another.

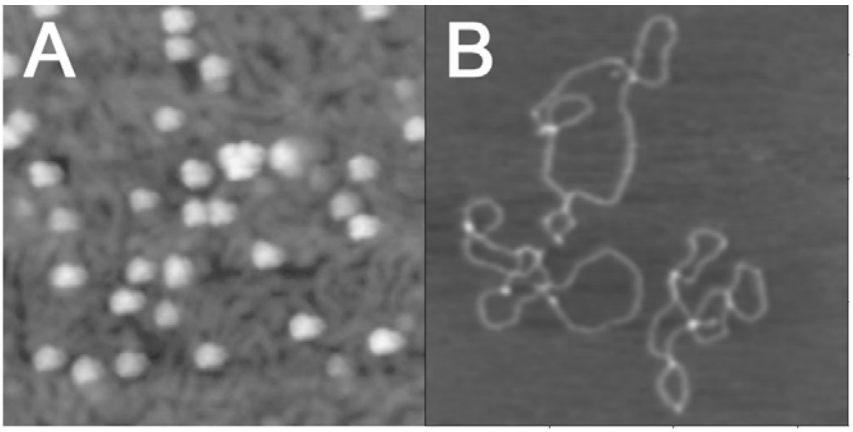


Fig. 1 Tapping-mode AFM images of air-dried samples on mica. (A). Single DNA gyrase molecules in the presence of the drug ciprofloxacin bound on top of a layer of DNA molecules (428 x 428 nm). (B) Supercoiled pBR322 plasmid DNA molecules (681 x 681 nm)

Nested-gene transcription model

"Nested gene" is a term coined for a gene that lies completely within the sequence of another gene. If the genes are arranged in a convergent orientation, the question of potential transcriptional collision arises. Nested genes therefore raise important implications for regulation of transcription and control of gene expression. Potential steric hindrance by RNA

polymerase transcribing the opposite strand and the possible generation of anti-sense RNAs have led to suggestions of co-regulation of expression. We are applying AFM to directly image two polymerases transcribing towards each other on opposite sides of a model DNA template. The template is based upon the highly conserved AMEL gene encoding for amelogenin (the principle component of the organic matrix of developing dental enamel), which is located entirely within (and in the opposite orientation to) the intron of a larger gene, ARHGAP6 (encoding a RhoGAP protein) on the human X chromosome. The initial AFM studies are concentrating on the study of the adsorption of DNA molecules to surfaces of varying chemistry to establish reliable protocols for imaging transcription processes in real-time.

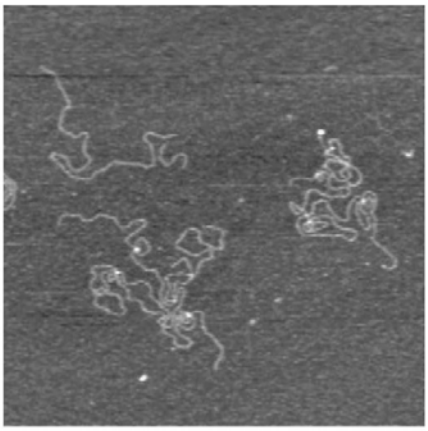


Fig. 2 Tapping-mode AFM image of linear DNA molecules adsorbed to mica functionalised with aminosilane (1130 x 1130 nm)

Collaborators

DNA gyrase project:
Jonathan Heddle and Anthony Maxwell (John Innes Centre)
Nested-gene Project:
Jennifer Kirkham and William Bonass (Oral Biology, Dental School, Leeds)

Funding

We gratefully acknowledge the EPSRC for financial support. Neil Thomson is an EPSRC Advanced Research Fellow.

The force producing mechanism of myosin

Peter Knight, Stan Burgess, Matt Walker, Neil Billington, F-K Jung, Kavitha Thirumurugan, Olusola Oke and John Trinick

We have been studying the conformations of the heads of myosin V molecules by electron microscopy and single particle image processing. We have examined molecules walking during ATPase activity attached to actin by both their heads. We have also looked at heads not attached to actin, both in the presence of ATP and in the absence of nucleotide. Unattached heads without ATP are relatively straight, whereas heads in the presence of ATP, in what is probably the ADP.Pi state, are strongly bent by ~90°, at what is called the motor domain/lever arm junction. This is the first direct demonstration of a massive conformational change in any myosin in response to ATP.

In molecules attached to actin during ATPase action, the trailing head is relatively straight, whereas the leading head is bent at the motor domain/lever arm junction. These two conformations are similar to the apo and ADP.Pi conformations referred to above, except that the leading head also appears to experience extra bending, probably due to tethering through the trail heads. This extra distortion is probably produced by Brownian motion and is likely to be important in the walking mechanism. Our pictures are the first of any myosin type to show the start of the power stroke on actin. The data are consistent with the 'tilting lever hypothesis' which states that force and movement in myosins is produced by a gross conformational change in the attached head.

We have also obtained micrographs of molecules of a mutant myosin V, where the number of calmodulin chains in the lever arm part of the molecule is varied from the wild-type number of 6. We examined molecules with 2, 4, and 8 calmodulins. The data show that binding of the second (lead) head has a strong preference for an actin subunits with a similar azimuth to the trail head. This suggests that the wild type molecule has evolved to span the helical repeat of the actin filament, which allows the molecule to walk straight. This is probably important in the function *in vivo*, preventing vesicle cargoes carried by myosin V colliding with membranes and other obstacles.

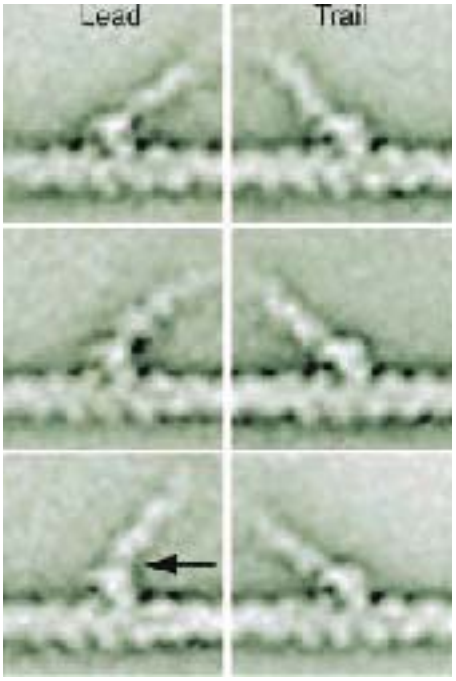


Fig. 1
The figure shows image averages of lead and trail heads on actin. Note these molecules are walking towards the left. The arrow points to the junction between the motor domain and lever arm where extra bending occurs.

Collaborators:
James Sellers, NIH; Howard White, Eastern Medical School; Joseph Chalovitch, Charlottesville.

Publications
Burgess, S., Walker, M., Wang, F., Seller, J. R., White, H. D., Knight, P. J., and Trinick, J. (2002). The prepower stroke conformation of myosin V. *Journal of Cell Biology* **159**, 983-991.

Burgess, S. A., Walker, M. L., Wang, F., Sellers, J. R., Knight, P. J., and Trinick, J. (2002). The pre-power stroke conformation of myosin V. *Biophysical Journal* **82**, 69.

Burgess, S., Yu, R., Walker, M. L., Trinick, J., Chalovich, J. M., and Knight, P. J. (2002). Structure of smooth muscle myosin in the switched-off state. *Biophysical Journal* **82**, 1734.

Funding
We acknowledge the support of the NIH (USA) and the BBRSC.

Self-association properties of the elastic region of the giant protein titin

Larissa Tskhovrebova, Ahmed Houmeida, Beatrix Thompson, Peter Knight and John Trinick

Titin is the largest known protein (~4 MDa). Single molecules span half striated muscle sarcomeres. In the myosin filament region of the sarcomere we have proposed that titin acts as a 'protein-ruler' to regulate exact myosin filament assembly. The remainder of the molecule forms an elastic connection between the end of the myosin filament and the Z-line. These connections are the main route of mechanical connectivity through relaxed muscle and they give muscle its passive elasticity. They also maintain myosin filaments midway between Z-lines during sarcomere length variations; this ensures an even development of force between the two halves of a myosin filament during active contraction.

Titin self-association

We have recently purified a 400 kDa proteolytic fragment of titin. N-terminal sequencing shows that the fragment begins at Ig domain number I22 in the molecule, which is the region of the molecule near the end of the thick filament *in situ*. We previously suggested that the six titin molecules that emerge from the thick filament end self-aggregate in register over about 100 nm to form distinctive structures called end-filaments with a 4 nm periodicity. In accord with this prediction, the 400 kDa fragment forms aggregates in low salt that are very similar in appearance to the native end-filaments.

Role of titin in muscle regulation

We have also suggested that the large charge in the part of the titin molecule that is elastic (the PEVK region) may affect calcium diffusion in the muscle, and thus that titin may have a role in controlling muscle contraction.

Publications

Tskhovrebova, L., and Trinick, J. (2002). Role of titin in muscle regulation. *Biophysical Journal* **82**, 1946.

Tskhovrebova, L., and Trinick, J. (2002). Role of titin in vertebrate striated muscle. *Proc. Roy. Soc. Lond. B* **357**, 199-206.

Funding

We acknowledge the support of the NIH (USA), BBRSC, British Heart Foundation and the Wellcome Trust.

Application of machine learning methods to aggregation, protein structure and SNP data

Vidhya Krishnan, Jennifer Siepen and David Westhead

In this post-genomic era the amount of biological data is increasing exponentially. Bioinformatics helps to extract relevant biological information from these voluminous data. Machine learning methods are computationally efficient techniques for learning valuable information from known data and can be used to make predictions on a new set of data. There are many machine learning approaches available such as neural networks, support vector machines (SVMs), decision trees and hidden Markov models, which are widely used in biological applications. In our group, different machine learning methods are implemented to address important biological problems.

Single nucleotide polymorphisms (SNPs) are subtle variations in a genomic DNA sequence of individuals of the same species. They play a key role in the pharmaceutical industry to understand variations in drug treatment responses between individuals at the molecular level. A large amount of available SNP data poses the questions: Can we predict SNPs; which are likely to be neutral; and which are likely to affect gene function? To address these questions a detailed comparison study is being done using decision trees and SVMs.

Decision tree and SVM approaches are also used to investigate protein structure. It is well established that recognition between exposed edges of beta sheets is an important mode of protein-protein interaction, and can have pathological consequences; for instance the aggregation of proteins into a fibrillar structure is associated with a number of diseases. Different protective mechanisms have evolved in the edge strands of beta sheets, preventing the aggregation and insolubility of many natural beta sheet proteins (Fig. 1). SVMs and decision trees are used to distinguish between edge strands and central strands based on amino acid sequence. This may be useful in the prediction of residues and mutations likely to be involved in aggregation and also as a first step in the prediction of folding topology.

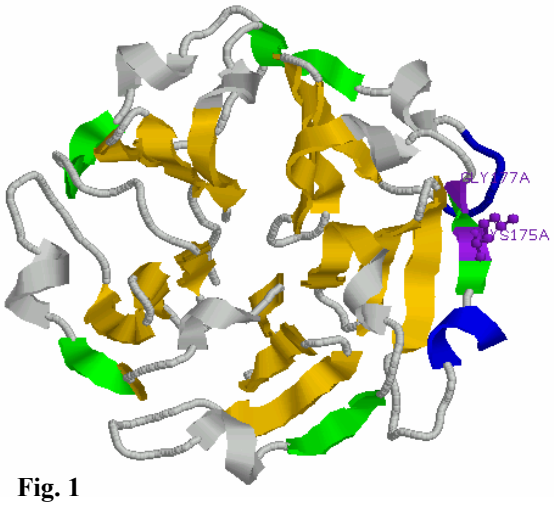


Fig. 1
Structure of the beta-propeller, 1TL2. Central β strands shown in yellow with edge strands of the blades in green. Further β sheet interactions are prevented by a charged lysine residue and a glycine bulge towards the strand edge and a nearby partially covering loop and collinear helix blocking the first β strand residue.

Publications:

Siepen, J.A. & Westhead, D.R. (2002) The Fibril_one on-line database: mutations, experimental conditions and trends associated with amyloid fibril formation. *Protein Science* **11,** 1862-1866

Collaborators:
Astrazeneca and Professor Sheena Radford, University of Leeds,

Funding:
MRC, Astrazeneca.

Automated metabolic reconstruction from genomic DNA

John W. Pinney and David R. Westhead

Metabolic reconstruction

The rapid proliferation of genome sequencing projects over the last ten years has resulted in an exponential growth in the amount of genomic DNA sequence available to biologists. The focus of genomics research is now moving towards the development of fast, accurate methods of extracting new knowledge from these data. One important target is the elucidation of an organism's metabolic pathway complement from its genome sequence, known as *metabolic reconstruction*.

Knowledge of the presence or absence of specific pathways in a given organism can help improve the quality of genome annotation by highlighting false positives and negatives in assigned gene function. If only one enzyme-encoding gene out of a pathway of several steps is found in an annotated genome, it is likely to be a mistaken assignment. Conversely, if all but one or two enzymes in a pathway have corresponding genes in the annotated genome, the missing steps are likely to be present amongst the unidentified genes and are worth hunting down.

Studying the metabolism of disease-causing organisms can also be an excellent means of identifying new drug targets. Many pathogenic bacteria and parasitic eukarya are the subjects of ongoing genome sequencing projects. If metabolic pathways can be identified which are essential in the pathogen but absent in the host, new drugs targeting the enzymes in these pathways are likely to be very effective.

The metaSHARK project

We are currently developing a comprehensive suite of programs for the representation and analysis of metabolic networks. The **meta**bolic **s**earc**h** **a**nd **r**econstruction **k**it includes an object-oriented database to store knowledge about networks of chemicals and reactions, as well as an automated system to search an unannotated genome for genes with significant similarity to known enzymes from other organisms. These genes are assigned a confidence score based on the strength of their similarity to the test sequences.

The structure of such a metabolic network is very effectively modelled by a type of graph called a *Petri net* (Fig. 1). This is a graph with two types of nodes, called *places* and *transitions*, linked by directed arcs. Chemicals are represented by place nodes, and reactions by transition nodes. Petri nets have been used extensively in computer science to represent complex systems, and have proven to be useful in studies of many kinds of biological network.

Analysis of the Petri net structure can reveal sets of reactions called *elementary modes* - pathways which are stoichiometrically and thermodynamically feasible. The evidence for the presence of each of these modes in an organism's metabolism can then be assessed using the scores of the mode's component reactions. This sort of information is vital for the purposes of identifying good drug targets in pathogens such as *Plasmodium falciparum*, the parasite causing the most virulent type of malaria. These organisms have evolved to defend themselves against the body's immune responses. As such, they are likely to possess robust backup systems so that essential metabolic processes operate even when the main pathways

are blocked. These backup pathways may provide a route of escape from a drug's effect, and could lead to the rapid development of drug resistance in the pathogen population.

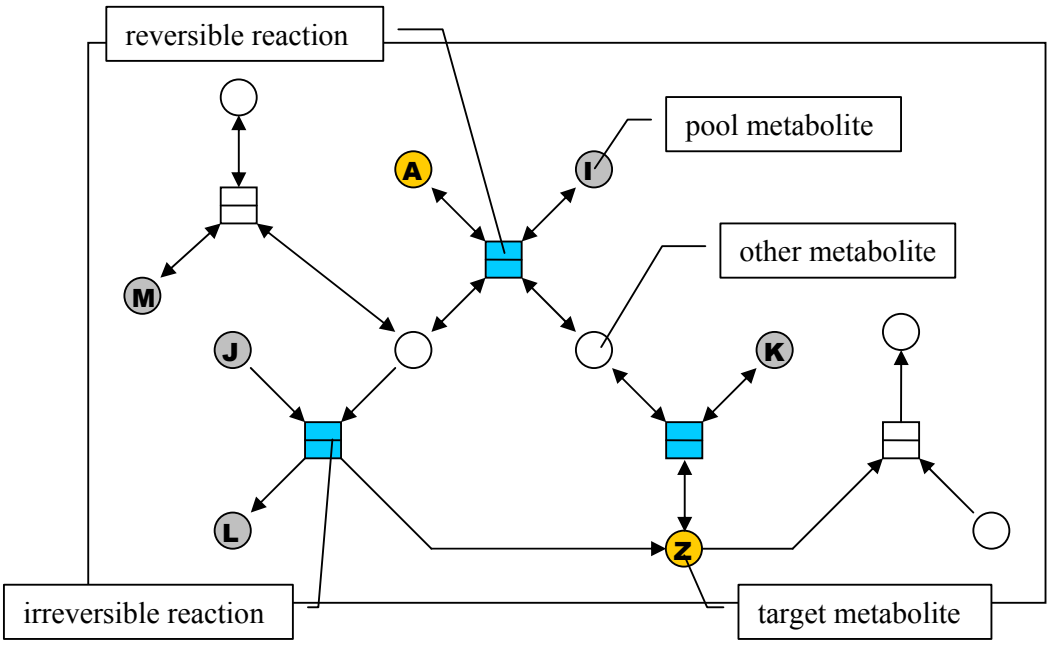


Fig. 1: Petri net diagram showing part of a hypothetical metabolic network. Metabolites are represented as circles; reactions as squares. The arcs between reactions and metabolites show the possible directions of each process: most metabolic reactions are reversible, but some are irreversible. If we definine a set of ubiquitous "pool" metabolites (shown in grey, eg. water, ATP, NADH etc.) and two or more "target" metabolites (shown in yellow, eg. glucose and pyruvate would be the target metabolites for glycolysis), the Petri net can be analysed to give a set of "elementary modes" which represent all thermodynamically and stoichiometrically feasible pathways between the target metabolites. As well as reproducing the well-known "textbook" versions of a pathway, such analyses often reveal previously unknown routes involving unexpected enzymes. In the example above, the reactions highlighted in blue make up an irreversible elementary mode with equation $A + I + J + K \rightarrow 2Z + L$

However, elementary mode analysis should be able to pinpoint enzymes which cannot be bypassed. These would be the most vulnerable parts of the metabolic network to attack, and hence the best choice of drug target.

Pathways in *Plasmodium*

In collaboration with Dr Glenn McConkey of the School of Biology, we are working towards the metabolic reconstruction of *Plasmodium falciparum* using the metaSHARK system. In particular, the pathways associated with nucleotide metabolism are of great interest since they are known to differ significantly from those in man. It is hoped that analysis of the recently-released complete parasite genome will reveal novel genes and pathways which can be verified by RNAi and DNA microarray experiments in Dr McConkey's group, ultimately leading to the identification of new drug targets for this killer disease.

Collaborators

Glenn A McConkey, School of Biology, University of Leeds.

Funding

We thank the MRC for funding this project.

Derivation and refinement of global sequence motifs for the integral membrane proteins

Michael Sadowski, J. Howard Parish and David R. Westhead

Previously, an alignment algorithm for matching flexible, sequence-length motifs ('signatures') to protein sequences has been described and a method of deriving these motifs using contact information derived from the crystal structures of various protein families was reported subsequently.

This project extends the previous work by providing a method of deriving such motifs for families where little or no structural information is available. Multiple sequence alignments constructed using predicted secondary structure information, are created for a minimal set of sequences from the family of interest, motifs being selected from these on the basis of conservation within columns of the alignment.

Motifs have been derived for the G-protein coupled receptors by this method and a selection process implementing a genetic algorithm, the details of which are in the forthcoming paper by Sadowski & Parish (2003). Current work is focused on the development of a genetic algorithm refinement process and on the extraction of motifs for bacterial drug transport proteins and their relatives.

Publications

Daniel, S.C., Parish, J.H., Ison, J.C. & Findlay, J.B.C. (1999): Alignment of a sparse protein signature with protein sequences: application to fold recognition for three small globulins. *FEBS Lett.* **459**, 349-352

Ison, J.C., Blades, M.J., Bleasby, A.J., Daniel, S.C., Parish, J.H. & Findlay, J.B.C. (2000): Key residues approach to the definition of protein families and analysis of sparse family signatures. *Proteins* **40**, 330-341

Sadowski M.I. & Parish J.H. (2003): Automated generation and refinement of protein signatures: case study with G-protein coupled receptors. *Bioinformatics*, in press.

Funding

We thank the MRC for funds to support this work.

Protein surface descriptions and function

James Bradford, Nicola D. Gold, Steven J. Pickering and David R. Westhead

Macromolecular docking

We are currently developing a novel machine vision-based algorithm to help solve the "docking problem", which can be defined as follows: given the coordinates of two proteins, predict whether they interact and the structure of their complex. Our method uses graph theory to identify regions of shape complimenarity between two protein surfaces, analogous to finding areas of similarity in the protein surface comparison project (see below). Recently, we have been looking at ways of using biological information to facilitate our algorithm and one strategy has been to predict the binding sites on the two proteins. Most binding site prediction methods assume that the largest cluster of highly conserved residues on the surface of the protein corresponds to the interface. With this in mind, we have carried out a thorough and systematic sequence-structure study of the pattern of conservation on the surface of seven protease-inhibitor docking test cases and related the results to binding site prediction. The proteases all displayed a significantly higher concentration of conserved vertices at the interface compared to the exposed surface in contrast to the inhibitors where the difference between interface and exposed surface was much less distinct and varied from case to case. Indeed, some inhibitors had a greater proportion of unconserved vertices at their interface than the exposed surface. We conclude that binding site prediction methods relying on a highly conserved interface would succeed when applied to our protease test cases but complications could arise with the inhibitors.

Protein surface comparisons

Locating a protein sequence or structure common to two or more proteins often implies that these proteins possess similar functionality. With this in mind, protein function is usually predicted through protein sequence or backbone structure comparison. However, protein functions such as catalysis or molecular recognition occur predominantly on or near the protein surface. Our current work therefore concentrates on protein surface matching, a novel technique that has the power to reveal further functional relationships between proteins, not necessarily apparent from comparisons of the protein fold, by using protein surface comparisons to identify similarly shaped regions of surface.

We have identified residue coordinates important in protein function and stored this information in a database SITESDB containing approximately 50,000 functional sites (ligand binding and/or active sites). This information has also been used to create a database of protein functional site surfaces. We have also developed three methods to compare a query functional site in a pair-wise fashion to the database of known templates all based on techniques in graph theory. In addition, to reduce program execution time the methods have been implemented using MPI (Message Passing Interface) for use on parallel machines such as a Beowulf cluster.

In each of the methods the first step is to find identifying features to be used by the graph comparison algorithm.

Method 1: Each functional site residue is represented with a co-ordinate at the residue's centre of mass.
Method 2: Following the generation of a Connolly protein surface for a functional site each functional site residue occurring on the protein surface is represented with a co-ordinate on the protein surface.

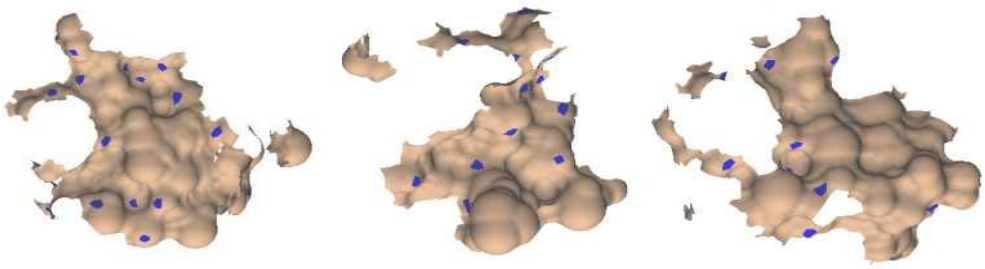
Method 3: Following the generation of a Connolly protein surface for a functional site a co-ordinate is used to represent certain shape properties (concave, convex, radius of curvature) or charge or hydrophobicity.

The co-ordinates of these features are then represented by vertices in mathematical graphs. Common vertices between two graphs are found using distance criteria and clique detection algorithms. The more vertices the two graphs have in common, the more similar the functional sites.

We believe that residues that lie on the protein surface at a functional site are more useful in assigning biochemical function than the use of the whole sequence. Using these residues allows for the detection of functional sites from evolutionarily related proteins where the sequence remains conserved at the functional site (method 2). For detecting more distantly related, or unrelated proteins, with similar function, we deduce the overall shape of the functional site from the orientation and position of various concave and convex regions within the site (method 3).

Early results indicate that we can successfully match related proteins within the same SCOP family as the query (Fig. 1). There are some unrelated matches that will be investigated further. We are also able to predict the function of sites that have not yet been classified in SCOP.

Fig. 1: Selected surface features of three eukaryotic serine proteases are shown in blue.



Trypsin (1btx) Elastase (1bma) Chymotrypsin (2gct)

Publications

Pickering, S.J., Bulpitt, A.J., Efford, N., Gold, N.D. and Westhead, D.R. (2001) AI-based algorithms for protein surface comparisons. *Computers and Chemistry*. **26**: 79-84.

Funding

We thank the BBSRC for funding this work.

Astbury Seminars 2002

Thursday, 17th January 2002
Dr. Louise Serpell, (Structural Medicine Unit, Cambridge Institute for Medical Research)
"Visualisation of the structure of amyloid"

Thursday, 14th February 2002
Dr. Adrian Mulholland, Department of Chemistry, Bristol)
"Modelling biological molecules and simulating enzyme-catalysed reactions"

Wednesday, 20th February 2002
Dr. Ben Luisi, (Department ofBiochemistry, Cambridge)
"Structural studies of RNA degradosome components"

Thursday, 7th March 2002
Prof. Randy Read, (Department of Haematology, Cambridge Institute for Medical Research)
"Shiga-like toxins and drug design"

Thursday, 21st March 2002
Prof. Ron Kopito, (Department of Biological Sciences, Stanford University)
"The dark side of folding: protein aggregation and neurodegenerative disease "

Thursday, 28th March 2002
Prof. Mike Ferguson, (Biological Chemistry & Molecular Microbiology, Dundee)
"Glycosylphosphatidylinositol (GPI) biosynthesis: a target for anti-parasite drug development"

Thursday, 2nd May 2002
Prof. Mervyn Miles, (SPM Group & IRC for Nanotechnology, University of Bristol.)
"Recent developments in scanning probe microscopy"

Friday, 5th July 2002
Prof. Leslie Poole, (Medical Center, Wake Forest University, North Carolina)
"The biochemical, physical and structural properties of AhpC, a bacterial peroxiredoxin"

Tuesday, 23rd July 2002
Prof. Bryan Cullen, (Duke University Medical Center)
"HIV-1 gene regulation: mechanisms and targeted inhibition"

Thursday, 22nd August 2002
Dr. David Millar, (Department of Molecular Biology, The Scripps Research Institute, La Jolla, California.)
"Application of time-resolved fluorescence spectroscopy to studies of DNA-protein interactions and RNA folding"

Tuesday, 1st October 2002
Prof. Phoebe A Rice, (Dept. of Biochemistry & Molecular Biology, The University of Chicago)
"Flipping and bending DNA: structural views of site-specific recombination and DNA bending"

Thursday, 17th October 2002
Prof. Erich Wanker, (Max Planck Institute for Molecular Genetics, Berlin)
"Huntington and other neurodegenerative disorders"

Thursday, 24th October 2002
Prof David Weaver, (Molecular Modeling Laboratory, Department of Physics, Tufts, USA)
"Using sequence and structure to predict protein folding kinetics"

Thursday, 7th November 2002
Prof. John W. Weisel, (University of Pennsylvania)
"Structure and interactions of blood clotting proteins and supramolecular complexes by TEM"

Thursday, 5th December 2002
Dr. David Klenerman, (University of Cambridge Chemical Laboratory)
"Single molecule spectroscopy"

We would like to acknowledge our corporate sponsors for 2002:

DIONEX Novagen® BELLE TECHNOLOGY Waters