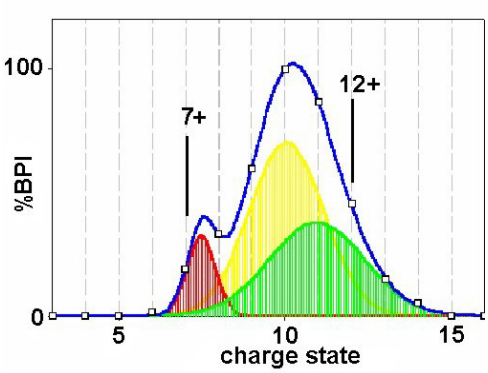
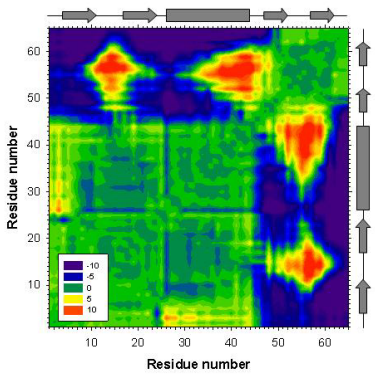
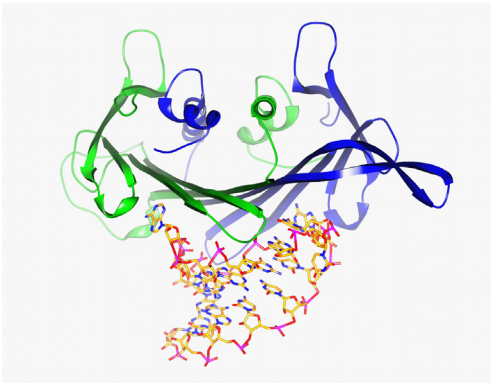
Astbury Centre for Structural Molecular Biology

University of Leeds







Annual Report 2004

Front cover illustration: A collage of pictures illustrating the work of the Astbury Centre. Upper left: The contribution of the charge state ions to the folded (red), partially folded (yellow) and unfolded (green) β_2-microglobulin conformers determined by ESI-Mass Spectrometry (see page 9); Upper right: Mechanical unfolding of proteins: a contour plot showing the difference in distance between every pair of amino-acids in protein L after pulling the protein apart by extensions of 1.6 Å before, and 1.6 Å after, the mechanical unfolding event. Pairs of residues that move further apart from each other during unfolding are coloured purple to green (-10 to 0Å), those that become closer to one another are shown in green to red (0 to 10 Å) (see page 23); Lower right: Dynamics on the ps-ns timescales for the backbone of *apo*-MS2W82R as detected by $\{^1H\}$ -^{15}N Heteronculear nOe experiments. The image shows a tubes representation of the backbone of MS2W82R [PDB code 1MSC] with $\{^1H\}$-^{15}N heteronuclear nOe intensity shown by shading from white to black. White indicates little mobility [nOe ~ 0.8-0.9] and black indicates significant mobility [minimum nOe = 0.13] (see page 107); Lower left: Ribbon diagram of an AB coat protein dimer from the MS2 phage (subunit A in blue, B in green) complexed with the wild-type MS2 RNA stem-loop operator (shown in stick format) (see page 95).

Acknowledgement
The Astbury Centre for Structural Molecular Biology thanks its many sponsors for support of the work and its members for writing these reports. 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

The Laws of Thermodynamics tell us that change is the only constant in life, but this last year has seen a great many changes within the Astbury Centre for Structural Molecular Biology and in the wider University. Two new University Interdisciplinary Institutes (UIIs) have been created, operating within the Astbury Centre. These are the Institute of Molecular Biophysics (IMB), headed by Alastair Smith and Sheena Radford, and the Institute of Bionanosciences (IoB), headed by myself. Together these two new ventures represent an investment of over £4.5 M in our science base, and we are very grateful for the University's commitment to our activities that this represents. At the time of writing, further institutional investment has been committed to develop the field of Chemical Genetics/Biology, lead by Adam Nelson.

The IMB supports University Research Fellowship positions, and we welcome Drs Emmanuele Paci, formerly at the University of Zurich, and Lars Jeuken, who was in the School of Physics, as the first appointees under this scheme. Dr David Brockwell has also been appointed to a lectureship within the IMB, supported by an EPSRC postgraduate training scheme award. The IMB is developing instrumentation for probing events in biological systems and modelling the processes that are revealed by these techniques. The IoB aims to exploit our fundamental knowledge about the structure and function of biomolecules and their complexes to create novel entities that might eventually have applied uses. Funding is supporting the purchase of new instrumentation, and the infrastructure staff to support them, as well as the creation of a core Bionanoscience laboratory housing dedicated research staff to assist projects in this area, supported by a large number of IoB-funded PhD students. I am confident that future editions of this Report will be full of the fruits of these new developments.

Members of the Astbury Centre continue to be very successful in raising external grant income. A special mention goes to Steve Baldwin, Simon Phillips, Mike McPherson, John Trinick, Steve Homans and Peter Henderson who are the Leeds' component of a national consortium to address the three-dimensional structures of membrane proteins. This Structural Genomics bid received £7.5 M from the BBSRC, of which the Leeds' share is £2.5M.

In a wider context, the University has a new Vice-Chancellor, and we welcome Prof. Michael Arthur to Leeds and applaud his vision of the pursuit of excellence across the University's activities. The Astbury Centre hosted Michael to a tour of our facilities before Christmas, and we were delighted to hear of his understanding and support for the types of interdisciplinary activity that the Astbury Centre represents. This is also an appropriate point to acknowledge the support of Michael's predecessor, Sir Alan Wilson, over the years. We wish Alan every success in his new role. 2004 was also a special year because the Centre hosted a visit from Bill Astbury's daughter and grand-daughter who were visiting Leeds from Sweden, where they now live. It was fascinating to hear first-hand accounts of Astbury's research and we're delighted that his family take such a continuing interest in our work.

The pages that follow describe 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. As always I am 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. The Astbury Centre has always been outward looking and this tradition continues with the many external collaborations acknowledged in these pages, from both 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 via 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. The Centre also continues to host a very successful seminar programme that illustrates aspects of work within the Centre.

Finally, I must thank the people who put so much effort into preparing this Report, especially Alan Berry and Rachael Taylor.

This annual report is also available as a 15.7Mb PDF document that can be downloaded from our web site.

Peter G. Stockley
Director, Astbury Centre for Structural Molecular Biology
Leeds, March 2005

Contents

Contributions indexed by Astbury Centre Principal Investigator

Analytical centrifuge facility

Andy Baron and Peter Stockley

Introduction

The Centre has two Beckman XL-I analytical ultracentrifuges installed in the Wellcome Trust JIF Centre for Biomolecular Interactions. Both instruments are equipped with absorbance and interference optics, 4-place, and 8-place rotors, 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 species distribution, mass, degree of asymmetry and association constants of interacting species.

Work carried out in 2004

The facility was used by a number of researchers groups, some making extensive AUC studies while others did simple experiments to determine the properties in free solution of their material. The results from one of these latter experiments is shown as an example.

Sedimentation velocity analysis of a GFP fusion protein

This work was part of a postgraduate research programme into directed evolution of aldolases being carried out by Chris Plummer in the laboratory of Dr Alan Berry. In order to select soluble mutants of GatY, an insoluble tagatose bisphosphate aldolase, fusions with green fluorescent protein (GFP) were used. The GFP serves as a visual screen for soluble mutants. Insoluble mutants would be cleared into inclusion bodies where GFP would not have time to fold and so be non-fluorescent. The soundness of this strategy was tested by constructing a GFP fusion with a soluble aldolase of the same class, fructose bisphosphate aldolase (FBP aldolase). Sedimentation velocity analysis was performed on this construct to check that the oligomeric state of the aldolase-GFP fusion was the same as the aldolase alone - in this case a dimer.

Three samples of the protein at 0.1, 0.25 and 0.5 mg/ml were centrifuged at 35000 rpm for 5 hours. The changing distribution of protein was recorded by Rayleigh interference optics and radial absorbance scans at 390nm - the absorbance peak of the GFP used. The whole boundary profiles were used to calculate the distribution of sedimentation coefficients [c(s)] by fitting to the Lamm equation using the program Sedfit v 8.9 (P. Schuck, NIH).

In all 6 datasets there was one major component having a sedimentation coefficient lying in the range 6.9 to 7.4 Svedbergs, the value being inversely correlated with concentration, as expected. The area underneath these peaks is proportional to the concentration of the component (Fig 1.)

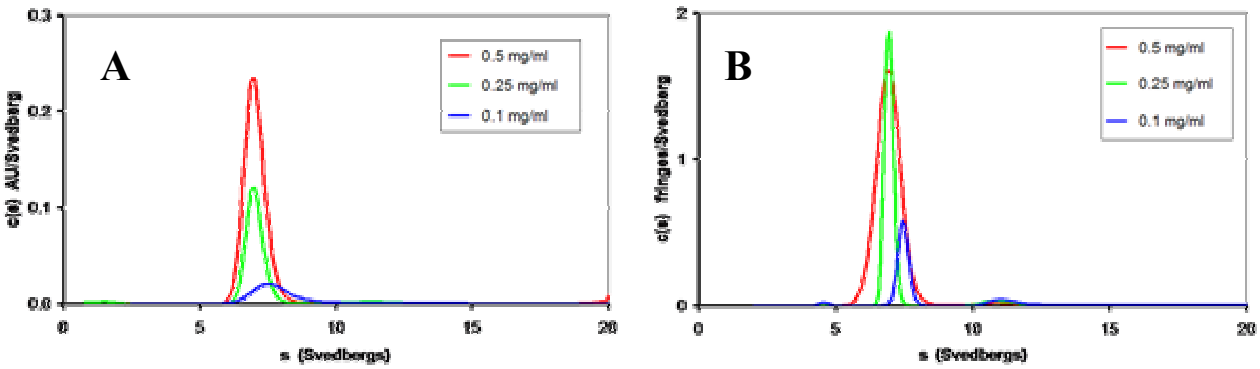


Fig 1. c(s) analyses A: Absorbance data, B: Interference data

Comparison of the sedimentation coefficient measured with that of a perfect sphere having the same mass as the protein indicated that a 7.1 S solute must be at least as large as the dimer (135 kDa), with a frictional ratio of 1.3. There were small amounts of material sedimenting faster than 7 S, but none smaller. In c(s) analyses of the interference data there is a peak at around 11 S which is absent in analyses of the A_{390} data. This is probably a non-GFP contaminant, or possibly unfolded fusion protein; the interference system measures concentration changes of total solute, whereas the absorbance system is selective.

This experiment showed that the natural oligomeric state of the control protein was maintained when fused with GFP. Therefore it is highly likely that the GFP fusions with GatY soluble mutants will also exist in the same oligomeric state as the unfused protein - in this case tetramer.

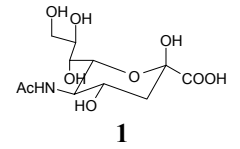
Funding
We gratefully acknowledge funding by HEFCE and the Wellcome Trust.

Structure-guided engineering of *N*-acetylneuraminic acid lyase

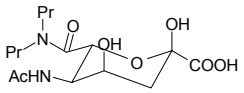
Gavin Williams, Thomas Woodall, Adam Nelson and Alan Berry

Introduction

Analogues of *N*-acetylneuraminic acid (**1**, sialic acid, NANA), particularly 6-dipropylcarboxamides (such as (**2**)), have been found to be selective and potent inhibitors of influenza sialidase. Sialic acid analogues are, however, difficult to synthesise by traditional chemical methods and the enzyme *N*-acetylneuraminic acid lyase (NAL) has been used for the synthesis of a number of analogues. Although a number of hexoses and pentoses and their analogues are substrates for NAL, condensations involving shorter aldehydes are less promising: L- and D-erythrose and threose react at between 0.3% and 5% of the rate of *N*-acetyl mannosamine, and two- and three-carbon aldehydes are not substrates. The activity of this enzyme towards 6-dipropylcarboxamides is also low. To overcome this problem, we have used structure-guided saturation mutagenesis to produce variants of NAL with improved activity and specificity towards 6-dipropylcarboxamides.



1



2

Results

Using the known crystal structure of the *E.coli* and *H. influenzae* NAL (Fig.1), we identified three residues which contact the 6-glycerol moiety of sialic acid and reasoned that mutagenesis of these residues would produce variants with increased specificity towards analogues of sialic acid which had hydrophobic groups in place of the polar glycerol component. Residues Asp-191, Glu-192 and Ser-208 were mutated to all the other 19 amino acids by saturation mutagenesis to create three libraries, D191X, E192X and S208X.

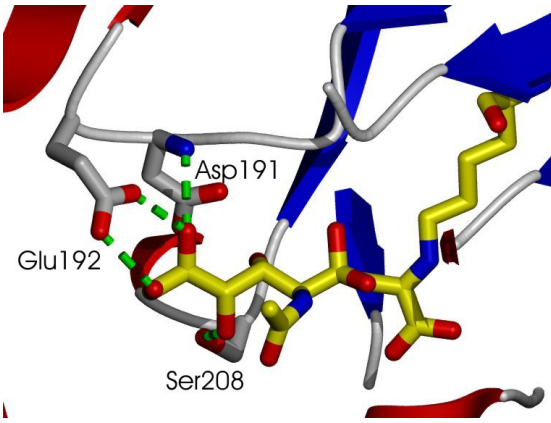


Fig. 1. The active site of *H. influenzae* NAL (from Barbosa *et al J.Mol. Biol*. (2000) **303**, 405-) showing the interactions of Asp-191, Glu-192 and Ser-208 with the glycerol moiety of the substrate.

Members of each library were tested for their ability to cleave the 6-dipropylamide (**2**). Only substitution at position 192 produced significant improvements in activity towards the dipropylamide, and a number of substitutions at this position resulted in significant switches in substrate specificity towards compound **2** (Fig 2A). One variant, E192N, was purified and characterized and showed a 49-fold improvement in catalytic efficiency towards the target analogue (**2**) and a 690-fold shift in specificity from sialic acid towards the analogue. The breadth of substrate specificity of the E192N variant was assessed by incubating the enzyme with pyruvate and various aldehydes and the reactions were followed synthetically. The results in Table 1 show that the E192N variant is a general purpose aldol catalyst for the production of a wide range of tertiary amide sialic acid analogues.

Another important property of any useful biocatalyst is the stereochemical purity with which the reaction is carried out. Naturally occurring NAL exhibits only poor facial selectivity during carbon-carbon formation, and as such, its scope as a general biocatalyst is limited, since products which are neither kinetic nor thermodynamically preferred are difficult, if not impossible to isolate. Engineering the stereochemical course of the NAL catalysed would

remove this limitation. We have used directed evolution to create a pair of stereochemically complementary variant lyases for the synthesis of sialic acid mimetics.

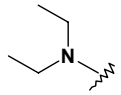
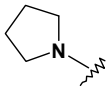
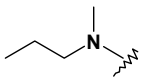
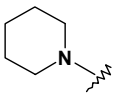
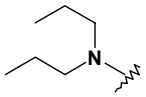
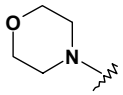
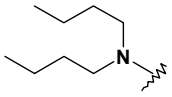
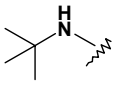
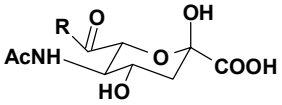
R-	Yield	R-	Yield
	37%		48%
	42%		55%
	42%		47%
	66%		35%

Table 1. Substrate specificity profile of the E192N variant NAL.
Aldehydes of general structure



were synthesized and incubated with pyruvate and the E192N enzyme. The yield of purified product is reported.

Initially, error-prone PCR identified functional residues that were located in the active site of NAL, and subsequently this inspired an intense program of structure-guided saturation mutagenesis (Fig. 2). Finally, two variants were obtained which were 48-fold and 52-fold stereoselective towards products with *R*- and *S*-configuration at C4 of the product, respectively. Wild-type NAL cannot be used for the synthesis of a 6-dipropylamide mimetic of sialic acid with *R* configuration at C4 because the product is not kinetically favoured nor thermodynamically more stable than the *S* product. However, the evolved *R* selective variant was used to synthesise this *R* configuration product allowing isolation of the rare diastereoisomer. The conversion of an essentially non-stereoselective aldolase into a pair of complementary biocatalysts will be of enormous interest to synthetic chemists, and these novel biocatalysts will be used for the synthesis of a range of clinically relevant sialic acid analogues.

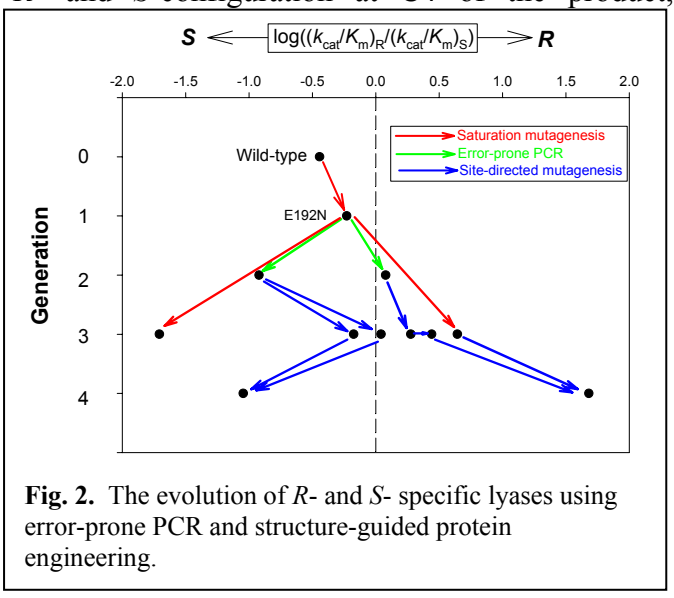


Fig. 2. The evolution of *R*- and *S*- specific lyases using error-prone PCR and structure-guided protein engineering.

Publications

Williams, G.J., Nelson, A.S. & Berry, A. (2004) Directed evolution of enzymes for biocatalysis and the life sciences. *Cell. Mol. Life. Sci.*, **61**: 3034-3046.

Woodhall, T., Williams, G.J., Berry, A. & Nelson, A.S. (2005) Directed evolution of an aldolase for the parallel synthesis of sialic acid mimetics. *Ang. Chem. Int Ed*, **44**: 2-5.

Woodhall, T., Williams, G.J. Berry, A. & Nelson, A. (2005) Synthesis of screening substrates for the directed evolution of sialic acid aldolase: Towards tailored enzymes for the preparation of influenza A inhibitor analogues. *Org. Biomol. Chem*, **In press.**

Funding

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

Structural studies of thermostable mutants of fructose-1, 6-bisphosphate aldolase

Elizabeth Bennett, Simon Phillips and Alan Berry

Introduction

Escherichia coli Class II fructose-1, 6- bisphosphate aldolase (FBP- aldolase) catalyses the reversible formation of fructose-1,6 bisphosphate from dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (G3P). This requires the formation of a carbon-carbon bond, a process executed in biological reactions with exquisite stereo-control and ease. These qualities are equally valuable in synthetic chemistry and the exploitation of biological catalysts (biocatalysts) in industrial processes is a desirable goal. However the implementation of biocatalysts faces a significant obstacle: protein stability is limited in non-biological environments, especially at high temperatures. Although many naturally thermophilic organisms exist, and function at temperatures exceeding 60°C (thermophilic) and 90°C (hyperthermophilic), most mesophilic-derived enzymes lose activity at temperatures above 40°. Comparative studies of proteins from mesophiles and thermophiles reveal structural differences, such as a greater extent of hydrophobic packing, especially in the core of the molecule, more salt bridges, improved capacity for hydrogen bonding and shorter loop regions. A family of thermostable variants of FBP-aldolase has been created by four rounds of directed evolution. The most thermostable enzyme shows a 90-fold increase in half-life at 53°C and a T_{50} which is increased by 12.2°C over the wild-type *E.coli* parent enzyme. Investigation of the residue changes introduced by mutagenesis ruled out formation of new disulphide bonds or changes in loop length, and modelling of the changes onto the known parent structure (*E.coli*) showed that significant mutations occurred distally from the active site.

Crystallographic studies of thermophilic FBP-aldolases created by directed evolution.

Structural studies of the four generations of mutant FBP-aldolases have been undertaken and crystallisation has been achieved for the first two generations using a combination of robotic screening and manual optimisation.

We have determined the structure of the first generation mutant, 1-44F2, using cryo-crystallography to 1.9Å resolution. The active site clearly shows the catalytic zinc and bound potassium and phosphate ions (Fig. 1), scavenged from the crystallisation conditions. The phosphate ions are located in positions equivalent to C1 and C6 of the enzyme's natural substrate, FBP. In addition, there is clear electron density close to the catalytic zinc, which most likely corresponds to glycerol bound from the cryo-protectant solution.

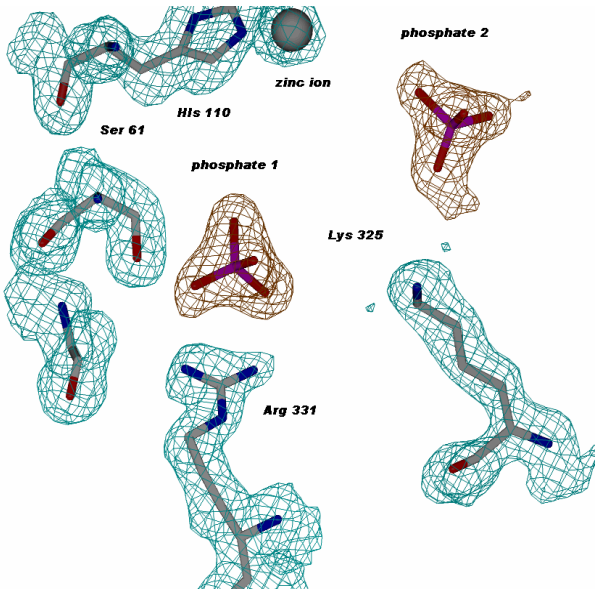


Figure 1 - Stick model showing two phosphate ions bound in the active site of 1-44F2 FBP-aldolase. Histidine 110 is shown co-ordinated to the zinc ion. Arginine 331and serine 61 surround phosphate 1. Lysine 325 is in the vicinity of phosphate 2. The molecule was crystallised in the absence of substrate and seems to have bound phosphate from the crystallisation mother liquor in place of fructose-1,6 bisphosphate.

This is the first time bound phosphate molecules have been seen in the Class II structure and their presence may offer a mechanistic insight.

Furthermore, the mutated residues are identifiable and all the mobile loops are clearly traceable. This is an important factor in the new structure as the enzyme is dynamic with loops known to flex during catalysis. A comparison of loop positions in the apo- versus substrate-bound forms may help to map the structural features of the catalytic cycle.

Crystals of the second generation mutant (Fig. 2) have been obtained under similar conditions and data have been collected at the Synchrotron Radiation Source, Daresbury, UK. Expression and purification of the remaining mutants is underway prior to crystallisation.

We have now reached an exciting stage of the research project where structural comparison of generations 1 to 4 of our directed evolution variants with each other and the wild-type parents should provide fundamental insights into mechanisms of thermostability.

Figure 2 – 2nd generation FBP-aldolase crystal grown in the absence of substrate or inhibitor. Conditions: 0.05M potassium dihydrogen phosphate, 23% PEG 8000 at 4°C in sitting drops.
Dimensions : 300x300x80microns.

Publications

Hao, J. and Berry, A. (2004) A thermostable variant of fructose bisphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents. *Protein engineering, design & selection* **17**, 689-697.

Funding

This work was funded by the BBSRC and the Wellcome Trust

Conformation dynamics and catalysis of aldolase studied by NMR

Tom Burnley, Stephan Paisey, Arnout Kalverda, Steve Homans and Alan Berry

Introduction

Conformational changes appear vital for both enzyme catalysis and regulation, however these essential motions are poorly understood at present. NMR relaxation experiments allow residue specific characterization of these dynamic motions. *Escherichia coli* Class II fructose-1,6-bisphosphate aldolase (FBP-aldolase) provides an excellent system, as it has been extensively studied at Leeds University via X-ray crystallography and site directed mutagenesis, coupled with kinetic analysis. These studies have revealed the essential role played by the protein's flexible loops during catalysis. FBP-A adopts the $(\alpha/\beta)_8$ barrel fold which is a common and versatile architecture, representing 10% of all known enzyme structures. Enzymes exhibiting an $(\alpha/\beta)_8$ barrel fold have been found for 61 different types of E.C. number, including all primary classes, with the exception of ligases. However, at present, no high resolution dynamic studies have been completed for such systems. The flexible structure provides an excellent platform for evolving novel enzymes and a thermostable mutant has already been previously evolved in Leeds. This work will provide the first general information on dynamics during catalysis of a "large" enzyme by modern NMR methods, and will provide the basis for the development of further novel catalysts.

Backbone amide resonance assignment

FBP-aldolase forms a 78kDa dimmer, and, as such, its size represents a significant challenge to study by NMR. To date, there only is one example of a larger polypeptide chain that has been fully assigned. A combination of HNCA, HN(CO)CA, HNCO, HN(CA)CO and HSQC TROSY based experiments (Fig. 1) using an isotopically enriched [^{2}H, ^{15}N, ^{13}C]-FBP-aldolase sample have been completed. This has resulted in a partial backbone amide assignment, which includes residues from the major β5-α7 flexible loop. This has been aided and confirmed via a number of selectively labeled samples including [^{15}N-Lys]-FBP-aldolase, [^{15}N-Glu]-FBP-aldolase, [^{15}N-Tyr]-FBP-aldolase and [^{15}N-Val]-FBP-aldolase samples. Residual dipolar coupling measurements have been measured on both [^{2}H, ^{15}N, ^{13}C]-FBP-aldolase and [^{15}N-Lys]-FBP-aldolase aligned in a dilute crystalline media which, in conjunction with previously obtained crystal structures, will assist in further backbone assignment.

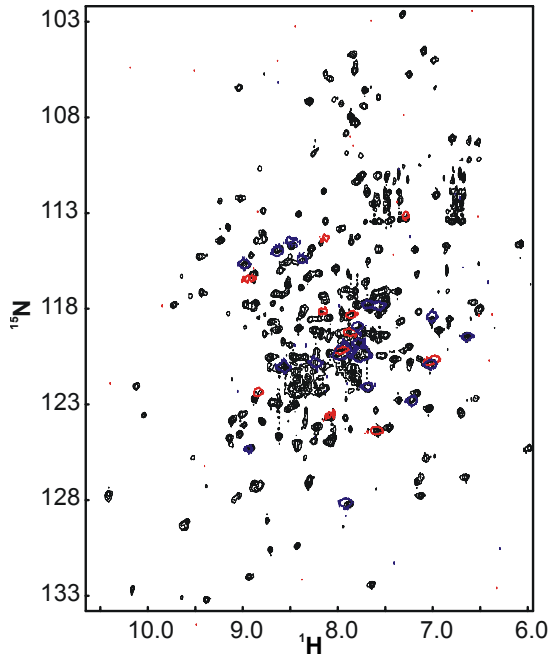


Fig 1. HSQC spectra of the amide resonances in [^{2}H, ^{15}N, ^{13}C]-FBP-aldolase measured at 900MHz are shown in black. Overlaying resonances from [^{15}N-Lys]-FBP-aldolase and [^{15}N-Tyr]-FBP-aldolase are shown in blue and red respectively (measured at 750MHz).

Characterization of nanosecond motions

T1, T2, and hetero-nuclear NOE relaxation rates have been attained and, for assigned residues, nanosecond motions have been quantified using Lipari-Szabo model-free formalism. This has confirmed the increased mobility of the β5-α7 loop with respect to other more stable regions of protein's architecture. Relaxation data has also been obtained in presence of a substrate analogue and inhibitor, phosphoglycolohydroxamate (PGH) (Fig. 2),

resulting in significant changes in both chemical shift and relaxation times. Relaxation data will also be obtained for natural substrate complexes, dihydroxyacetone phosphate, glyceraldehyde 3-phosphate, and FBP. This, coupled to a full assignment, will allow a complete description of enzyme motions during the catalytic cycle.

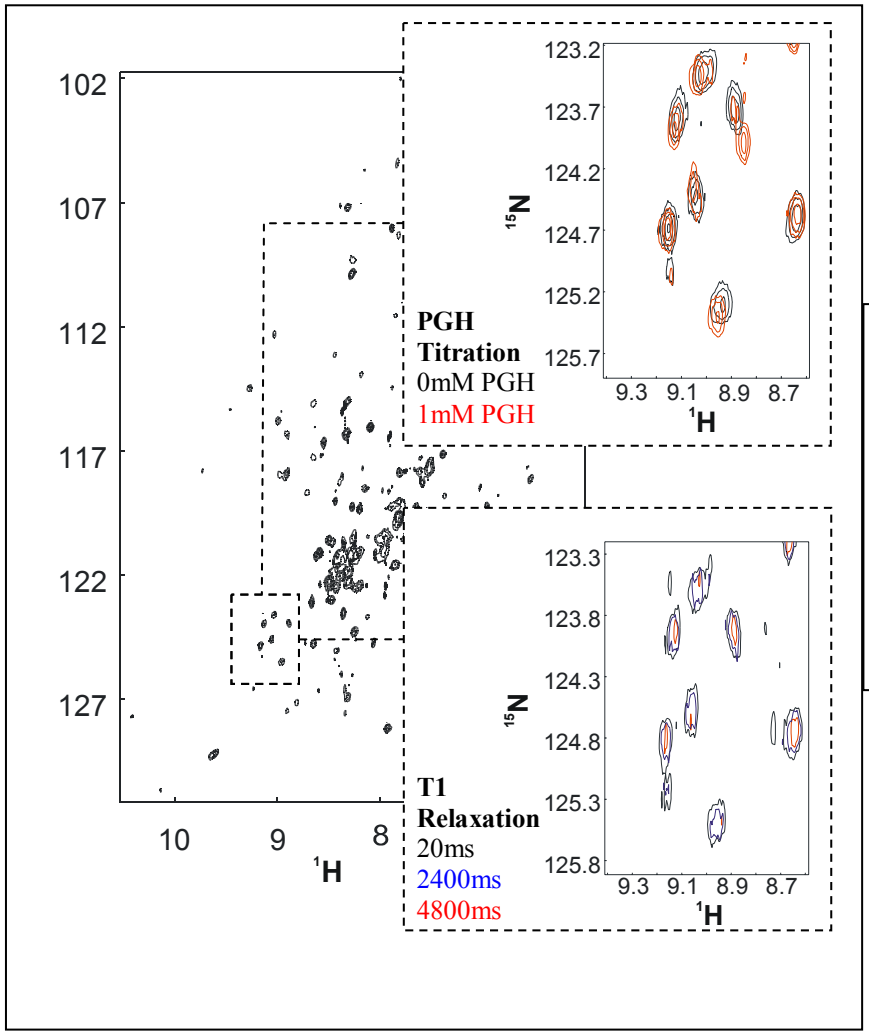


Fig 2. Example region of [^{2}H, ^{15}N, ^{13}C]-FBP-aldolase HSQC spectrum which shows, top, changes in amide resonance chemical shift in the presence and absence of PGH and, bottom, differing longitudinal (T1) relaxation rates for the uncomplexed form. All data was collected at 750MH.

Funding
This work is funded by the BBSRC and the Wellcome Trust.

Mass spectrometry facility

Alison E. Ashcroft

Overview of facility
The Mass Spectrometry (MS) Facility has a **Q-Tof** orthogonal acceleration quadrupole-time-of-flight tandem instrument with nano-electrospray ionisation (ESI) and on-line capillary HPLC, a **Platform II** (ESI) quadrupole instrument with on-line HPLC and CE, a **TSQ 7000** ESI tandem quadrupole instrument, and a surface enhanced laser desorption ionisation/matrix assisted laser desorption ionisation (SELDI/MALDI) **ProteinChip** mass spectrometer.

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, and also with other groups and external collaborators.

Research
The research involves the application of MS to the structural elucidation of biomolecules:

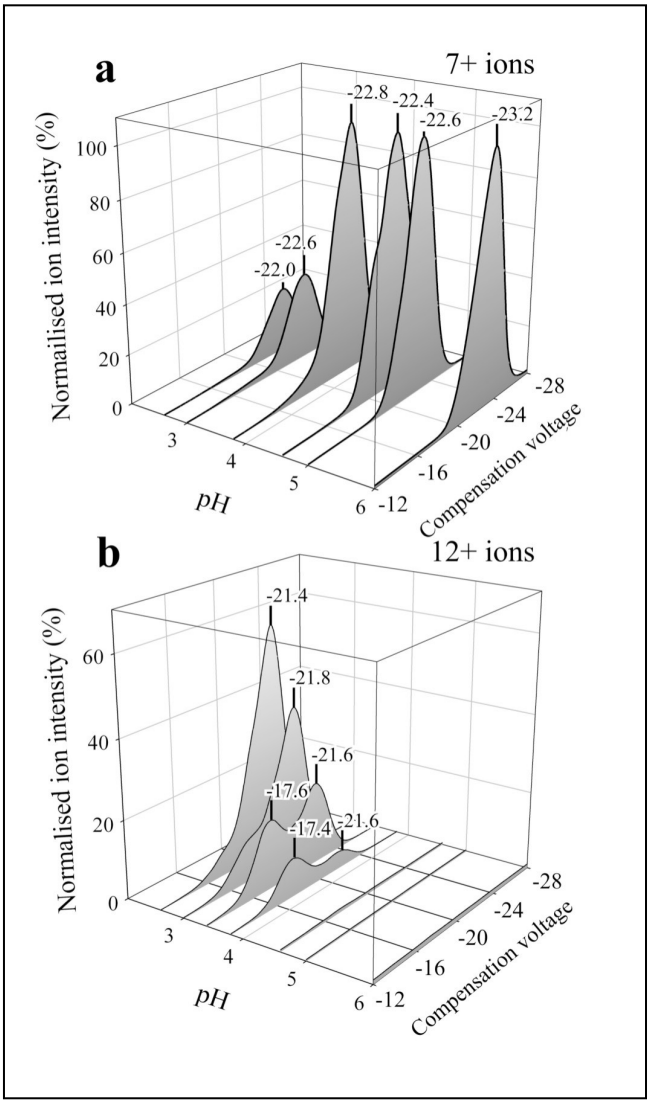
i). Protein folding. Protein folding is an intriguing area of biochemistry and protein mis-folding is thought to be a contributing factor to several diseases. Working with Sheena Radford's group, ESI-MS is being used to investigate β_2-microglobulin conformations using charge state distribution analysis,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

ii). Protein-ligand non-covalent interactions and macromolecular assembly. In collaboration with Peter Stockley, Sheena Radford and Nicola 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 amyloid studies.

iii). Reaction monitoring. We use MS to measure the uptake of ATP by the muscle protein myosin (Howard White (Eastern Virginia Medical School, USA) and John Trinick). Myosin has a motor domain that interacts with filaments of F-actin and splits ATP to generate force and movement. We also monitor enzyme proteolyses on-line.

iv). Structural elucidation and proteomics. Tandem MS (MS/MS) sequencing of proteins and peptides is an important bioanalytical technique. Several proteomics-related projects are in progress, the functional analysis of preproneuropeptide genes from the *Drosophila* genome (Elwyn Isaac, Biology), an investigation into parasite manipulation of host sex (Alison Dunn, Biology) and studies into the peroxisomal protein import machinery (Alison Baker). Other structural analysis projects include the characterisation of lipopolysaccharides (Deidre Devine, Oral Biology).

v). Mass spectrometry method development. To further our structural molecular biology studies, we are investigating the use of Ion Mobility Spectrometry coupled to mass spectrometry as a potential method to separate co-populated protein conformers and for the structural characterisation of small peptides. Using a high-Field Asymmetric waveform Ion Mobility Spectrometry (FAIMS) device we have separated co-populated β_2-microglobulin conformers at different pH according to the mobility of their charge state ions.



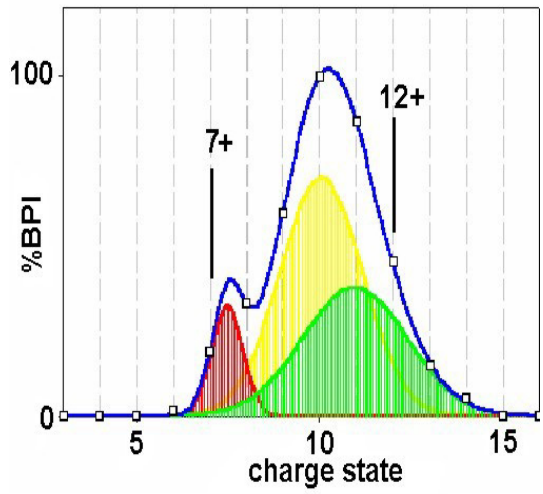


Fig. 1 (a,b) ESI-FAIMS-MS of β_2-microglobulin showing the 7+ (folded protein) and 12+ (partially folded and unfolded protein) charge state ions over the range pH2-6[3]. (c) contribution of the charge state ions to the folded (red), partially folded (yellow) and unfolded (green) β_2-microglobulin conformers determined by ESI-MS[2].

This is illustrated in Fig. 1. Panel (a) shows the 7+ charge state ions of β_2m (*m/z* 1695), and (b) the 12+ charge state ions of β_2m (*m/z* 989), acquired over the range pH 2.0 to 6.0 using ESI-FAIMS-MS with a compensation voltage (CV) ramp. Each trace is normalized to the base peak in the spectrum at that pH value. The data indicate that the 7+ charge state ions are associated only with a folded conformer (CV -22.6V) which decreases in population at low pH, whilst the 12+ ions are associated with both a partially folded (CV -17.5V) and an unfolded conformer (CV -21.6V) which are populated at low pH. This is in accordance with our previous work on charge state distributions[2]; an example of which (Figure 1(c)) shows the contributions from the various charge states, including the 7+ and 12+ ions, to the folded (red), partially folded (yellow) and unfolded (green) conformers at pH 2.6.

Publications

Rodgers-Gray, T.P., Ashcroft, A.E., Smith, J.E. & Dunn, A.M. (2004) Mechanisms of parasite-induced sex reversal in *Gammarus duebeni, Int. J. Parasitology*, **34,** 747-753.

Borysik, A.J.H., Radford, S.E. & Ashcroft, A.E. (2004) Co-populated conformational ensembles of β_2-microglobulin uncovered quantitatively by ESI-MS, *J. Biol Chem.,* **279**, 27069-27077.

Borysik, A.J.H., Read, P., Little, D.R., Bateman, R.H., Radford, S.E. & Ashcroft, A.E. (2004) Separation of β_2-microglobulin conformers by high-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled to electrospray mass spectrometry, *Rapid Commun. Mass Spectrom.,* **18**, 2229-2234.

Ashcroft, A.E. & Read, P.Peptide structures investigated by FAIMS coupled to ESI-MS, British Mass Spectrometry Society, 5th – 8th September 2004, The University of Derby, UK.

Khan, A., Ashcroft, A.E., Korchazhkina, O.V. & Exley, C. (2004) Metal-mediated formation of fibrillar ABri amyloid, *J. Inorg. Biochem.,* **98,** 2006 - 2010.

Collaborators/Group members

Antoni Borysik (BBSRC/CASE (Micromass UK Ltd) PhD studentship with Sheena Radford); Simona Francese (Marie Curie PhD studentship with Nicola Stonehouse & Peter Stockley); Michelle Morgan (MRC PhD studentship with Deidre Devine, Oral Biology); Rebecca Rose (BBSRC/CASE (Micromass UK Ltd) PhD studentship with Sheena Radford); Andrew Smith (Wellcome Trust post-doctoral fellow with Sheena Radford & Peter Stockley).

Funding

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

Herpes viral-host cell interactions which regulate viral gene expression

Jim Boyne, Rhoswyn Griffiths and Adrian Whitehouse

Gamma-2 herpes viruses are an increasingly important sub-family of herpes viruses with oncogenic potential, particularly as a result of the identification of the first human gamma-2 herpes virus, Kaposi's sarcoma-associated herpes virus (KSHV). KSHV has rapidly become the focus of intensive research as epidemiological studies suggest it is the etiologic agent of Kaposi's sarcoma, the most common AIDs-related malignancy. In addition, the presence of the virus has been detected in a variety of lympho-proliferative disorders including primary effusion lymphoma and multicentric Castleman's disease. However at present, analysis of KSHV gene function is hampered by the lack of a permissive cell culture system. Therefore, the ability to easily grow and manipulate the prototype gamma-2 herpes virus, HVS, *in vitro* has made this virus an attractive model for the analysis of gamma-2 herpes viruses in general. Therefore, we have a major research focus investigating the virus-host cell interactions which regulate the early events in gamma-2 herpes virus replication cycles, in particular HVS and more recently KSHV.

The interaction of the major transcription control protein, ORF 50 and viral promoters. The ORF 50 protein is the latent-lytic switch gene in gamma- herpes viruses and transactivates delayed-early gene expression. It functions as a sequence specific transactivator, binding to an A/T rich ORF 50 response element with DE promoters (Fig. 1a). We have demonstrated that ORF 50 contains a DNA binding domain that has homology to an AT-hook DNA binding motif. The AT-hook is a small DNA-binding protein motif that was first described in the non-histone chromosomal protein HMGA, and allows binding to the minor groove of short stretches of AT-rich DNA. The AT-hook has a core consensus sequence of Pro-Arg-Gly-Arg-Pro (with R-G-R-P being invariant), flanked on either side by a number of positively charged lysine/arginine residues. The core of the AT-hook peptide motif is highly conserved in evolution from bacteria to humans and is found in one or more copies in a large number of other, HMGA proteins, many of which are transcription factors or components of chromatin remodelling complexes. Deletion analysis of this domain reduces ORF 50-mediated transactivation of the DE ORF 6 and ORF 57 promoters by 100% and 90%, respectively. Furthermore, gel retardation experiments demonstrated that the AT-hook motif was required for binding the ORF 50 response element in the promoters of DE genes. Single site-directed mutagenesis of the AT-hook revealed that mutation of the glycine residue at position 408 to an alanine reduced ORF 50 transactivation of the ORF 57 promoter by 40%. Moreover, mutation of multiple basic residues in conjunction with the glycine residue within the core element of the AT-Hook abolishes ORF 50-mediated transactivation. In addition, the p50GFPΔAT-hook mutant was capable of functioning as a *trans*-dominant

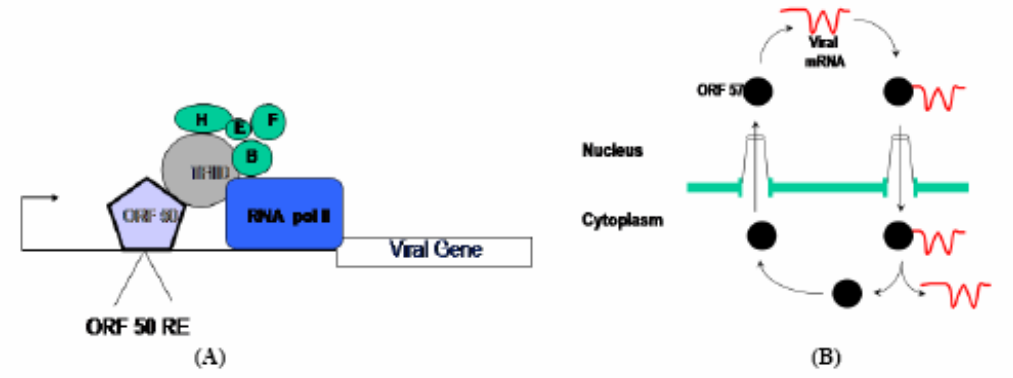


Fig. 1. (a) The ORF50 protein functions as a sequence specific transactivator. (b) ORF57 functions in viral mRNA nuclear export.

mutant leading to a reduction in virus production of approximately 50% compared to wild-type ORF 50.We are presently characterising these protein-DNA interactions using surface plasmon resonance (SPR) and fluorescence resonance energy transfer (FRET).

The interaction of the nucleocytoplasmic shuttle protein, ORF 57 and viral mRNA.
The ORF 57 protein encodes a nuclear cytoplasmic shuttle protein which mediates the nuclear export of viral mRNAs. We have recent analysis demonstrating that ORF 57 has the ability to bind viral RNA, shuttle between the nucleus and cytoplasm and is required for efficient nuclear export of viral transcripts (Fig. 1b). Moreover, we have shown that ORF57 shuttles between the nucleus and cytoplasm in an CRM-1 independent manner. ORF 57 interacts with the mRNA export factor REF and two other components of the exon-junction complex, Y14 and Magoh. The association of ORF57 with REF stimulates recruitment of the cellular mRNA export factor TAP, and HVS infection triggers the relocalisation of REF and TAP from the nuclear speckles to several large clumps within the cell. Using a dominant negative form of TAP and RNA interference to deplete TAP, we show that it is essential for bulk mRNA export in mammalian cells and is required for ORF57 mediated viral RNA export. Furthermore, we show that disruption of TAP reduces viral replication. These data indicate that γ-2 herpes viruses utilise ORF57 to recruit components of the exon-junction complex and subsequently TAP to promote viral RNA export via the cellular mRNA export pathway. We now aim to analyse the domains required for these interactions in more detail using structural analysis.

Publications.

Calderwood, M.A., Hall, K.T., Matthews, D.A. and Whitehouse, A. (2004). The Herpesvirus saimiri ORF 73 protein co-localises with host cell mitotic chromosomes and self associates via its C-terminus. *Journal of General Virology,* 85, 147-153.

Walters, M.S., Hall, K.T. and Whitehouse, A. (2004). The HVS ORF 50 (Rta) Protein encodes an AT-Hook required for binding to the ORF 50 Response Element in delayed early promoters. *Journal of Virology,* 78, 4936-4942.

Walters, M.S. Hall, K.T. and Whitehouse, A. (2005). The herpesvirus saimiri Rta protein autostimulates via binding to a non-consensus response element. *Journal of General Virology,* 86, 581-587.

Williams, B., Boyne, J.R., Goodwin, D.J., Roaden, L.R., Wilson, S.A. and Whitehouse, A. (2005). The prototype gamma-2 herpesvirus nucleocytoplasmic shuttle protein, ORF 57, transports viral RNA via the cellular mRNA export pathway. *Biochemical Journal,* In Press.

Wakenshaw, L., Walters, M.S. and Whitehouse, A. Herpesvirus saimiri Rta and cEBPα act synergistically to activate the DNA polymerase promoter. Submitted to *Journal of Virology.*

Collaborators.
Stuart Wilson, University of Sheffield.

Funding.
This work has been funded in parts by the BBSRC, MRC, YCR and Royal Society.

Molecular mechanism of Staphylococcal plasmid transfer

Jamie A. Caryl and Christopher 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*. Although not self-transmissible, it can be mobilised by a co-resident self-transmissible plasmid such as pGO1. Typically for 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.

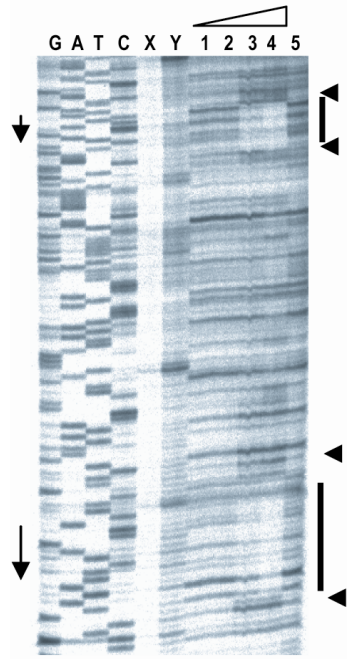


Fig. 1. DNaseI footprinting of MobC at *oriT*. Titration of MobC. Vertical bars indicate protection. Arrows indicate consensus direct repeats.

Recent findings

MobC binds to *oriT* at two sites containing a 9 bp consensus sequence (Fig. 1), and additionally within the *mobC* gene at a 7 bp conserved sequence. The addition of MobA to the MobC-*oriT* complex modified the observed footprint profile at the nick site and the adjacent MobC binding site. The *oriT* has recently been functionally characterised with respect to these sites. A 78 bp region of *oriT*, encompassing the nick site and the proximal MobC binding site (MCB) has been used as a nicking substrate (Fig. 2). Mutations in both MCB and the region modified by the addition of MobA (SRA) abolish nicking, thus demonstrating that nicking by the MobA relaxase requires the binding of MobC, and the presence of an intact SRA sequence.

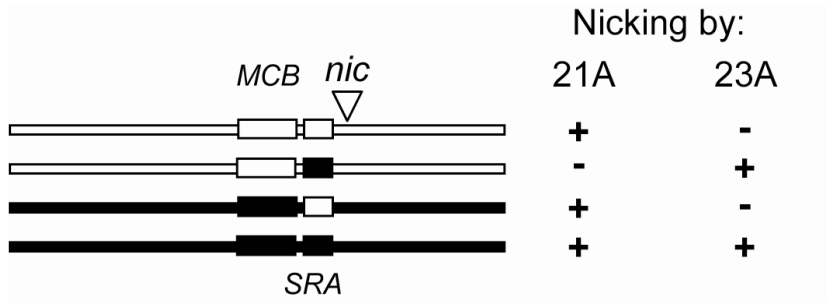


Fig. 2. Schematic detailing SRA-swap between pC221 *ori*T (white) and pC223 *ori*T (black), and the nickability of recombinant substrates by pC221 MobA 21A or pC223 MobA (23A)

The SRA sequence is not conserved between the related *oriT*s of pC221 and pC223. It was thought this sequence may therefore be functionally important in substrate recognition. The SRA regions (which differ by 4 bp) were swapped between the pC221 78 bp nick target and a comparable nick target of pC223 (Fig. 2). Exchange of this region was indeed found to swap

the substrate specificity, thus SRA represents a recognition sequence by which the relaxases recognise their specific substrate. Determination of the corresponding specificity domain in the MobA protein and the nature of any MobA-MobC interactions are currently under investigation.

Publications

Caryl, J.A., Smith, M.C.A. and Thomas, C.D. (2004). Reconstitution of a staphylococcal plasmid-protein relaxation complex *in vitro*. *J. Bacteriol*. **186**: 3374-3383.

Caryl, J.A. and Thomas, C.D. (2005) Initial events in small staphylococcal plasmid transfer. *In*, Thomas, C.M. *et al*. (eds). Plasmid Biology 2004: International Symposium on Molecular Biology of Bacterial Plasmids and other Mobile Genetic Elements. *Plasmid* **53**: p.47-48.

Acknowledgements & Funding

We thank Val Sergeant for technical support. This work has been funded by a grant from BBSRC.

DNA:DNA interactions mediate sequence specificity in the termination of plasmid replication.

Catherine Joce and Christopher D. Thomas

Introduction

RepD is the protein required for the initiation of replication of plasmid pC221. This replication is via a rolling-circle process; after nicking the (+) strand of the replication origin to provide a primer for DNA synthesis, the protein becomes covalently attached to the 5' end of the DNA. The covalent protein-DNA complex then tracks with the replisome until the next encounter with the origin. Here, RepD mediates a concerted cleavage-religation reaction to terminate replication: this includes the religation of both the displaced single-stranded circle and the nascent DNA strand of the double-stranded daughter plasmid.

The role of the inverted complementary repeat

The cleavage site falls between the arms of an inverted complementary repeat (ICR). Current literature is based on the presumption that nicking, and hence initiation of replication, requires the extrusion of a stem-loop structure as the functional DNA target. However, this model is challenged by the results of experiments with purified RepD protein and either double-stranded plasmids or single-stranded oligonucleotides as substrates. In both cases efficient nicking is still possible, even when the ICR is completely disrupted.

The ICR is a consequence of, not a requirement for, the action of RepD

We have recently re-appraised the religation reaction using a model heterodimeric intermediate, RepD/D*. After cleavage of a single-stranded oligonucleotide substrate with RepD, one of the protein subunits becomes substituted with a covalent DNA adduct at the active site (Fig. 1). This intermediate is isolated, then challenged with an acceptor oligonucleotide to reverse the reaction; religated products are amplified in a variation of the SELEX approach prior to cloning and DNA sequence analysis.

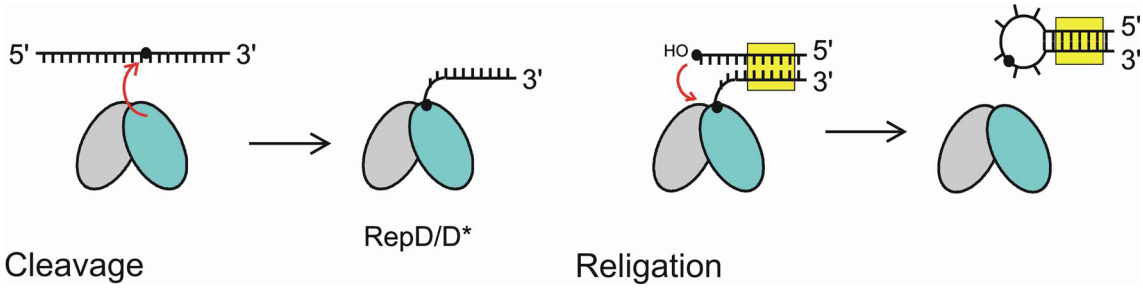


Figure 1. Cleavage and religation of oligonucleotides via the heterodimeric intermediate RepD/D*. The boxed (yellow) region indicates randomised sequences in the oligonucleotide substrates.

Using randomised sequences in either the acceptor substrate, or within both acceptor and the covalent adduct of RepD*, we have found that a significant proportion of recovered, religated products display base complementarity despite being randomised to begin with (Table 1).

Table 1. Sample sequences obtained following religation

Sequence	**Sequence** (5' → 3'; ' indicates nick)
wild-type	..AAAACCGGCTACTCT'AATAGCCGGTTAA..
I-T1	..AAACCGGTCTACTCT'AATAGCCGGTTAA..
I-U1	..AACCGGCGCTACTCT'AATAGCCGGTTAA..
II-T2	..AAGGCCGGCTACTCT'AATAGCCGCTGAA..
II-R	..AAGCTCAGCTACTCT'AAATAGCTGAGTC..

Bases initially randomised in the substrate are underlined

Complementarity between a randomised acceptor substrate and the wild-type RepD/D* intermediate could be explained by DNA sequence specificity inherent within RepD. However, the recovery of products in which the complementary sequences are altered in both acceptor substrate *and* RepD/D* intermediate argues

strongly in favour of a DNA-mediated base pairing between the DNA adduct of RepD/D* and the acceptor substrate, as depicted in Figure 1.

The RepD/D* complex may thus be using base pairing with the covalent DNA adduct in conjunction with the conserved nick site to identify the target for termination of replication. To our knowledge, this is the first example of a covalent DNA adduct being utilised by a protein for discrimination of sequence in such a way. Because base paring is involved, the product of the reaction will possess an inverted complementary repeat, even though one is not required for the initial cleavage reaction to occur. Over many generations, the extent of the base-paired region may reach an optimum length, as found *in vivo*. The SELEX approaches demonstrated in this study are currently being redesigned to investigate this.

Acknowledgements
We thank Val Sergeant for technical support, and Denise Ashworth for DNA sequence analysis.

Funding
This work is supported by the Wellcome Trust.

Crystallographic studies of a type IV topoisomerase from *Staphylococcus aureus*

Stephen Carr, George Makris, Simon E. V. Phillips and Chris D. Thomas

Background

DNA topoisomerases are ubiquitous enzymes responsible for resolving topological problems arising during DNA transcription, recombination, replication and chromosome partitioning. Topoisomerase IV (topo IV) of *Staphylococcus aureus* is a heterotetrameric protein composed of two homodimeric subunits: GrlA, which is responsible for DNA binding, cleavage of both strands (type II class enzyme) and the religation, and GrlB, which hydrolyses ATP allowing enzyme turnover. The action of topo IV results in a reduction in the superhelical density.

Previously, GrlA (90 kDa) had been purified and shown to be functional *in vitro*. With a view to identifying domains with discrete functions within GrlA partial proteolysis was performed and a stable 56 kDa fragment was generated (GrlA56). This was shown to retain DNA nicking activity, but was unable to relegate DNA.

Recent findings

GrlA56 was subjected to large-scale crystallographic screening using a "Douglas instruments" Oryx 6 robot. Initially small clusters of plate-like crystals were observed, these were readily optimised into large, discrete plates (Fig. 1a).

Diffraction data has been collected at the ESRF (Grenoble, France) to a maximum resolution of 2.8 Å (Fig. 1b). The space group was found to be $P2_1$ and to contain one dimer per asymmetric unit.

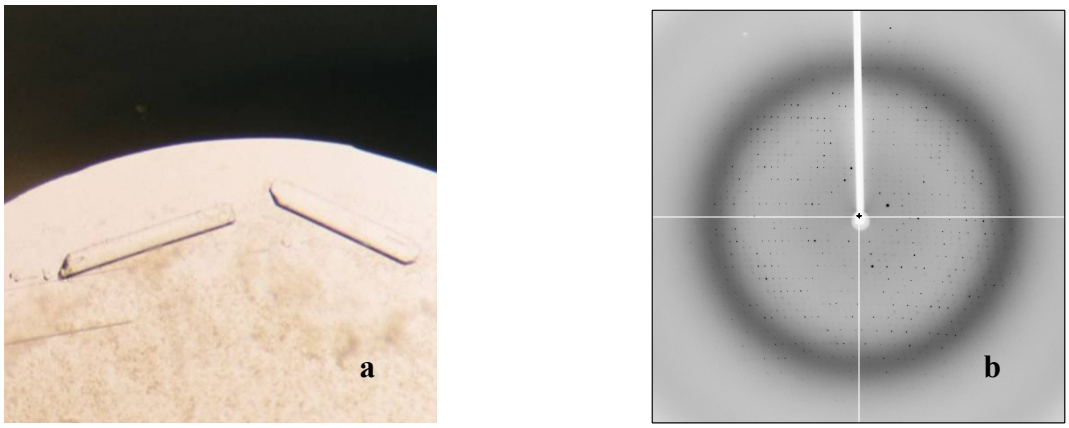


Figure 1a- Crystals of GrA56, **b**- Diffraction data collected from GrlA56 crystals

Molecular replacement using a 59 kDa fragment of a type II topoisomerase from *E. coli* resulted in interpretable electron density, from which 50% of the GrlA56 has been built. Platinyl and selenomethionyl derivatives GrlA56 have been produced with a view to generating additional phase information to complete the structure solution.

Acknowledgements and funding

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

UVRR studies of fast protein folding initiated by microsecond mixing

Iñigo Rodriguez, Sergiu Masca, Clare Friel, Sheena Radford and Alastair Smith

Introduction

The question of how a polypeptide chain can spontaneously fold into a highly ordered three-dimensional structure is of fundamental importance to many areas of biochemistry, molecular biology and biotechnology. The structural and kinetic information of the early events occurring on the sub-millisecond time scale in the folding of a polypeptide along its folding pathway can provide valuable information key to understand how a protein folds. Ultraviolet resonance Raman spectroscopy (UVRR) is a powerful technique to study the aromatic aminoacids tryptophan and tyrosine. Structural and conformational information of these important residues is provided by UVRR: the degree of hydrogen bonding of both tyrosine and tryptophan; the relative orientation of the tryptophan side-chain; and additionally, the degree of solvent exposure of the tryptohan and tyrosine residues can be obtained. Thus the local environment of tryptophan and tyrosine residues can be monitored in detail during the early stages of folding when combined UVRR spectroscopy and ultra-rapid mixing technology are used.

Integration of the microsecond mixing and the UVRR system

The ultra-rapid mixing apparatus we have developed is detailed in Fig. 1. The instrument has a 'T' shape mixer design built within a 45 mm quartz cell. It consists of a 200 μm diameter channel that joints perpendicularly with the 200 x 200 μm observational channel. A stainless steel wire introduced in the 200 x 200 μm inlet channel reduces the cross-section of the channel in order to obtain high flow velocities in the mixing region and confines the mixing chamber to a minimum. Syringe movements are controlled by a PC controlled stepper motor driver capable of working at different linear motor speeds. The mixing cell is mounted beneath the objective of the UVRR system on a precision translation stage. The progress of the mixing reaction is followed downstream from the mixer along the flow direction, which is translated into time according the flow rate and dimensions of the observational channel.

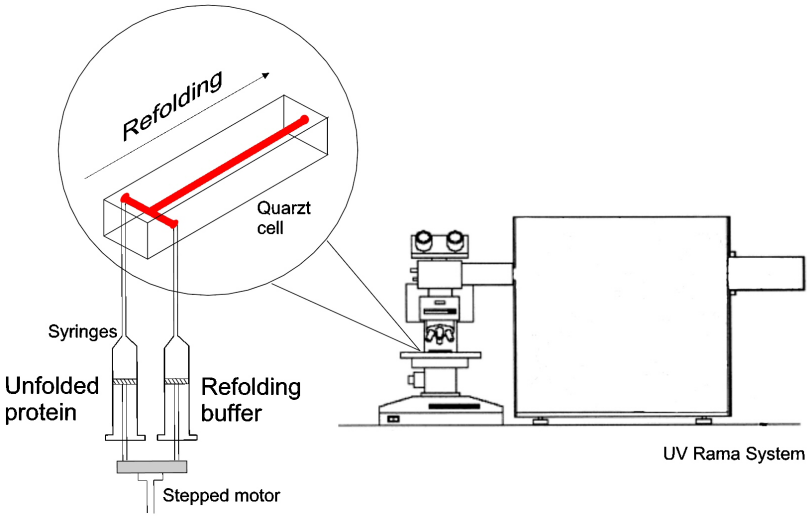


Figure 1: Continuous-flow mixing apparatus and the UVRR system.

UVRR spectra are obtained with a commercial Raman microscope, a Renishaw RM 1000, adapted for operation with deep UV laser wavelengths (229 and 244 nm laser excitation provided by an intra-cavity frequency-doubled argon ion laser). The laser beam is directed via steering mirrors onto the first Rayleigh rejection filter, this reflects the laser light into the microscope and down on to the mixing cell. Backscattered light from the sample is collected and collimated by the objective, directed through the Rayleigh rejection filter which transmits most of the Raman scattered light on to the diffraction grating before being redirected by a prism to the active surface of the CCD camera.

The time resolution of a fast mixing device is determined by the instrumental dead time, which depends critically on the time required to achieve complete mixing of the two reagents, the flow velocity, and the volume between the mixing region and the point of observation. The mixing efficiency was determined with the Raman spectral change caused by protonation of imidazole (Im) to imidazolium (ImH^+). The Raman bands of Im at 1260 and 1326 cm^{-1} are definitely distinguished from those of imidazolium at 1216 and 1457 cm^{-1}. When Im is mixed with citric acid, the UVRR bands of ImH^+ appears at 1457 and 1216 cm^{-1}, and simultaneously, the two strongest Im bands at 1260 and 1326 cm^{-1} became weaker. The kinetic spectra plotted in Fig. 2 show the mixing dead times achieved mixing equal volumes of Im and citric acid at 0.4 and 0.5 ml/s flow rate (Figs. 2A and B respectively), and the dead times observed mixing 1/10 volumes of Im and citric acid at 0.5 and 0.7 ml/s (Figs. 2C and D).

UVRR indicates that the environment of Trp75 is non-native in partially folded variants of Im7* at equilibrium.

In order to study the conformational properties of the on-pathway folding intermediate of Im7* in more detail and, in particular, the origins of the non-native interactions that stabilise this species and give rise to its unusual and characteristic hyper-fluorescence, a series of Im7* variants that closely resemble the kinetic intermediate were designed. The UVRR data presented in Fig. 3 cast new light on the structural properties of the partially folded states of Im7*, particularly in the environment of the single tryptophan residue that gives rise to the unusual fluorescence properties of these species. The UVRR data clearly indicate that the underlying structural changes induced by the mutations in the Im7* variants, L53AI54A, H3G6 and YY, result in an increase in the hydrophobicity of the environment local to Trp75 compared with wild-type Im7*. The increase in fluorescence emission intensity reported for these variants compared with the native state of wild-type Im7* is consistent with these data. Our current time resolved fluorescence measurements together with previous fluorescence quenching data indicate a decrease in the overall solvent accessibility of Trp75 in the variants L53AI54A and YY, whilst the accessibility of Trp75 in the variant H3G6 remains close to that of wild-type Im7*.

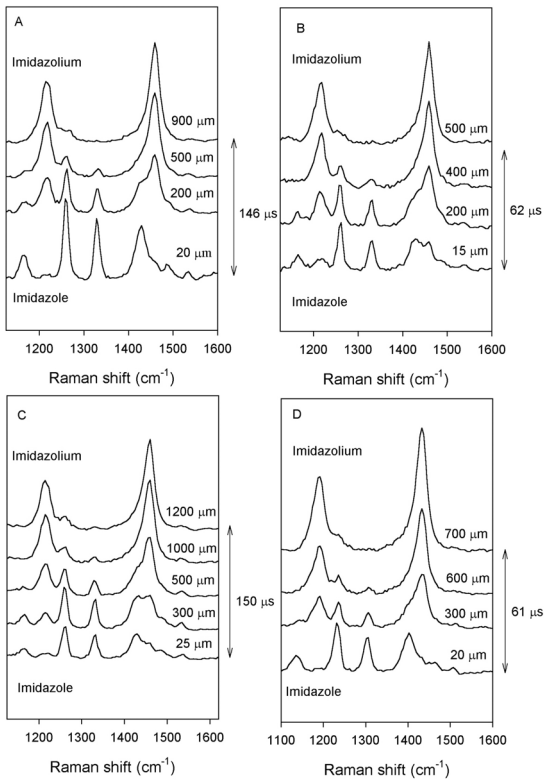


Fig. 2: 229 nm UVRR spectra of the mixing reaction of imidazole 100 mM with citric acid 100 mM. (A) and (B) correspond to equal volumes of reactants at 0.4 and 0.5 ml/s flow rate (C) and (D) are the mixing spectra obtained with 1/10 dilution of Im in citric acid with 0.5 and 0.7 ml/s . Dead times are calculated from the distance between the appearance of imidazole and imidazolium bands and the mean flow velocity of the solution in the observable channel.

UVRR is known to be a sensitive probe of the solvent accessible surface area of tyrosine and tryptophan residues in proteins - the lower the area exposed, the greater the Raman band intensities are expected to be. The UVRR data indicate an increased hydrophobicity local to the indole ring in all of these variants. The UVRR studies presented here, therefore, indicate that a newly formed hydrophobic environment is a shared property of all of the variants of Im7* studied and reveal that subtle changes local to the environment of these residues can lead to significant changes in their fluorescence properties. These data substantiate the results of ϕ-value analysis by providing direct evidence that the environment local to Trp75 in partially folded Im7* is both non-native and hydrophobic in nature.

Future Work

We will now examine the evolution of the intermediate populated transiently during folding of wild-type Im7* combining UVRR spectroscopy with ultra-rapid mixing. Structural and environmental properties concerning the side chains of the aromatic residues tryptophan and tyrosine will be obtained from this intermediate and compared with our present UVRR data at equilibrium.

Publications

Rodriguez-Mendieta, I. R., Spence, G. R., Gell, C., Radford, S. E. and Smith, D. A. (2005) Ultraviolet resonance Raman studies reveal the environment of tryptophan and tyrosine residues in the native and partially folded states of the E Colicin-binding immunity protein Im7**.** *Biochemistry.* In press.

Funding

We acknowledge the BBSRC, the Wellcome Trust and the University of Leeds for financial support.

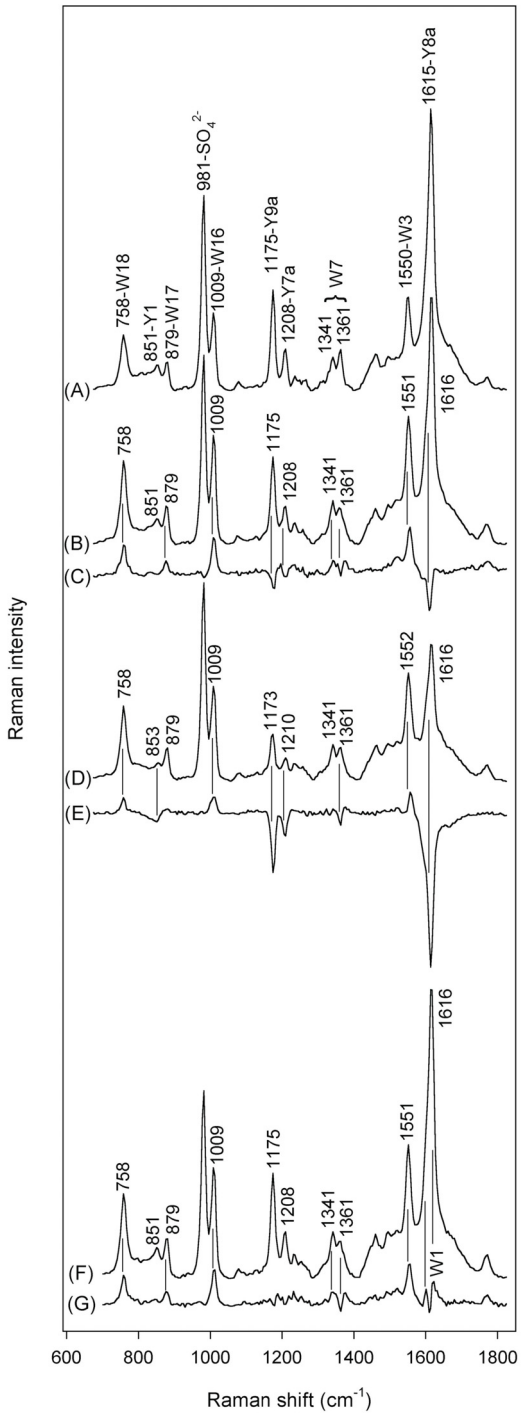


Fig. 3: 229 nm excited UVRR spectra of Im7* (A); L53AI54A and the difference spectrum L53AI54A – Im7* (B and C); H3G6 and the difference spectrum H3G6 – Im7* (D and E), and YY and the difference spectrum YY – Im7* (F and G).

The dynamic properties of single biomolecules

Masaru Kawakami, Katherine Byrne, Bhavin Khatri, David Brockwell, David Sadler, Sheena Radford, Tom McLeish and Alastair Smith

Measuring the thermal noise response of single molecules

The Brownian motion of single polysaccharide molecules under tension has been measured using AFM and used to extract a measure of the energy dissipation of the motion (see Fig. 1). This has contributions from work done against the solvent, and internal work done (against bond bending for example), as well as the molecular stiffness. The internal friction of single molecules is sensitive to conformational change and has hitherto not been accessible by experiment.

When single cellulose molecules are extended, the chain uncoils following the predictions of the freely jointed chain model, and the internal friction of cellulose is found to rise steadily with applied force. However, in the case of dextran, a plateau is observed in the force extension curve, due to a force-induced conformational change in the pyranose ring, which flips from the chair to the more extended boat conformation. This is reflected by a minimum in the internal friction of the molecule. There is also a minimum in chain stiffness due to the extra length released by the conformational change. The minimum in internal friction reflects the increased ease with which pyranose rings can change conformation. It is hoped that internal friction measurements will reveal previously unseen conformational changes in more complex molecules such as proteins during folding.

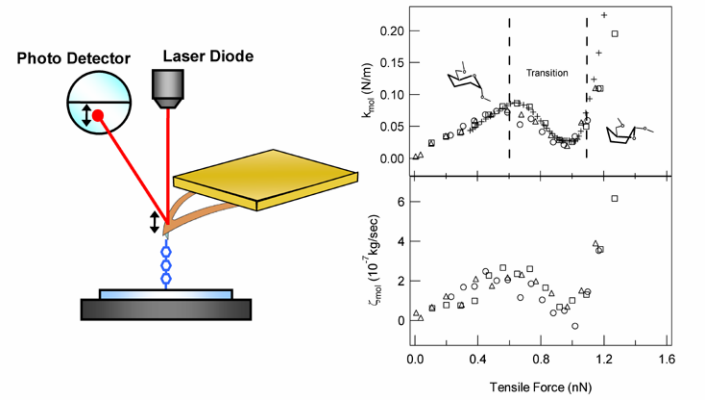


Fig. 1: Measuring the dynamic mechanical properties of single molecules with an AFM cantilever. The variation in single molecule dextran elastic constant and internal friction with applied force.

Tethering protein concatamers

To apply the method to single protein concatamers, we require a durable means of tethering the molecules. Single protein molecules can be suspended between a cantilever and substrate using non-specific tethering, but the tether breaks before sufficient data can be captured to give an accurate noise spectrum. Tethering molecules through covalent bonds should enable longer duration experiments to be carried out with ease, and has the added advantage that the exact location of the tether point is known. A protocol to tether a modified 5-domain concatamer of I27 between a gold coated cantilever and gold substrate, via distinct covalent links is being developed. The gold-binding functionality of the C-terminal cysteine residue, and the ability of N-terminal histidine residues to bind succinamide will be exploited.

Publication

Kawakami, M., Byrne, K., Khatri, B., Mcleish, T.C.B., Radford, S.E. and Smith, D.A. (2004) Viscoelastic properties of single polysaccharide molecules determined by analysis of thermally driven oscillations of an atomic force microscope cantilever. *Langmiur.* **20**, 9299-9303

Funding: This work was funded by the EPSRC.

Mechanically unfolding the small, topologically simple protein L.

Eleanore Hann, Emanuele Paci, David Sadler, Dan West, Sheena Radford,
Alastair Smith and David Brockwell

Introduction

Mechanical force is ubiquitous in biology. The role that some proteins or their complexes play in resisting or reacting to a mechanical extension on the macroscopic scale is well understood. Some proteins resist or respond to force at the sub-cellular level, providing scaffolds or acting as force sensitive triggers. Mechanically unfolding proteins using the AFM now allows the mechanical properties of these proteins to be characterised at the single molecule level. Our current understanding of the unfolding process suggests that the geometry of extension and protein topology are important factors which define the mechanical resistance of proteins. The greatest mechanical strength is found in proteins that have parallel and directly hydrogen bonded β-strands at their termini. However, such proteins display a broad range of unfolding forces under similar extension rates that are difficult to rationalize.

Protein L is mechanically resistant

In order to quantify the factors which modulate protein mechanical resistance, it is necessary to perform a systematic study of a large number of proteins with different folds. Most mechanical studies to date have focused on immunoglobulin and fibronectin type domains. To further test the hypothesis that the extension of parallel, directly hydrogen bonded terminal β-strands correlates with high unfolding forces and to identify any other features which endow mechanical resistance, we commenced a study of protein L, a small 62 amino-acid protein with a simple α+β topology.

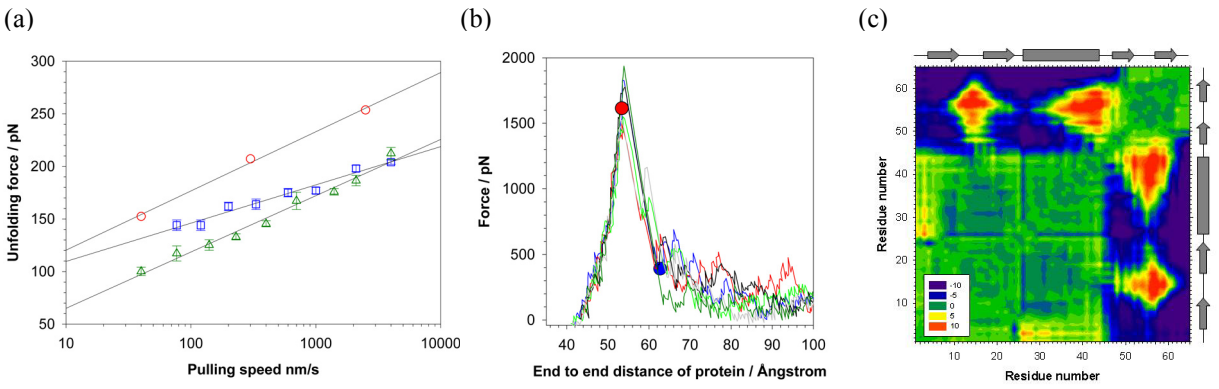


Figure 1: (a) Pulling speed dependence of the unfolding force for $(pL)_5$ (green triangles), $(I27^*)_5$ (blue squares) and $(ubiquitin)_9$ (red circles). Data for $(I27^*)_5$ and $(ubiquitin)_9$ taken from Brockwell *et al*. (2002) *Biophys J*. **83**:458-472 and Carrion-Vazquez *et al*. (2003) *Nature Struct. Biol*. **10**:738-743, respectively. (b) Constant velocity molecular dynamics simulations of the forced unfolding of protein L at 4×10^9 nm/s. The similarity of each trajectory (different coloured lines) suggests that protein L unfolds *via* a small number of routes through the unfolding energy landscape. (c) Contour plot showing the difference in distance between every pair of amino-acids in protein L at an extension of 1.6 Å before and 1.6 Å after the mechanical unfolding event (red and blue filled circles in (b), respectively). Residue numbers (left hand side and bottom) are shown opposite cartoons (right hand side and top) depicting the type of secondary structure element that each residue occupies in the native state (rectangle: α-helix, arrow β-strand). Pairs of residues that move further apart from each other during unfolding are coloured purple to green (-10 to 0Å), those that become closer to one another are shown in green to red (0 to 10 Å).

In accord with the above hypothesis, analysis of the mechanical unfolding parameters of a pentameric homopolymer of protein L $(pL)_5$ showed that the protein is significantly resistant

to extension at all speeds tested (Fig. 1(a)). This finding provides strong evidence that mechanical strength is determined predominantly by topology and not by evolved function, in agreement with earlier experiments on non-force bearing proteins such as the E2lip3 domain from *E.coli* studied previously in this laboratory. The gradient of a force *versus* the logarithm of the pulling speed plot (Fig. 1(a)) gives information on the distance between the native and transition states. Comparing the force *versus* log pulling speed plots for $(pL)_5$ and $(I27^*)_5$ shows that the unfolding transition state for these proteins occur at different extensions from their native states, as expected for proteins with completely different folds. Interestingly, ubiquitin, a protein within the same fold family as protein L, shows an identical pulling speed dependency, but unfolds at a significantly higher force, even though this protein has fewer hydrogen bonds between its N- and C-terminal strands.

Simulations reveal a simple unfolding mechanism for protein L

Molecular dynamics simulations can be used to gain insight into the structural origin of the mechanical resistance of proteins in atomistic detail. The force-extension profiles of replicate trajectories of protein L are highly reproducible suggesting that this protein unfolds *via* an unusually well defined transition state that is represented by a narrow structural ensemble (Fig. 1(b)). A distance difference map was calculated to highlight the contacts which are broken when the protein traverses the unfolding transition state barrier. The resulting diagram (Fig. 1(c)) is striking, showing that protein L unfolds by the shearing apart of two distinct structural units. This mechanism is similar to that proposed for ubiquitin. Importantly, the number of long-range contacts that span the two unfolding units in protein L is both significantly smaller (22 and 38, respectively) and in fewer clusters than those for ubiquitin. Thus, although each protein has to be extended to a similar extent to reach the transition state to unfolding, a significantly greater force may be required for ubiquitin to reach the transition point as this protein shows greater co-operatvity across the surfaces which are to be sheared.

These data suggest that the mechanism of mechanical unfolding is conserved in proteins within the same fold family and demonstrate that whilst the topology and presence of a hydrogen bonded clamp are of central importance in determining mechanical strength, hydrophobic interactions also play an important role in modulating the mechanical resistance of these similar proteins. Further experiments using protein L, its homologues and other proteins with related folds, combined with site-directed mutagenesis studies, are now underway to determine and quantify the balance of these effects in determining the mechanical stability of proteins.

Publications

Brockwell, D.J., Beddard, G.S., Paci, E., West, D.K., Olmsted, P.D., Smith, D.A. and Radford, S.E. Mechanically unfolding the small, topologically simple protein L. Submitted *Biophys. J.* (2005).

Collaborators

Godfrey Beddard, School of Chemistry, University of Leeds
Peter Olmsted, School of Physics and Astronomy, University of Leeds

Acknowledgements

We thank Keith Ainley for technical support and the BBSRC, EPSRC and the Wellcome Trust for funding. David Brockwell is an EPSRC funded White Rose Doctoral Training Centre lecturer and Sheena Radford is a BBSRC Professorial Fellow.

Machine learning to predict gene and protein function

Andrew Garrow, James Bradford, Matthew Care and David Westhead

Introduction

Machine learning techniques are being applied to several biological problems, in collaboration with groups in computer science and statistics. Projects employ a variety of learning methods including support vector machines, decision trees and Bayesian networks, and the applications range through protein structure prediction, the prediction of gene function and the effects of mutations, and the prediction of protein interactions. Following our earlier work in these areas, this year has seen the start of a major new effort in Bayesian network learning, which will provide a new avenue of attack to predict protein interactions and the effects of mutations, and open the new problem of predicting the relatedness of gene function from '-omics' data using the Gene Ontology.

Protein function prediction and classification using uncertainty

The aim of this project is to investigate the use of Bayesian networks to integrate information, express relationships and make inferences or predictions on biological problems, motivated by data generation in genomics and proteomics. This aim will be met through the following objectives:

- to demonstrate the use of Bayesian networks for prediction of functional effects of single nucleotide polymorphisms (SNPs) and other mutations;
- to construct a Bayesian network for prediction of protein-protein interaction interfaces;
- to research the implementation of an existing classification ontology such as the Gene Ontology (GO) as a Bayesian network to handle uncertain data and relate functional categories;
- and, in all cases, to compare Bayesian networks with other methods, including support vector machines (SVMs), decision trees and standard neural networks in terms of prediction performance and usability issues.
-

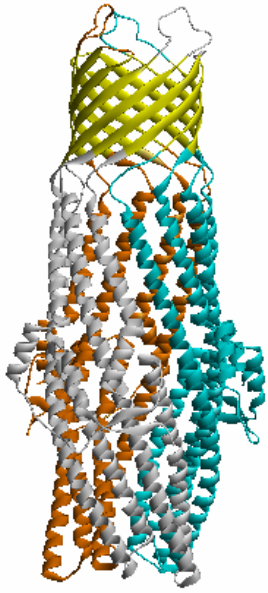
The most advanced aspect of this project currently is the prediction of the effects of non-synonymous single nucleotide polymorphisms (nsSNPs). This has been studied by various research groups using a variety of probabilistic and machine learning tools. All of the methods use a range of structural and sequence attributes to try and predict deleterious mutations, those affecting protein function.

This project aims to use Bayesian networks for SNP prediction problems, using a diverse set of attributes to predict the effects of mutations. Bayesian networks have several advantages over other machine learning techniques. Firstly the structure of the Bayesian network is a pictorial representation of the assumptions made for the prediction process, showing the causal links between the attributes and the possible effect. In addition Bayesian networks can function with reasonable accuracy even if there are missing attributes in the sample for which a prediction is to be made. Thus they can be trained using both protein structural and sequence attributes without any punishment for proteins for which structural information has not been derived.

An additional project goal is to use the final Bayesian networks to produce a SNP prediction web server. This would allow SNP queries to be submitted to a server, with the predictions and derived attributes returned by email. This would then be followed up by a

comprehensive SNP study using a variety of the best Bayesian network models on a dataset derived from the OMIM database.

Searching genomes for transmembrane barrel proteins



Transmembrane barrel (TMB) proteins are a functionally important and diverse group of molecules found spanning the outer membranes of Gram negative and acid fast Gram positive bacteria, mitochondria and chloroplasts. Structurally they are well understood with entries from over 23 families in the protein databank (PDB). However, unlike with alpha-helical transmembrane proteins, development of TMB computational screening techniques has proven difficult with TM strands composed of a short and aliphatic, inside-outside dyad repeat motif.

In this project high accuracy composition based discrimination algorithms are being developed using a number of machine learning techniques (e.g. support vector machines (SVMs) and genetic algorithms). Another related project is focusing on development of Hidden Markov Models for detection of transmembrane strands.

Collaborations

Drs. Andy Bulpitt and Chris Needham in the School of Computing, University of Leeds.
Dr Alison Agnew in the School of Biology, University of Leeds

Publications

Bradford J.R. & Westhead D.R. (2005) Improved prediction of protein-protein binding sites using support vector machines. *Bioinformatics* (in press).

Bradford J.R. & Westhead D.R. (2003) Asymmetric mutation rates at enzyme-inhibitor interfaces: Implications for the protein-protein docking problem. *Protein Science* **12**: 2099-2103.

Krishnan, V.G., Westhead, D.R. (2003) A comparative study of machine learning methods to predict the effects of Single Nucleotide Polymorphisms on protein function. *Bioinformatics* **19**: 2199-2209.

Siepen, J.A., Radford, S.E. and Westhead, D.R. (2003) β edge strands in protein structure prediction and aggregation. *Protein Sci* **12**: 2348-2359.

Funding: This work is funded by the MRC, BBSRC, and the BBSRC E-Science Initiative

Reconstruction and analysis of biochemical networks

Chris Hyland, Liz Gaskell, John Pinney, Glenn McConkey and David Westhead

Introduction

The rapid proliferation of genome sequencing projects over the last ten years has resulted in an exponential growth in the amount of genomic DNA 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 have developed 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 (*metaSHARK*) 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. Metabolites 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. The confidence of gene predictions can be improved by the incorporation of other forms of genomic data, such as gene expression data, to show whether a predicted gene is expressed under a particular condition or at a particular time point in an organism's life cycle. Pathways in the network and elementary modes can then be ranked according to their biological relevance based on the combined expression levels of each gene in the pathway. This data can also be used to produce a list of candidate genes for enzymatic functions that appear to be missing from a particular pathway.

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. Analysis of the recently-released complete parasite genome is revealing novel genes and pathways which are being verified by RNAi experiments in Dr McConkey's group. We hope

that this work will ultimately lead to the identification of new drug targets for this killer disease.

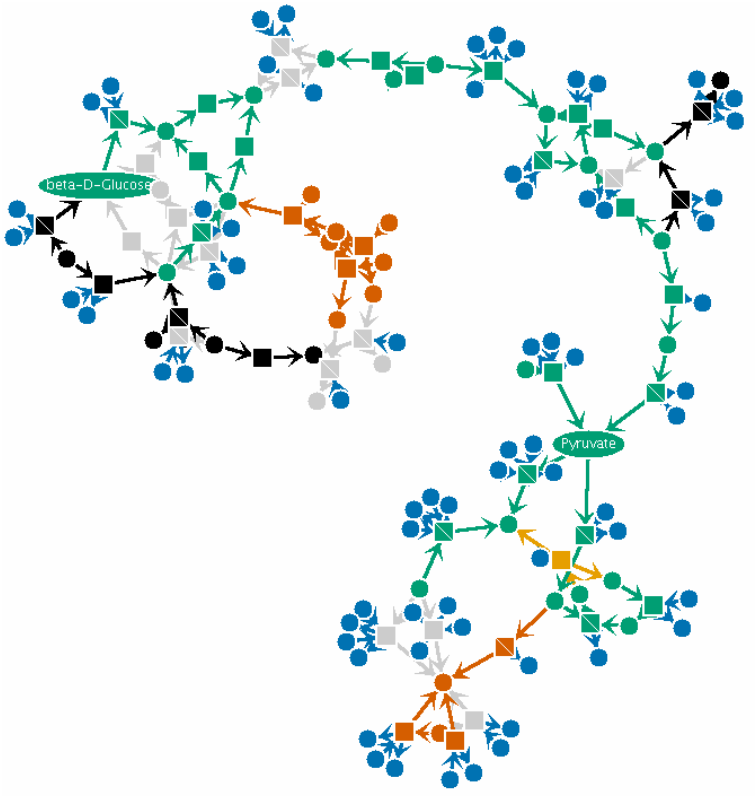


Fig. 1: Part of an automated metabolic reconstruction of the avian intracellular parasite *Eimeria tenella*, visualised as a Petri net. Metabolites are represented as circles; reactions as squares. Directed arcs between nodes show the effect of a reaction as the consumption and production of different metabolites. Blue circles represent ubiquitous ("pool") metabolites, such as water, ATP, NADH etc.. Nodes in green show reactions catalysed by enzymes for which good evidence has been found in the *E. tenella* genome. Nodes in red show reactions catalysed by enzymes for which only tentative evidence has been found. Grey nodes show that no evidence for the catalysing enzymes was found, whilst black nodes indicate that no model data was available in our database, hence no gene search could be performed. The reconstruction clearly shows the presence of a glycolytic pathway in this organism as a route in green from beta-D-glucose to pyruvate. The Petri net structure of the reconstructed network will aid in the development of automated pathway detection algorithms using elementary mode theory, as well as other forms of network analysis.

Publications

Pinney, J.W., Shirley, M.W., McConkey, G.A. and Westhead, D.R. (2005) metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Res.* (in press).

McConkey, G.A., Pinney, J.W., Westhead, D.R., Plueckhahn, K., Fitzpatrick, T.B., Macheroux, P. and Kappes, B. (2004) Annotating the *Plasmodium* genome and the enigma of the shikimate pathway. *Trends in Parasitology* **20**:60-65

Pinney, J.W., Westhead, D.R. and McConkey, G.A. (2003) Petri Net representations in systems biology. *Biochem. Soc. Trans.* **31**:1513-1515

Funding

We thank the MRC and BBSRC for funding this project.

Storage and analysis of microarray data

Chih-hung Jen, Ioannis Michalopoulos, Archana Sharma-Oates, Iain Manfield, Phil Gilmartin, Noel Buckley, Phil Quirke and David Westhead

Introduction

The group has a growing interest in data from post genomic research, including microarray based measurements of gene expression, and, more recently, tissue microarrays. Work is collaborative with local experimental groups who generate data, and we are responsible for three aspects of these projects: (i) appropriate storage and archiving of data according to international standards, and efforts to advance these standards; (ii) data analysis using methods of multivariate statistics; and (iii) the use of private and public data to make predictions and motivate experimental verification or refutation.

Plant post genomics

Microarray analyses are being carried out to identify *in vivo* targets of plant GATA transcription factors in *Arabidopsis thaliana*. GATA transcription factors are Type IV zinc finger proteins, found in all eukaryotes. 29 *Arabidopsis thaliana* GATA genes are cloned, all sharing a $CX_2CX_{18}CX_2C$ zinc finger domain. Their mammalian orthologs show specificity of the conserved promoter element GATAAGG. *Arabidopsis thaliana* GATA-2 and GATA-4 are Phytochrome A regulated, as opposed to GATA-1.

Identifying coexpressed genes from the microarray data can be used to assign potential functions to new genes and help the discovery of transcriptional regulation networks. Currently, the coexpressed genes are usually analysed by many sophisticated clustering algorithms e.g. SOM, hierarchical clustering, k-means clustering. However, these clustering approaches usually depend on the distance cut-off value or arbitrary k value to group the genes, and these criteria do not really indicate the significance of the similarity within the clusters. Besides, they assign particular genes to only one cluster that may cause loss information where genes may have multiple biological roles or respond to different transcription factors.

In order to identify the *in vivo* potential targets of *Arabidopsis* GATA family transcription factors using microarray data and avoid the drawbacks of clustering algorithms, we propose a novel robust approach of assessing the significance of relationships in expression. We developed a new WWW-based *Arabidopsis* Co-Expression Tool (ACT) for plant gene analysis, based on large *Arabidopsis thaliana* public microarray data sets consisting of 322 Affymetrix arrays (ATH1) from 51 different experiments, obtained from the Nottingham Arabidopsis Stock Centre (NASC). The co-expression analysis tool allows users to identify genes whose expression patterns are correlated across selected experiments or the complete data set. The output is the Pearson correlation coefficient, or r-value, which is a scale-invariant measure of expression similarity, and this is accompanied by probability (p) and expect (E) values reflecting statistical significance against a background of random chance correlations. The E value is calculated as a product of the number of genes on the array and the p value. The correlation coefficient (r) is used to rank the genes in descending order of correlation with the driver. In addition to r, p and E values, the output includes Affymetrix probe ID, AGI code and current annotation for each gene. Genes with strongly correlated expression patterns are likely to be under similar transcription regulatory mechanisms, or involved in related biological processes. We illustrate the applications of the software by analysing genes encoding functionally related proteins, as well as pathways involved in plant responses to environmental stimuli. The resource is freely available at http://www.arabidopsis.leeds.ac.uk/.

Based on the r-value derived from the correlation analysis tool, we can reveal the GATA coexpressed genes with confidence. An example result of the top fifty genes coexpressed with GATA-1 is shown in Table 1 and co-correlation plot of GATA2 and GATA-4 is shown in Fig. 1.

256916_atAT3G24050GATA transcription factor 1 (AtGATA-1)

Probe Setr-value	p-value	e-value	GeneID	Annotation
250274_at 0.683184	1.4e-45	2.9e-41	AT5G13020	expressed protein
253780_at 0.649790	5.2e-40	1.1e-35	AT4G28400	protein phosphatase 2C (PP2C), putative
256853_at 0.645309	2.6e-39	5.7e-35	AT3G18640	hypothetical protein
256183_at 0.628998	7.2e-37	1.6e-32	AT1G51660	mitogen-activated protein kinase kinase (MAPKK), putative (MKK4)
255095_at 0.610592	2.8e-34	6.1e-30	AT4G08500	mitogen-activated protein kinase kinase, putative
253204_at 0.609395	4.0e-34	8.8e-30	AT4G34460	transducin / G-protein beta-subunit (AGB1)
263274_at 0.605274	1.4e-33	3.2e-29	AT2G11520	protein kinase family
258979_at 0.602097	3.8e-33	8.4e-29	AT3G09440	heat shock protein hsc70-3 (hsc70.3)
252449_at 0.595278	3.0e-32	6.5e-28	AT3G47060	FtsH protease, putative
250994_at 0.584791	6.3e-31	1.4e-26	AT5G02490	heat shock protein hsc70-2 (hsc70.2) (hsp70-2)
246292_at 0.583759	8.4e-31	1.8e-26	AT3G56860	RNA recognition motif (RRM) - containing protein
246529_at 0.582779	1.1e-30	2.4e-26	AT5G15730	serine/threonine protein kinase, putative
248195_at 0.582351	1.3e-30	2.8e-26	AT5G54110	VAMP (vesicle-associated membrane protein)-associated protein family
267341_at 0.580437	2.2e-30	4.7e-26	AT2G44200	expressed protein
252670_at 0.577386	5.1e-30	1.1e-25	AT3G44110	DnaJ protein AtJ3
250350_at 0.575414	8.8e-30	1.9e-25	AT5G12010	expressed protein
248131_at 0.572371	2.0e-29	4.5e-25	AT5G54830	expressed protein
245986_at 0.569217	4.8e-29	1.1e-24	AT5G13160	protein kinase family
267009_at 0.568045	6.6e-29	1.4e-24	AT2G39260	middle domain of eukaryotic initiation factor 4G (MIF4G) domain-containing protein
256820_at 0.566753	9.3e-29	2.0e-24	AT3G22170	far-red impaired response protein -related
247054_at 0.566549	9.9e-29	2.2e-24	AT5G66730	zinc finger protein
257484_at 0.565352	1.4e-28	3.0e-24	AT1G01650	expressed protein
261743_s_at 0.564090	1.9e-28	4.2e-24	AT1G08420	protein serine/threonine phosphatase alpha -related
249613_at 0.563312	2.3e-28	5.1e-24	AT5G37380	DnaJ protein family
252906_at 0.561615	3.7e-28	8.0e-24	AT4G39640	gamma-glutamyltransferase -related
257883_at 0.561464	3.8e-28	8.4e-24	AT3G16940	calmodulin-binding protein
246221_at 0.560417	5.0e-28	1.1e-23	AT4G37120	step II splicing factor - like protein
262408_at 0.559422	6.5e-28	1.4e-23	AT1G34750	protein phosphatase 2C (PP2C), putative
261520_at 0.559319	6.7e-28	1.5e-23	AT1G71820	SEC6 protein
266800_at 0.558941	7.4e-28	1.6e-23	AT2G22880	hypothetical protein
255605_at 0.554410	2.4e-27	5.3e-23	AT4G01090	expressed protein
252862_at 0.553816	2.8e-27	6.2e-23	AT4G39830	L-ascorbate oxidase, putative
249988_at 0.553640	2.9e-27	6.4e-23	AT5G18310	expressed protein
259202_at 0.552617	3.8e-27	8.4e-23	AT3G09100	mRNA capping enzyme, RNA guanylyltransferase -related
263457_at 0.551520	5.1e-27	1.1e-22	AT2G22300	ethylene-induced calmodulin-binding protein, putative
259341_at 0.551249	5.4e-27	1.2e-22	AT3G03740	expressed protein
251861_at 0.551161	5.6e-27	1.2e-22	AT3G54810	GATA zinc finger protein
255280_at 0.550891	5.9e-27	1.3e-22	AT4G04960	receptor lectin kinase, putative
256185_at 0.549733	8.0e-27	1.7e-22	AT1G51700	Dof zinc finger protein
248268_at 0.547811	1.3e-26	2.8e-22	AT5G53480	importin beta, putative
247874_at 0.547468	1.4e-26	3.1e-22	AT5G57710	101 kDa heat shock protein; HSP101-related protein
262054_s_at 0.547074	1.6e-26	3.4e-22	AT1G79920	heat shock protein hsp70, putative
251683_at 0.546546	1.8e-26	3.9e-22	AT3G57120	protein kinase family
249730_at 0.546272	1.9e-26	4.2e-22	AT5G24430	calcium-dependent protein kinase, putative (CDPK)
247811_at 0.545759	2.2e-26	4.8e-22	AT5G58430	leucine zipper-containing protein
248698_at 0.545115	2.6e-26	5.6e-22	AT5G48380	leucine rich repeat protein family
264436_at 0.544316	3.1e-26	6.8e-22	AT1G10370	glutathione transferase, putative
254432_at 0.543187	4.1e-26	9.0e-22	AT4G20830	FAD-linked oxidoreductase family
252206_at 0.541739	5.9e-26	1.3e-21	AT3G50360	caltractin (centrin)
266964_at 0.539990	9.1e-26	2.0e-21	AT2G39480	ABC transporter family protein

Table 1: Top fifty correlated genes to GATA-1 *Arabidopsis* gene.

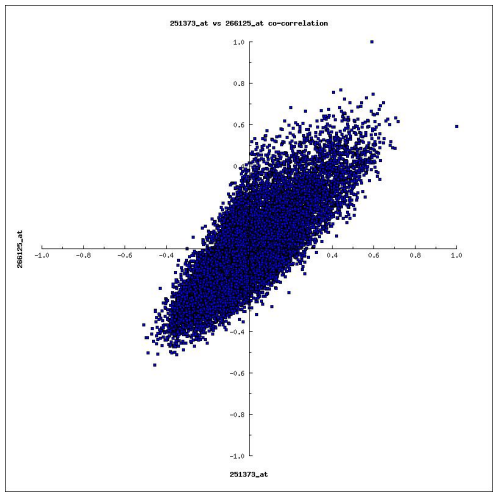


Fig. 1: Co-correlation scatter plot of GATA-2 against GATA-4.

Human cancer pathology

Human tissue samples are obtained from the individuals in a clinical trial (both cancerous and normal tissues from the same person) for biochemical and histopathological analyses.

Depending on the nature of the trial these tissue samples undergo extensive characterisation using a number of high-throughput molecular biology techniques. The high-throughput techniques most commonly used in cancer research are cDNA microarrays, comparative genomic hybridisation arrays and tissue microarrays. The purpose of the cDNA microarray approach is to gain an insight into the expression levels of all the predicted genes in the human genome with the aim of identifying a set of genes related to a clinical outcome that may be either up or down regulated in tumour verses normal tissue. Comparative genomic hybridisation (CGH-arrays) arrays are used to study chromosomal instability at a genome level within tumour verses normal tissues. TMA is a technique that enables the analysis of a large cohort of clinical specimens in a single experiment thereby studying the molecular alterations (at the DNA, RNA, or protein level) in thousands of tissue specimens in parallel. The aim of cDNA microarray and CGH-array techniques are to either identify biomarkers that can be verified by TMA.

Our involvement in this research involves analysis of the cDNA microarray and CGH-array data using statistical approaches, and the development of storage and analysis software for tissue microarray experiments, and area where we are contributing to the development of international standards, and the integration of this data with MIAME compliant microarray databases.

TMAs are used in the laboratory to assess on a large-scale the diagnostic and therapeutic significance of various genes and proteins in colorectal tumour samples. A relational database has been designed and implemented in MySQL. The information stored in the database include TMA design constructs, tissue staining protocols, the results including images scanned from digital slide scanners and the pathology reports associated with each tumour sample. Additional information includes experiment authors, dates of each experiment, quality of cores on each TMA slide and the storage location of each TMA within the laboratory. This database is interfaced with the World Wide Web (WWW) thereby enabling users to query and assimilate their own data into the database.

Collaborators
Profs. P.M.Gilmartin and N. Buckley, Dr. P. Devlin (plant project), Prof. P. Quirke (human cancer pathology).

Funding
This work is funded by the BBSRC.

Computational molecular modelling of protein structure and function

Binbin Liu, Sally Mardikian, Richard Jackson and Dave Westhead

Introduction

The application of molecular modelling to study the behaviour of proteins in cells is useful in understanding the interactions they form with small molecules (ligands), and also their functions within the cell. Molecular modelling allows us to test our current knowledge of these systems and our ability to predict their behaviour *in silico*. We have been using computational techniques to study the electrostatics of conduction in potassium channels (in collaboration with Prof. M. Boyett) and in the prediction of protein binding affinity to ligands (in collaboration with Drs. R.M. Jackson and V. Gillett).

Modelling the electrostatics of conduction in a voltage-gate potassium channel and an inward-rectifier potassium channel

K^+ channels, integral membrane proteins, are universal regulators of cellular function both in excitable and non-excitable cells. Voltage-gated K^+ (Kv) channels and inward-rectifier K^+ (Kir) channels form two important families of K^+ channels. The role of these two families of K^+ channels includes the generation of the action potential and resting potential. The electrophysiological functions of key amino acids in Kv1.4 and Kir3.1/Kir3.4 channels have been studied in site-directed mutagenesis experiments. H509 and K540, located at the extracellular mouth of the Kv1.4 channel, for example, affect slow C-type inactivation on changing the pH of the environment. In the Kir3.1/Kir3.4 channel, neutralisation of R149, E139 and D173 has profound effects on, for example, the ligand (polyamine)-binding underlying inward rectification.

We have performed structure-function evaluation and electrostatics calculations for the Kv1.4 and Kir3.1/Kir3.4 channels at the atomic level using computational biology approaches. 3D models of Kv1.4 (S5, p-loop and S6 domains only) and Kir3.1/Kir3.4 channels were constructed by a standard homology modelling procedure. The electrostatic potential profiles of the channels were calculated by applying the Finite Difference Poisson-Boltzmann equation (FDPB) to the Kv1.4 and Kir3.1/Kir3.4 channels with the residues of interest in different protonation states. In addition, the electrostatic potential profile of the Kv1.4 and Kir3.1/Kir3.4 channels with different numbers of K^+ ions within the selectivity filter region was calculated.

The electrostatic potential profile along the axis of the channels shows that the electrostatic potential is generally negative along the pore of the channels, and that the electrostatic potential is lowest (i.e. most negative) in the selectivity filter region. Therefore, the selectivity filter does provide a suitable environment for K^+ ions to cross the high energy barrier of the lipid membrane. When H509 in the Kv1.4 channel carries a positive charge (as opposed to being neutral) the electrostatic potential is increased (i.e. made more positive) by ~38 %. On neutralising K540 in Kv1.4, the electrostatic potential is decreased (i.e. made more negative) by ~45 %. It is concluded that the residues of interest could be affecting the electrophysiological properties of the channels by altering the electrostatic potential of the channel pore (the electrostatic potential is expected to influence the K^+ occupancy of the channel).

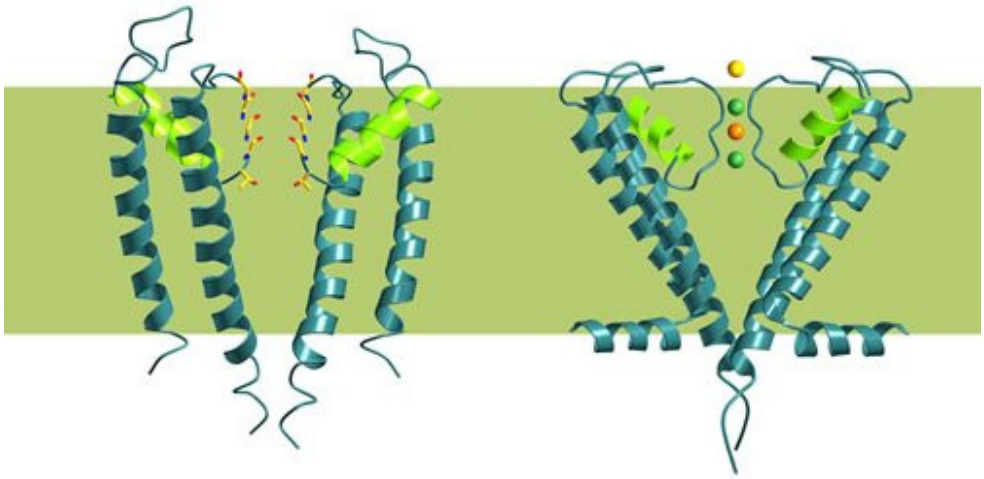


Fig. 1. Cartoon representation of transmembrane helices of potassium channel with two subunits. The selectivity filter is shown in stick (left) and four potassium ions are bound to S0 and the selectivity filter (right)

The application of multi-objective genetic algorithms to protein-ligand docking

Ligand docking is a process which computationally predicts the correct bound conformation of a given protein-ligand complex from atomic coordinates. A docking procedure employs search techniques which produce different conformations and orientations of a ligand on a target protein. A scoring function is then used to score the generated conformations. This is usually done by calculating the binding energy between ligand and protein. From the score values it is possible to infer at which confirmation the most optimal binding occurs. Docking score values are at best empirical functions showing some correction with observed binding conformations and energies. They often involve weighted combinations of possibly competing objectives, including van der Waals energy, electrostatics and desolvation. In this project we are investigating the use of multi-objective genetic algorithms in the docking problem, to investigate optimal balance between the components of the score function and its deviation between test cases, and to research the possibility of controlled consensus docking based on more than one different score function.

Collaborators

Dr.Val Gillet, University of Sheffield
Prof. Mark Boyett, University of Manchester.

Funding

Binbin Liu is funded by an Emma and Leslie Reid Scholarship from the University of Leeds and a scholarship from the University of Manchester. Sally Mardikian is funded by the Medical Research Council.

Investigation into the role of lysosomal proteolysis in Dialysis Related Amyloidosis

Isobel Morten, Sheena Radford and Eric Hewitt

Introduction

β_2-microglobulin (β_2m) is one of approximately twenty proteins that aggregate to form highly ordered amyloid fibrils *in vivo*. β_2m is a small 99 residue soluble protein, which is non-covalently bound to a membrane-integrated heavy chain, forming a major histocompatibility complex (MHC) class I molecule. MHC class I molecules are expressed on the surface of all nucleated cells and present peptide fragments, derived from intracellular proteins, to cytotoxic T lymphocytes. *In vivo*, β_2m is continuously shed from nucleated cells as part of its normal catabolic cycle, into the serum. β_2m is then transported to the proximal tubule of the kidney, where it is degraded and excreted. As a consequence of renal failure the β_2m serum concentration increases by up to 60-fold. This high β_2m serum concentration causes free β_2m to associate, forming insoluble amyloid fibrils, which typically accumulate in the musculoskeletal system. As a consequence uremic patients who have been dialysed for 10-15 years develop dialysis-related amyloidosis (DRA), a debilitating arthritic-like condition.

Hypothesis

How β_2m fibrils form *in vivo* is unknown, but *in vitro* studies have shown that incubation of β_2m at acidic pH or the removal of the N-terminal six residues (ΔN6β_2m) induces rapid fibril formation. We hypothesise that cells within the joint capsule internalise β_2m, which enters the endocytic pathway and accumulates in the lysosomes. We propose that the acidic microenvironment of lysosomes (pH 4.5), coupled with its high concentration of proteases, may stimulate fibril formation.

Internalisation of β_2m by a model cell line

HeLa cells were used as a model cell line to analyse the internalisation of human β_2m. These cells were incubated with monomeric human β_2m at the concentration typically seen in the serum of uremic patients (50μg/mL) (Fig. 1). Internalisation of β_2m was analysed using immunofluorescence microscopy, and localisation of internalised β_2m in the endosomal pathway was analysed by co-staining with antibodies specific for early endosomal antigen-1 (EEA-1) and vesicle associated membrane protein-7 (VAMP7). Internalised human monomeric β_2m accumulates in perinuclear regions and co-localises with both EEA-1 and VAMP7. These data are consistent with sorting of the internalised β_2m to lysosomes.

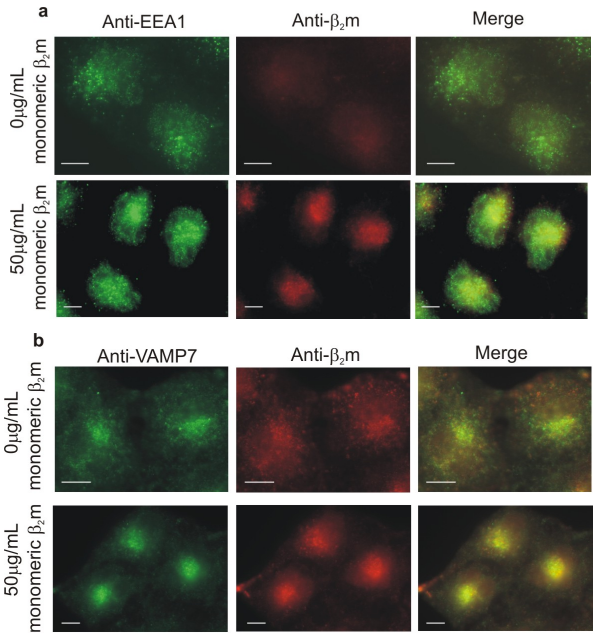


Fig. 1 – Immunofluorescence microscopy images of HeLa cells which have been incubated in the presence and absence of 50μg/mL monomeric human β_2m for 1 hour at 37°C/5%CO_2. β_2m was detected by staining with a β_2m specific antibody and its localisation determined by co-staining with antibodies to (a) EEA-1 and (b) VAMP7. 10μm scale bar.

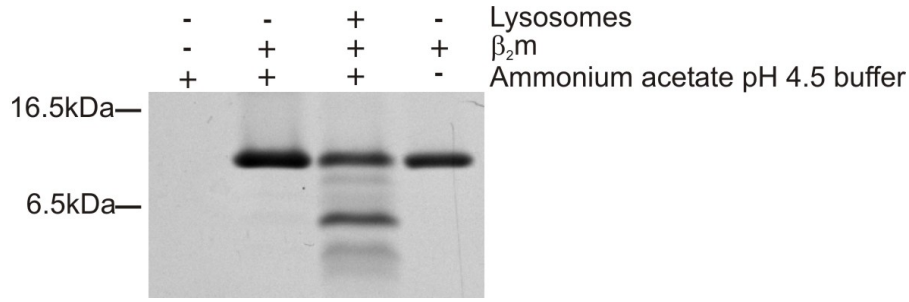


Fig. 2 – Tris-tricene gel of β_2m digested by lysosomes, isolated by subcellular fractionation from HeLa cells.

Lysosomal digestion of β_2m

To determine the effect that the lysosomal environment has on β_2m, lysosomes were isolated from HeLa cells by subcellular fractionation and these fractions used to digest monomeric human β_2m (Fig. 2). We are currently identifying the discrete bands generated and investigating whether these truncated β_2m species form amyloid fibrils both *in vivo* and *in vitro*.

Publications

Morten, I.J., Hewitt, E.W. and Radford, S.E. (2004) β_2-microglobulin and dialysis-related amyloidosis in *Protein Misfolding, Aggregation and Conformational Diseases*, Eds. Uversky, V.N. and Fink, A.L., Kluwer Academic/Plenum Publishers.

Funding

We gratefully acknowledge The National Kidney Research Fund and the BBSRC for financial support. SER is a BBSRC Professorial Research Fellow.

Biophysical characterisation of the coronavirus nucleoprotein

Kelly Spencer and Julian Hiscox

Introduction

Coronaviruses are a group of positive strand RNA viruses which can cause severe respiratory disease and gastrointestinal illnesses in both humans and animals. Principal amongst these viruses are severe acute respiratory syndrome coronavirus (SARS-CoV) and avian infectious bronchitis virus (IBV). During virus replication, a variety of proteins are synthesised including the viral RNA binding protein, nucleoprotein (N protein), a multi-functional protein with roles in both the virus life cycle and modulating host cell function. One of the key functions of N protein is to bind viral RNA and selectively incorporate this into virus particles. Using a combination of mass spectrometry coupled with surface plasmon resonance we have shown that phosphorylation of N protein plays a key role in the modulation and specificity of RNA binding. Indeed, we were the first group to fine map the positions of phosphates on the coronavirus N protein. These phosphates clustered in two groups to regions of the protein predicted to be involved in RNA binding and cell signalling. Using bioinformatic analysis (Fig. 1), in conjunction with circular dichroism and NMR we are currently investigating whether RNA binding affects the structure of N protein and to fine map the viral RNA binding domain(s).

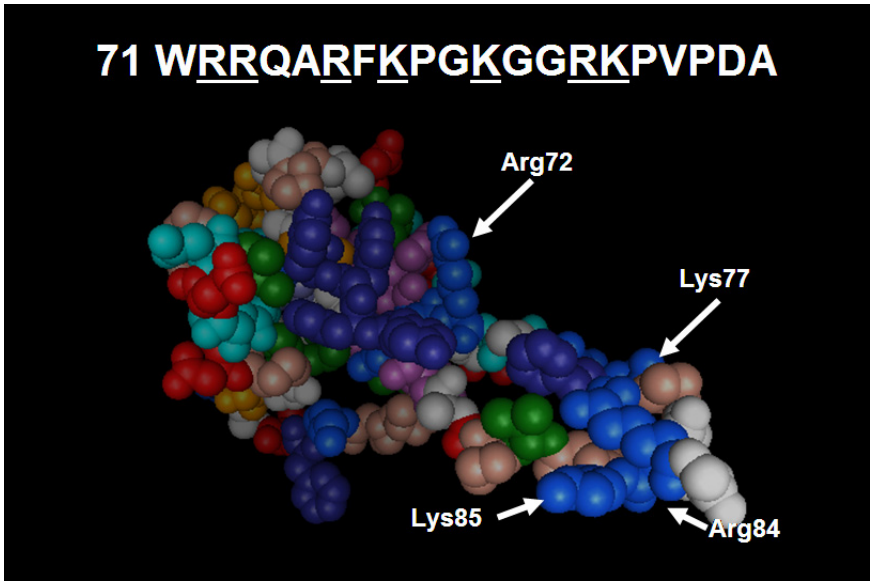


Fig. 1. Predicted structure of the N-terminal region of IBV N protein. Residues thought to be involved in RNA binding are highlighted.

Publications

Chen, H., Gill, A., Dove, B. K., Emmett, S.R., Ritchie, M.A. and. Hiscox, J.A. (2005) Mass spectroscopic characterization of the coronavirus infectious bronchitis virus nucleoprotein and elucidation of the role of phosphorylation in RNA binding using surface plasmon resonance. *Journal of Virology*. **79**, 1164-1179.

Funding

This work was funded by the BBSRC and Guildhay Ltd.

Cellular characterisation of the coronavirus nucleoprotein - delineating cell cycle control and nucleolar targeting

Brian Dove, Sally Harrison, Mark Reed, Jae-Hwan You and Julian A. Hiscox

Introduction

The nucleolus is a dynamic sub-nuclear structure involved in ribosome subunit biogenesis and cell cycle control. The specific mechanism by which proteins localise to the nucleolus and regulate the cell cycle is unknown. We have been using viral proteins as model systems to investigate the signalling involved in such pathways, specifically focusing on the coronavirus nucleoprotein (N protein). Coronaviruses are a group of positive strand RNA viruses which can cause severe respiratory disease and gastrointestinal illnesses in both humans and animals. Principal amongst these viruses are severe acute respiratory syndrome coronavirus (SARS-CoV) and avian infectious bronchitis virus (IBV). During virus replication a variety of proteins are synthesised including the viral RNA binding protein, N protein, a multi-functional protein with roles in both the virus life cycle and modulating host cell function. Using a combination of deletion and site specific mutagenesis we have identified potential nuclear localisation and nucleolar retention signals in the coronavirus N protein (Fig. 1). Our data also indicates that the cell cycle is perturbed in virus infected cells and that p53 is redistributed from the nucleus/nucleolus to the cytoplasm. This underlines current thinking that the nucleolus acts as a cellular stress sensor through the action of p53.

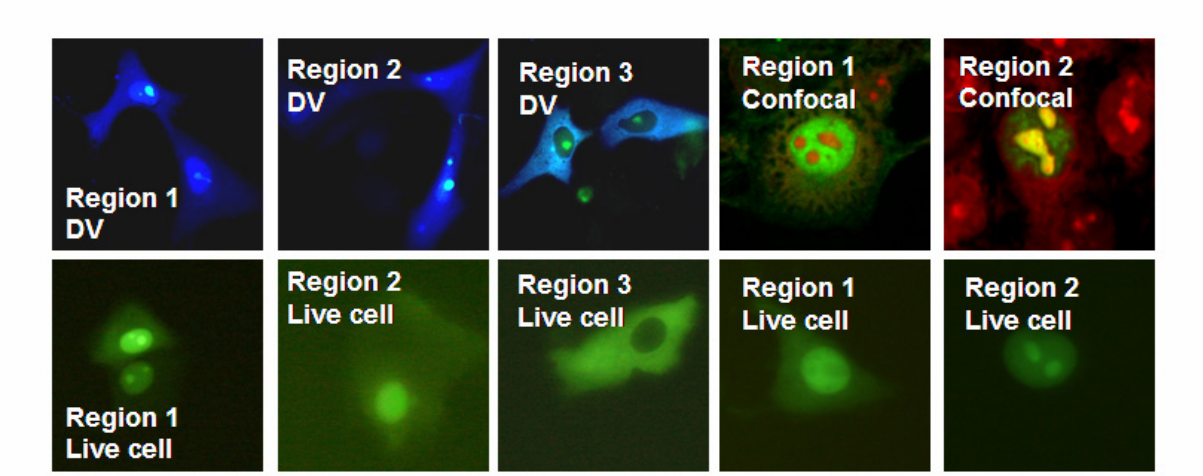


Fig. 1. Both confocal imaging (top row) and live cell microscopy are used to study the sub-cellular localisation of N protein. In this case N protein is either tagged with ECFP or EGFP. Confocal microscopy is particularly powerful as the localisation of multiple tagged proteins can be studied simultaneously.

Publications

Emmett, S.R., Dove, B.K., Mahoney, L., Wurm, T. and Hiscox, J.A. (2004) The cell cycle and virus infection. *Methods in Molecular Biology*. **296,** 197-218.

Funding

This work was funded by the BBSRC and Intervet.

Proteomics and allied technologies

Jeff Keen

Introduction

Detailed characterisation of proteins is a critical component of an active research environment in structural biology, providing information on the integrity of proteins of interest. Such work requires appropriate instrumentation and Leeds has acquired the essential equipment and expertise to enable thorough 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). The main emphasis of the laboratory has moved towards proteomic analysis of biological systems over the last couple of years and it is currently engaged in a number of collaborative projects.

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 the 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 the components of protein complexes involved in biochemical pathways or signal transduction.

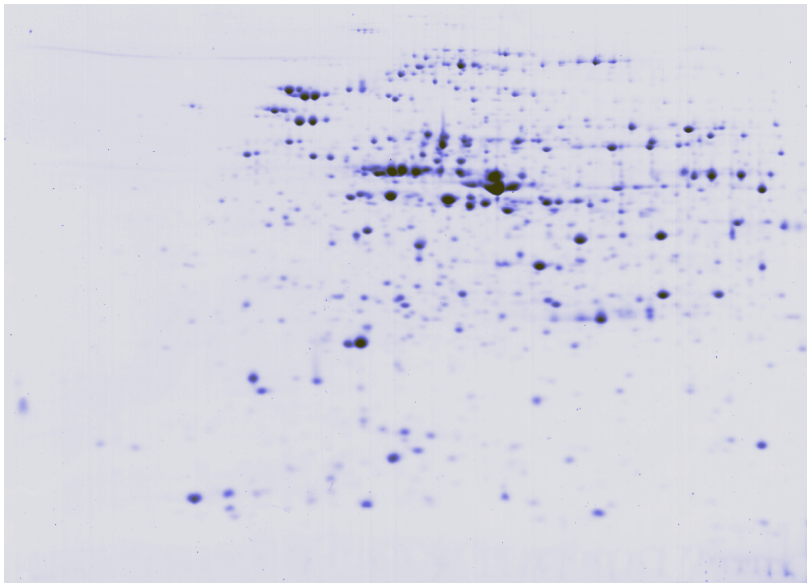


Fig. 1. 2-Dimensional electrophoresis of *E. coli* extract.

The core 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, generating an expression profile for each condition under investigation that can be used to select proteins displaying differential expression. Proteins of interest can then be analysed independently to provide their identification. This usually involves digestion of the protein and analysis of the peptide mixture 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.

Following an award in 2000 from HEFCE (through the JREI scheme), a comprehensive Proteomics Facility has been 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) are separated by high-resolution 2DE, the protein patterns visualised through staining of the gels and digital images acquired. Sophisticated software is then 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 are selected for identification. Robotic equipment is 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 is used to screen these digests, producing lists of peptide masses that are used to search databases to identify the parent protein. Digests that do not provide a protein identity can be put forward for Q-Tof™-MS/MS analysis 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 analysis of human plasma for clinical investigations (Grant and Carter), investigation of the secretory response in mammalian mast cells (Findlay), biomarker discovery for oesophageal adenocarcinoma (Wild and Hardie), investigation of the response of *Burkholderia cepacia* to challenge with antimicrobial peptides (Devine), identification of bacterial enzymes involved in bioremediation (Knapp), identification of proteins involved in microbial biofilm formation (Sandoe), plant responses to abiotic stress at range margins (Kunin) and identification of binding partners for mammalian ion channels (Wray).

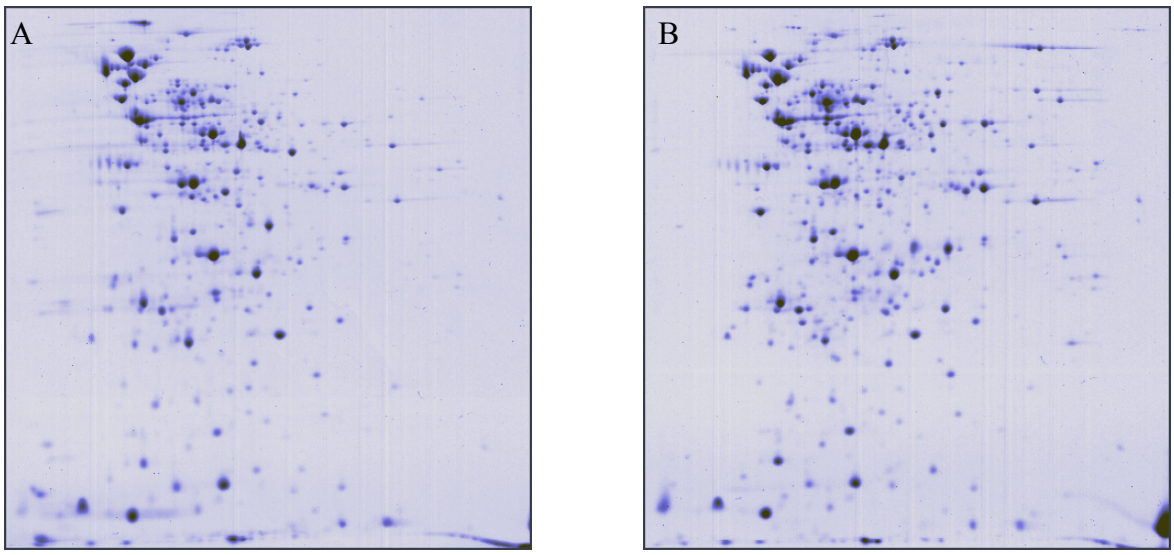


Fig. 2. 2-Dimensional electrophoresis of soluble fractions of *Enterococcus faecalis* grown in standard culture conditions (A) or in serum (B).

Protein sequencing
The N- and C-terminal sequencing facility (funded through JREI 1999) has attracted samples for analysis from various national and international academic institutions and commercial enterprises. 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 unique within the UK, with few instruments worldwide. Interest in the characterisation of protein C-termini for quality assessment or delineation of cleavage sites between structural domains remains, but limitations due to the requirement for large amounts of material (nmol level) and the provision of only a few residues of information restrict its application. The continuation of this facility has been cast in doubt by the withdrawal of manufacturer support for the instrument and the increasing unavailability of specialised reagents.

Collaborators
John Findlay, School of Biochemistry and Microbiology, Leeds.
Jon Sandoe, School of Biochemistry and Microbiology, Leeds.
Peter Grant and Angela Carter, Academic Unit of Molecular Vascular Medicine, Leeds.
Chris Wild and Laura Hardie, Molecular Epidemiology Unit, Leeds.
Bill Kunin, School of Biology, Leeds.

References
Parkinson, N.M, Conyers, C., Keen, J., MacNicoll, A., Smith, I., Audsley, N. and Weaver, R. (2004) Towards a comprehensive view of the primary structure of venom proteins from the parasitoid wasp *Pimpla hypochondriaca. Insect Biochem. Mol. Biol.* **34:** 565-571.

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. and Vanderpump, M.P. (2003) Ghrelin can bind to a species of high-density lipoprotein containing paraoxonase. *J. Biol. Chem.* **278:** 8877-8880.

Parkinson, N., Conyers, C., Keen, J., MacNicoll, A., Smith, I. & Weaver, R. (2003) cDNAs encoding large venom proteins from the parasitoid wasp *Pimpla hypochondriaca* identified by random sequence analysis. *Comp. Biochem. Physiol. C* **134:** 513-520.

Houmeida, A., Thompson, B., Burgess, S., Keen, J., Thirumurugan, K., Tskhovrebova, L., Knight, P.J. and Trinick, J. (2003) Preparation of synthetic titin end-filaments. *Biophys. J.* **84 Suppl.:** 563A.

Funding
We acknowledge the support of BBSRC, NERC, Bio-Rad Laboratories, Waters (Micromass), the HUPO Plasma Proteome Project and the University of Leeds.

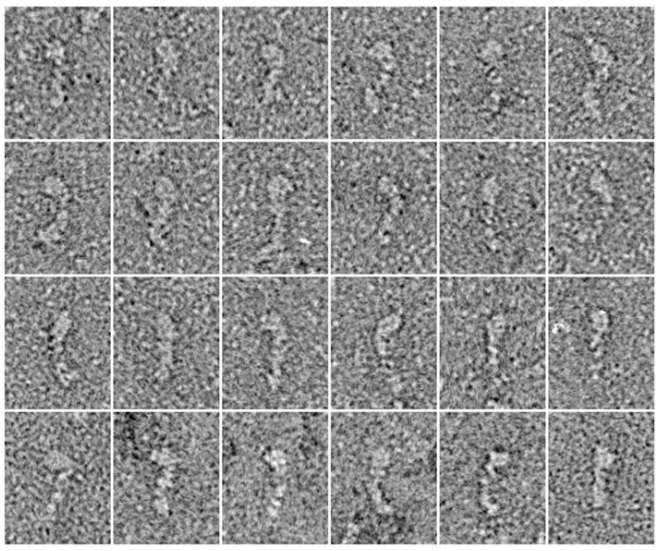
Class I myosins

Matthew Walker and John Trinick

Myosins are identified on the basis of sequence similarities in their motor domain regions, which bind actin and have ATPase activity. Extending from the motor domain is a single α-helix, which binds light chains of the calmodulin class. This acts as a lever arm and produces motion by changing its angle relative to the motor domain. At the end of the lever arm the molecule tail binds cargo. The vast majority of work on myosin has focussed on the type found in muscle, which is called class II. Muscle myosin is a dimeric molecule, with two heads, each with a motor domain and a lever arm. The heads are dimerised through the tail, which is a coiled-coil a-helix and polymerises to form muscle thick filaments.

Recent years have seen the identification of 17 other, non-muscle myosin classes. Based on their sequences, these have similar motor domains, but vary widely in other respects. They may, or may not, dimerise and are also likely to vary in kinetic properties, regulatory mechanisms and cargo binding. Class I myosins are found across a large range of organisms, from amoebe to humans. They are thought to be involved in a range of cellular mechanisms where long persisting forces are needed; for example, in bracing and tensioning the cytoskeleton. The myosin I power stoke is thought to occur in two steps, coupled respectively to Pi and ADP release. ADP release may be strain-sensitive and this, together with a slow ATP-induced dissociation from actin, suggests that myosin I is uniquely suited for maintenance of tension. Studies in lower eukaryotes indicate the importance of myosin I in cell motility, establishment of polarity, and actin organization. A family of myosins I exists in higher cells involved, for instance, in hearing.

We have studied myosin I in collaboration with Drs Lynn Coluccio and Walter Stafford (Boston Biomedical Research Institute, Boston, Massachusetts). Our electron microscopy of purified myosin revealed a monomeric molecule. Consistent with our data, analytical ultracentrifugation gave a molecular mass of 213 kg/mol and dimensions of 28 x 4 nm. Microscopy also suggests the presence of more than one actin binding site in each molecule.



Montage of myosin I molecules. The tadpole-shaped molecules are orientated with their motor domains (which bind actin and cleave ATP) at the top. The pictures appear to show flexibility between the motor domain and the lever arm which transmits force. Each molecule is ~22nm long.

Publication

Stafford, W. F., Walker, M.L., Trinick, J.A. and Coluccio, L.M. (2005) Mammalian class I myosin, Myo1b, is monomeric and cross-links actin filaments as determined by hydrodynamic studies and electron microscopy. *Biophysical Journal*, **88**, 384-391.

Collaborators

Drs Lynn Coluccio and Walter Stafford, Boston Biomedical Research Institute, Boston, Massachusetts.

Funding: This work is supported by the NIH (USA) and the BBSRC.

Class VI myosins

Matthew Walker and John Trinick

We have continued to collaborate on class VI myosin with Dr John Kendrick-Jones (MRC Laboratory of Molecular Biology (Cambridge) and with Dr Claudia Veigel (National Institute for Medical Research, Mill Hill). Myosin VI is unusual in that it walks along actin filaments in the opposite direction to all other myosins so far studied. The mechanism of its reverse gear is therefore of great interest.

The tail of the myosin VI molecule was predicted to dimerise in a coiled-coil a-helix, similar to muscle myosin II. Many papers analysing the kinetic and walking mechanism have been published describing preparations where dimerisation was enforced by inclusion of a GCN4 leucine zipper sequence. However, last year we published the surprising result that the full length molecule is monomeric, both *in vitro* and *in vivo*. Using optical tweezers, our colleagues at Mill Hill showed that the power stroke of myosin VI is much larger than predicted for a molecule with a lever arm containing only 2 light chains (18 vs 5 nm). This suggests that the angular throw of the lever may be substantially larger than seen in other myosins, or that part of the molecule tail may participate in the lever action.

During 2004, papers from other laboratories continued to appear exploring the properties of the obligate dimer. It may be that myosin VI will prove to be dimeric under some conditions *in vivo,* however, there is currently no direct evidence other than the sequence prediction that the molecule dimerises.

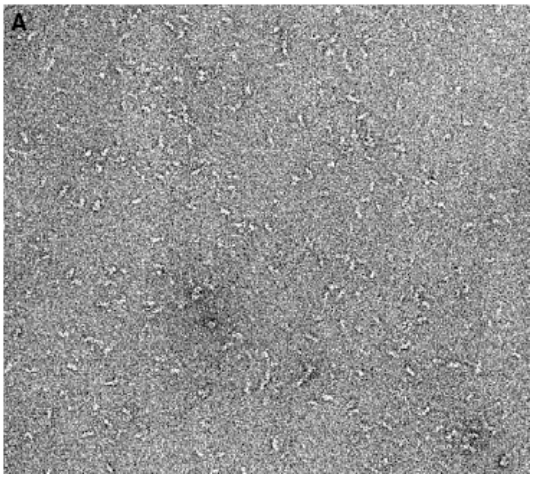


Figure: Field of negatively stained monomeric myosin VI molecules (each about 16 nm long.

Publications

Lister, I., Schmitz, S., Walker, M., Trinick, J., Veigel, C. and Kendrick-Jones, J. (2004) Myosin VI is a non-processive monomer with a large working stroke. *EMBO J*, **23**, 1729-1738.

Roberts, R., Lister, I., Schmitz, S., Walker, M., Veigel, C., Trinick, J., Buss, F. and Kendrick-Jones, J. (2004) Myosin VI: cellular functions and motor properties. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 1931-1944.

Lister, I., Roberts, R., Schmitz, S., Walker, M., Trinick, J., Veigel, C., Buss, F. and Kendrick-Jones, J. (2004) Myosin VI: a multifunctional motor. *Biochemical Society Transactions*, **3**, 685-688.

Collaborators

Dr John Kendrick-Jones (MRC Laboratory of Molecular Biology (Cambridge) and with Dr Claudia Veigel (National Institute for Medical Research, Mill Hill)

Funding

This work is funded by the BBSRC and NIH (USA).

Titin

Larissa Tskhovrebova, Ahmed Houmeida, Nasir Khan and John Trinick

Titin is the largest protein yet described (chain weight 3-3.7 MDa) and the third most abundant protein in muscle. More than half of the titin molecule is bound to muscle thick filaments in the sarcomere, where we have suggested it regulates exact assembly of the 294 myosin molecules known to comprise the filament. The remainder of the titin molecule forms an elastic connection between the end of the thick filament and the Z-line (Fig. 1). These connections are the main source of the passive elasticity of muscle. They also ensure that thick filaments stay in the middle of the sarcomere, which ensures that even forces are developed by myosin in each half of the filament.

Studies of titin elasticity are of considerable interest and have notably been pursued in many single molecule studies (including some by ourselves) using optical tweezers and AFM. It has been temping to assume that the single molecule data can be scaled to explained directly muscle passive elasticity. However, several factors suggest that such direct extrapolation may not be possible, including interactions of the elastic region of titin *in situ*, both with itself and with other proteins. We have shown that a section of the titin molecule ~100 nm long, located in the elastic region near the end of the myosin, self-associates to form a hexameric assembly. This demonstrates that a substantial fraction of the elastic part of titin does not function as independent single molecules. Effects of molecular crowding and confinement within the lattice of actin (thin) filaments are also under consideration, which would also be expected to modulate the the behaviour of the molecule.

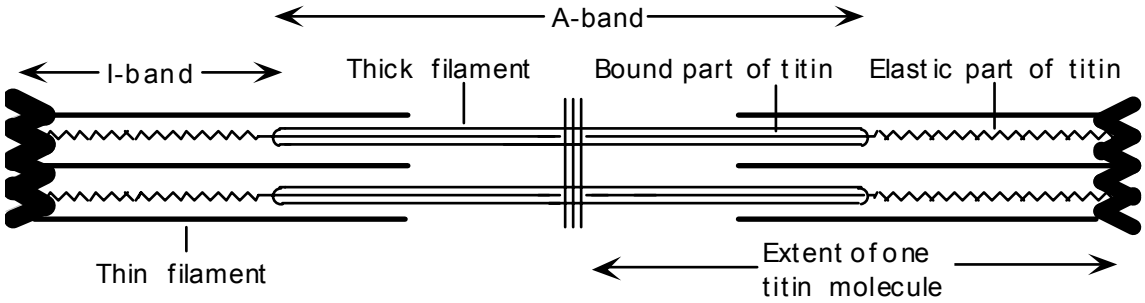


Fig. 1. Schematic of the location of titin in muscle

Publication

Tskhovrebova, L. and Trinick, J. (2004). Properties of titin immunoglobulin and fibronectin-3 domains. *J. Biol. Chem*, 279:46351-46354.

Funding

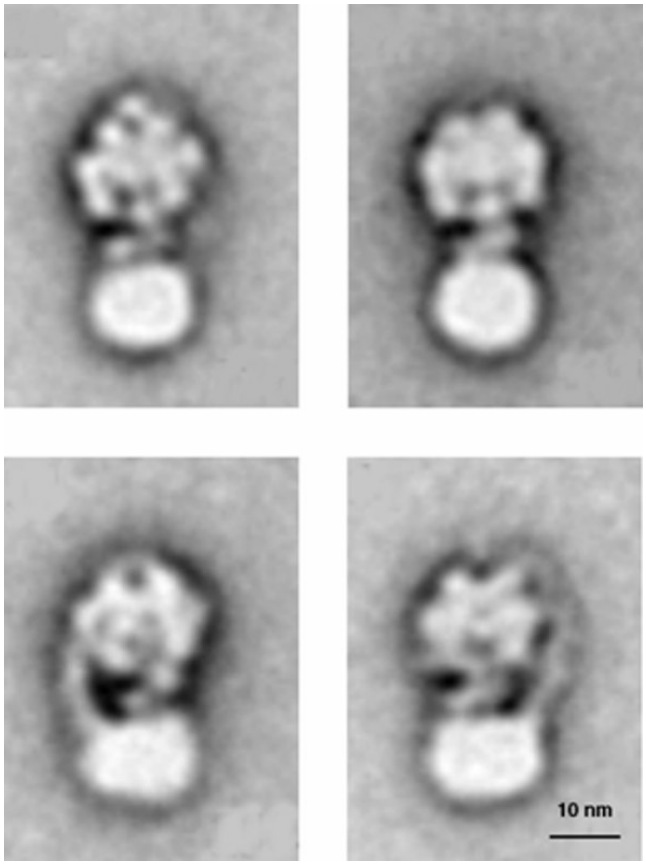
This work is funded by the British Heart Foundation.

Vacuolar ATPase

Chun Feng Song, Michael Harrison, John Findlay and John Trinick

The vacuolar ATPase (V-ATPase) is a rotary molecular motor that acidifies subcellular compartments. It is related to the F_1F_0-ATPase, but is much less well characterised. We have used negative stain electron microscopy and single particle image processing to examine the V-ATPases from the yeast *Saccharomyces cerevisiae* and the bacterium *Enterococcus hirae*. Image averages after alignment and classification of datasets containing 5,000-10,000 particles show both complexes with a similar overall appearance of major soluble (V_1) and membrane (V_0) domains.

Comparisons between the two types reveal substantial differences. The bacterial V_1 appears roughly conical, whereas the yeast V_1 has a less regular profile and shows more internal details, including evidence of a central channel. The yeast V_0 is wider than in the bacterial preparation (12 vs 10 nm) probably due to differences in subunit composition. The region between V_1 and V_0 appears considerably more detailed in the yeast complex, probably due to the presence of subunits found only in eukaryotic V-ATPases. Density extends from here in one (or sometimes two) stalks that go around V_1 and join its top. A stalk was only occasionally seen in the bacterial preparations. Monoclonal antibody labelling of the N-terminal part of the yeast Vma2p subunit was observed directly near the end of V_1, which is consistent with the predicted structure and orientation of this subunit.



Electron microscopy and single particle averaging of Saccharomyces V-ATPase. The panels show image averages after classification of 5500 images of negatively stained V-ATPase into homogeneous groups. Each panel is composed of about 100 images. Two major domains, V_1 and V_0 (lower), are seen, both of which are thought to be rotary motors. These are connected by central and peripheral stalks. There are 13 subunits in the complex totalling 650kDa.

Funding
This work is supported by the BBSRC.

Establishing links in the regulation of antibiotic production

Gabriel Uguru, Jane Towle, Jonathan Stead, Simon Baumberg and Kenneth McDowall

Introduction

Mycelial bacteria of the genus *Streptomyces* and their relatives are of great medical and commercial importance through their production of some 70% of clinically useful antibiotics, in addition to many other therapeutic agents such as antihelminthics and anticancer agents. Moreover, such substances are only a subset of the vast range of secondary metabolites produced by this group. Many, if not most, *Streptomyces* species are capable of producing more than one secondary metabolite, and the timing of the production of secondary metabolites and the quantities generated are exquisitely sensitive to growth and environmental conditions. Where effects on production have been studied in detail, they appear largely to reflect changes in the level of transcription of genes encoding activators that are pathway-specific and found along with the corresponding biosynthetic genes as part of a cluster. As a means of establishing the regulatory pathways that control antibiotic production by this important bacterial genus, the model species *S. coelicolor* and its relatives have been screened to identify regulatory mutants. This has led to the identification of numerous genes in *S. coelicolor* that affect the production of antibiotics. Remarkably, despite the successful characterisation of these genes, there is still no overall understanding of the regulatory network that integrates the various environmental and growth signals to bring about the production of antibiotics in *S. coelicolor*. However, as the initiation of secondary metabolism in most cases has been shown to be associated with increased transcription of a pathway-specific activator, the question of what regulates secondary metabolism can be addressed by identifying and characterizing transcription factor(s) that regulate the promoter of the pathway-specific regulator(s).

Identification of a transcriptional activator of the pathway-specific activator of the actinorhodin biosynthetic genes in *S. coelicolor*

Using electrophoretic mobility shifts as an assay, we have partial purified and then identified a protein from extracts of *S. coelicolor* that binds with specificity to the promoter of *actII-ORF4*, the gene that activates transcription of the biosynthetic genes of actinorhodin, a blue-pigmented polyketide antibiotic. Disruption of the corresponding gene was found to reduce greatly the production of actinorhodin (Fig. 1), but had no detectable effect on the production of undecylprodigiosin and the calcium-dependent antibiotic. These results indicate that this gene, which has been designated *atr*A (actinorhodin transcriptional regulator), has specificity with regard to the biosynthetic genes it influences. AtrA is evolutionarily conserved and we have also shown that it can bind *in vitro* to the promoter of *str*R, a transcriptional activator unrelated to *actII-ORF4* that is the final regulator of streptomycin production in *S. griseus*. These findings provide further evidence that the path leading to the expression of pathway-specific activators of antibiotic biosynthesis genes in disparate *Streptomyces* may share evolutionarily conserved components in at least some cases, even though the final activators are not; and suggests that the regulation of streptomycin production, which serves an important paradigm, may be more complex than represented by current models. Very

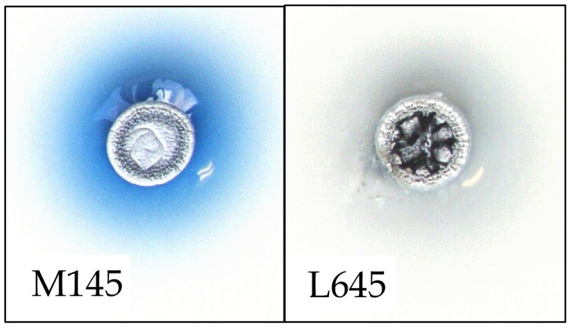


Fig. 1. Phenotypes of mycelial patches of the wild type (M145) and Δ*atrA* mutant (L645).

recently, we have found that AtrA mutants no longer produce actinorhodin in response to the addition of extracts of γ-butyrolactones, small signaling molecules that appear to coordinate the responses of mycelial populations.

The identification of transcription factors using a 2D proteomics approach

The above example illustrates that protein-nucleic acid interactions are central to many regulatory systems. Remarkably, the vast majority of protein-nucleic acid interactions continue to be identified and characterised using standard biochemical techniques. Identifying that a particular segment of nucleic acid is bound by a cellular activity can be relatively straightforward. The bottleneck is often identifying the gene product(s) that is responsible. Traditionally, identification requires extensive purification of the binding activity using liquid chromatography. We have recently developed a proteomics approach that allows the identification of proteins that interact with specific nucleic-acid targets, without requiring their extensive purification. Proof-of-principle is shown in Fig. 2.

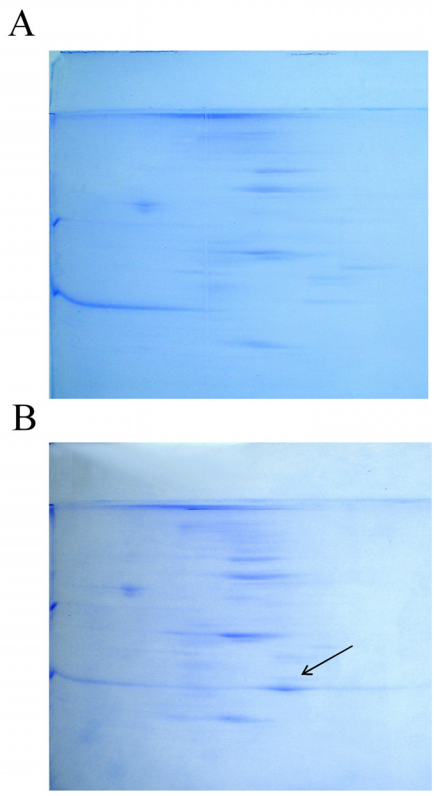


Fig. 2. Identification of a DNA-binding protein in a crude extract of *E. coli* by 2D gel electrophoresis. An electrophoretic mobility shift assay replaces isoelectric focusing in the 1st dimension. Panels A and B were run without and with DNA in the 1st dimension, respectively. An arrow indicates the position of the unique band that corresponds to the DNA-binding protein.

In brief, an electrophoretic mobility shift assay replaces isoelectric focusing in the 1st dimension of 2D SDS/polyacrylamide gels. The formation of a complex not only changes the mobility of the nucleic acid, it also alters the mobility of the protein binding partner. By setting up 1st dimension separations with, and without, the nucleic acid target, and then comparing the bands produced after running the 2nd dimension, it is possible to detect a unique spot that corresponds to the nucleic acid-binding protein. This technique is now being applied to identify additional components of the regulatory network that controls antibiotic production in *S. coelicolor* and other systems of interest to our laboratory.

Funding

This work is funded by the BBSRC.

Structure-led studies of a nuclease central to RNA decay and processing

Yulia Redko, Jonathan Stead and Kenneth McDowall

Introduction

The *Escherichia coli* endoribonuclease RNaseE is required for the rapid decay of mRNA and the correct processing of precursors of ribosomal and transfer RNAs. It is the archetype of a family that is widespread in bacteria and plastids. *E. coli* RNaseE is a relatively large protein of 1061 amino acids. The N-terminal half (NTH; residues 1 to 529) is the ribonucleolytic centre. RNaseE appears to initiate the decay of many, if not most, transcripts in *E. coli* generating upstream products that have a 3'-OH group and are rapidly attacked by 3' exonucleases, and downstream products that have a 5' monophosphate, which greatly stimulates further cleavages by RNaseE. The C-terminal half (CTH; 530 to 1061) contains an ancillary RNA-binding domain and multiple sites that serve as a platform for the assembly of the degradosome complex, which is arguably the main centre for RNA processing and decay in *E. coli*. The other major components of this non-covalent assembly are the 3' exonuclease polynucleotide phosphorylase (PNPase), the RhlB helicase, and the glycolytic enzyme enolase. Components of this complex have been shown to interact functionally, e.g. RhlB assists PNPase to progress through stable stem-loop structures. Our objective is to understand, at the molecular level, all the factors that contribute to the cleavage of transcripts by RNaseE within the context of the multi-enzyme degradosome complex. Structural and biochemical studies are being integrated with an analysis of RNA processing and decay *in vivo*. Our approach has been to start by thoroughly characterising the structure of the N-terminal catalytic domain and the effects on hydrolysis of modifying functional groups within RNA. These studies provide a solid base from which to analyse in more detail the recognition and hydrolysis of RNA by mutating key amino acids, and to begin to probe the structural and functional interplay of enzymes within the degradosome by extending our analysis to include domains within the C-terminal half of RNaseE.

The zinc link and structure-led studies

Recently we have begun to investigate interactions required for the formation of tetramer, and the role of these interactions in mediating the hydrolysis of RNA. The starting point for this aspect was the observation that in *E. coli* RNaseE there is a sequence, CPxCxGxG, between residues 404 to 410, which is evolutionarily conserved in homologues and found in the metal coordinating sites of the chaperonin DnaJ and the DNA mismatch repair protein MutH (also called MutR). In DnaJ and MutH, two copies of the pattern are used to coordinate a single zinc ion in a tetrahedral arrangement. These observations lead us to test whether the motif might play a similar role in RNaseE. We made cysteine to alanine substitutions at 404 and 407 and found, using sedimentation velocity AUC and a discontinuous enzyme assay, that mutation of either of these conserved residues resulted in the disruption of the tetramer into dimers and a greater than 200-fold reduction in the rates of hydrolysis of substrates regardless of whether they have a phosphate or hydroxyl group at the 5' end. Additionally, our collaborator, Dr Ben Luisi (University of Cambridge) has shown that the NTH-RNaseE does indeed contain zinc, with a ratio of protomer to zinc of 2:1, and obtained data consistent with a tetrathiol co-ordination scheme expected for metal co-ordination by four

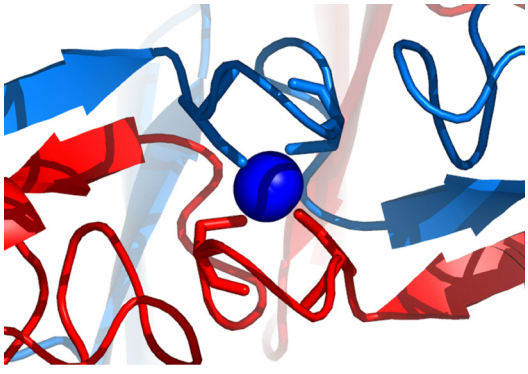


Fig. 1. The Zn Link. Different colours are used to distinguish the two protomers in a pair. Structure solved by the laboratory of Dr. Ben Luisi (University of Cambridge)

cysteine side chains using synchrotron X-ray fluorescence (with Dr. J. Gunter Grossmann, CCLRC Daresbury Laboratory), micro particle induced X-ray emission (with Dr. Elspeth Garman, University of Oxford) and measurements of extended X-ray absorption (with Dr. Lorrie Murphy, University of Bangor), respectively. Taken together, these data are consistent with a model in which the RNase E tetramer is arranged around two, non-equivalent interfaces: The zinc-mediated interface being required for the organisation of the catalytic site but not RNA binding.

We have also started to investigate whether zinc binding may have a role in regulating RNaseE activity *in vivo*. Although the study is incomplete, we are able, through studying the effects of treating preparations of RNaseE and cultures of *E. coli* with diamide (a thiol-specific oxidising agent), to conclude that zinc-binding is redox sensitive and that RNaseE activity is reduced, and mRNA is stabilised in cells during diamide stress. This is a nice example of how structural studies have suggested a possible mechanism by which gene expression is regulated *in vivo*. Dr. Ben Luisi has also been successful in solving crystal structures of the catalytic domain of RNase E as trapped allosteric intermediates with RNA substrates. Four subunits of RNase E associate into an inter-woven quaternary structure that is consolidated through the coordination of two zinc ions that are shared between subunit pairs (Fig. 1). The structure offers explanation for why the quaternary structure is required for activity, and how the recognition of the 5' terminus of the substrate triggers an allosteric transition to initiate catalysis. Mechanistic models are currently being explored through the mutagenesis of the NTH-RNaseE.

Publications

Callaghan, A.J., Redko, Y., Murphy, L., Grossmann, J.G., Yates, D., Garman, E., Ilag, L., Robinson, C.V., Symmons, M.F., McDowall, K. and Luisi, B.F. (2005) The 'Zn-link': A metal-sharing interface that organizes the quaternary structure and catalytic site of the endoribonuclease, RNase E. *Biochemistry* – in press.

Funding

This work was funded by the BBSRC.

Electrodes for redox-active membrane proteins

Steve Evans, Richard Bushby, Simon Connell, Peter Henderson and Lars Jeuken

Introduction

Redox proteins, which are estimated to account for a quarter of all proteins, perform a myriad of functions in biology. They shuttle electrons and catalyse redox reactions in many vital processes, including photosynthesis and metabolism. Dynamic electrochemical techniques have proven to be powerful tools to study these proteins. The thermodynamics and kinetics can be studied in detail if they are electrochemically connected or 'wired' to the electrode surface. The main challenge is to adsorb proteins in their native state on the electrode while efficiently exchanging electrons. Because membrane proteins are more difficult to manipulate experimentally than globular proteins, less work has been reported on the electrochemistry of these proteins. Here, we report a novel approach to link membrane proteins to an electrode surface.

Cholesterol tethers to 'wire' membrane proteins

We have used cholesterol tethers to bind crude membrane extracts (membrane vesicles) of *B. subtilis* to a gold electrode surface. Atomic force microscopy (AFM, Fig. 1), including force measurements, electrochemical impedance spectroscopy (EIS) and surface plasmon resonance (SPR) revealed that the membrane vesicles are 'flattened' upon adsorption, but otherwise remain intact. The natural co-enzyme (i.e., menaquinone-7 [MQ-7]), which is located in these vesicles, can be oxidised and reduced electrochemically. The membrane protein, succinate menaquinone oxidoreductase (SQR), remains in the vesicles and is able to reduce fumarate using MQ-7 as mediator (Fig. 2). The catalysis of the reverse reaction (oxidation of succinate), which is the natural catalytic function of SQR, is almost absent with MQ-7. However, adding the co-enzyme ubiquinone, which has a reduction potential that is about 0.2 V higher, restores the succinate oxidation activity. These results corroborate previous reports that *B. subtilis* uses transmembrane electrical and proton gradients to supply additional energy to oxidise succinate using MQ-7 as electron acceptor.

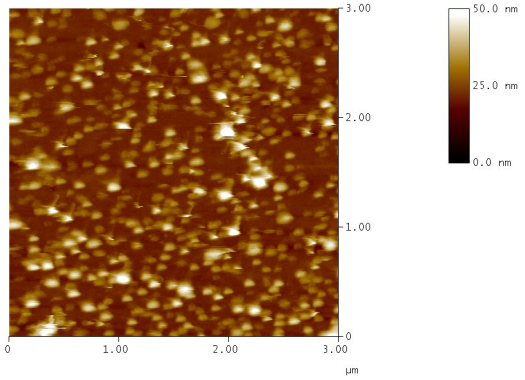


Fig 1. AFM image of a cholesterol-modified gold electrode after adsorption of membrane vesicles of *B. subtilis*

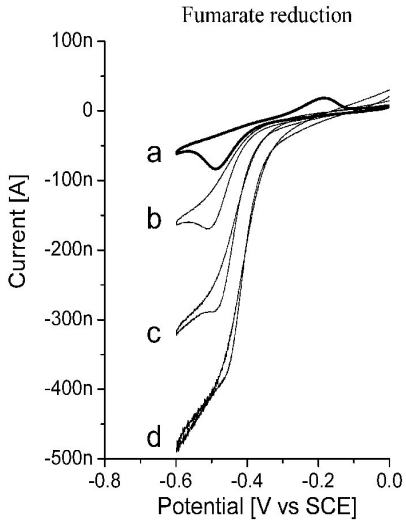


Fig. 2. CVs (1 mV/s) of a cholestrol-modified gold electrode with adsorbed membrane vesicles in the absence (a) and presence (b - d) of 2 mM fumarate at different temperatures, (a-b) 20, (c) 30 and (d) 40◦ C

Publications

Jeuken, L.J.C., Connell, S.D., Nurnabi, M., O'Reilly, J., Henderson, P.J.F., Evans, S.D. and Bushby, R.J. (2005) Direct electrochemical interaction between a modified gold electrode and a bacterial membrane extract, *Langmuir*, In Press.

Collaborators and Funding

Dr. Celicia Hägerhäll (University of Lund)
BBSRC for funding (David Phillips fellowship).

Studies on the Hepatitis C Virus non-structural NS5A protein.

Andrew Street, Holly Shelton, Andrew Macdonald, Chris McCormick, David Brown, Nicholas Burgoyne, Richard Jackson and Mark Harris.

Hepatitis C virus (HCV) infects ~170 million individuals and is a major cause of chronic liver disease, including fibrosis, cirrhosis and hepatocellular carcinoma. 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 which is cleaved into 10 individual polypeptides by a combination of host cell and virus specific proteases. The molecular mechanisms of pathogenesis remain to be elucidated. To this end, our laboratory is interested in understanding the interactions between the NS5A protein and host cell signal transduction pathways.

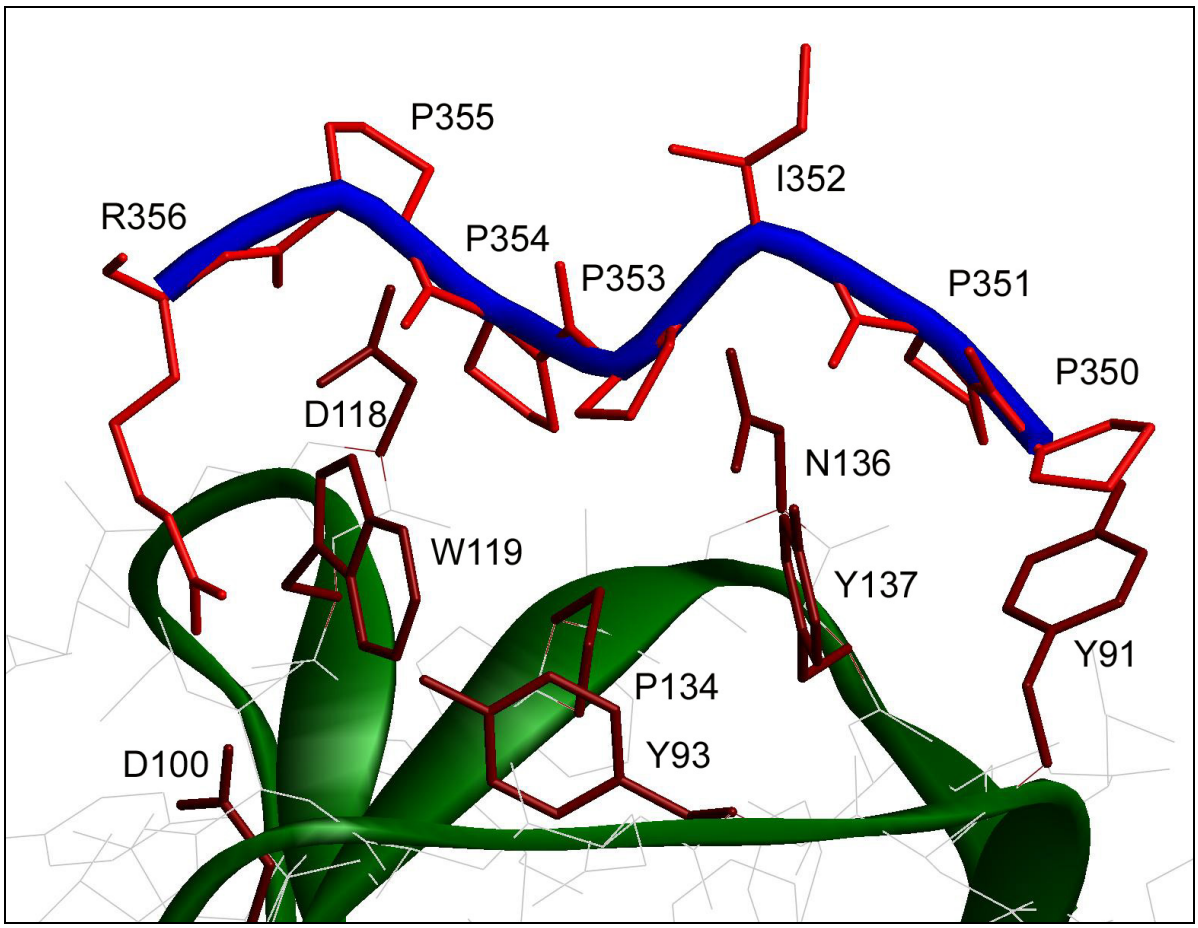


Fig 1: **Molecular modelling of the interaction between the NS5A polyproline motif and the SH3 domain of the Fyn tyrosine kinase.** The blue ribbon represents the backbone of the donor peptide, the red sticks show the atoms of the modelled peptide (residues are labelled in white). The green ribbon represents the backbone of the SH3 domain, atoms of the residues involved in direct interaction with the modelled peptide are shown as dark red sticks. Note that Arg356 in NS5A forms a salt bridge with Asp100 in the SH3 domain – the importance of this has been confirmed by mutagenesis **(5)**. The donor scaffold for this analysis was derived from the published structure of the HIV-1 Nef protein complexed to FynSH3 (Arold *et al* (1997) *Structure* **5**, 1361-1372).

We have shown that NS5A perturbs two key mitogenic pathways within the cell; firstly NS5A blocks the Ras-ERK MAP kinase pathway and, secondly, NS5A stimulates the activity of phosphatidylinositol 3-kinase (PI3K) resulting in the activation of PI3K signalling pathways. Recently we have shown that this promotes cell survival and activitates the proto-

oncogene β-catenin. The latter event has implications for the link between HCV and hepatocellular carcinoma.

We are currently probing the interactions between NS5A and the SH3 domains of members of the Src family of protein tyrosine kinases using a range of techniques including bioinformatics **(see Fig 1)** and surface plasmon resonance, coupled with extensive mutagenesis of both NS5A and SH3 domains. We are also investigating the implications of the interactions for both virus replication and modulation of host cell function.

Publications:

Macdonald, A., Crowder, K., Street, A., McCormick, C. and Harris, M. (2004) The hepatitis C virus NS5A protein binds to members of the Src family of tyrosine kinases and regulates kinase activity. *Journal of General Virology* **85**, 721-729.

Street, A., Macdonald, A., Crowder K. and Harris, M. (2004) The Hepatitis C Virus NS5A protein activates a phosphoinositide 3-kinase dependent survival signalling cascade. *Journal of Biological Chemistry* **279**, 12232-12241.

Macdonald, A., Chan, J.K.Y. and Harris, M. (2005) Perturbation of EGFR complex formation and Ras signalling in hepatoma cells harbouring the hepatitis C virus subgenomic replicon. *Journal of General Virology*, in press.

Street, A., Macdonald, A., McCormick C. and Harris, M. (2005) Hepatitis C virus NS5A-mediated activation of phosphoinositide-3-kinase results in stabilisation of cellular β-catenin and stimulation of β-catenin responsive transcription. *Journal of Virology*, in press

Macdonald, A., Mazaleyrat, S., McCormick, C., Street, A., Burgoyne, N., Jackson, R. M., Cazeaux, V., Saksela, K. and Harris, M. (2005) Further studies on hepatitis C virus NS5A-SH3 domain interactions: Identification of residues critical for binding, and implications for viral RNA replication and modulation of cell signalling. *Journal of General Virology*, in press.

Collaborators:

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

Funding:

This work is funded by the BBSRC, MRC and Wellcome Trust.

Effective sampling of protein conformation space: Exploring local structure propensities

Geraint L. Thomas and Martin J. Parker

Introduction

Modern secondary structure prediction algorithms typically rely on associating homologies between the query sequence and template sequences of known structure. While fast, they are necessarily limited to sequences exhibiting discernable homology with a restricted set of template sequences and structures. Moreover, they do not give a thermodynamic measure of a sequence's conformational tendency. This requires knowledge of the temperature-dependent Boltzmann distribution of the conformation space. Reduced models of proteins, which use statistical and/or physics-based potentials, are quite successful at predicting low complexity folds. Even with this reduced complexity however, sampling methods must be employed to estimate Boltzmann-weighted parameters. Simple energy minimisation may be insufficient in cases of marginal stability, where the time-averaged conformation comprises a diverse range of structures.

Metropolis Monte Carlo (MMC) methods are commonly used for estimating the canonical potential energy surface (PES) of high dimensional systems, in particular protein folding. MMC simulations of protein folding are complicated by two primary factors: the vast size of the configuration space, and the complexity of the PES. The complexity of the PES arises from the heterogeneous interplay of myriad interactions operating over a range of interatomic distances. The resultant PES exhibits a rugged topography with configurations that may be structurally similar but greatly different in energy. Below critical temperatures, tunnelling problems are encountered and the simulation becomes quasi-ergodic in behaviour, i.e. trajectories between structurally diverse but statistically important energy minima have low transition probabilities. The trapping of the simulation in isolated minima leads to statistical inaccuracy in the sample, as the system does not reach equilibrium. The problem is insidious as it is hard to discern quasi-ergodic behaviour in a Monte Carlo simulation.

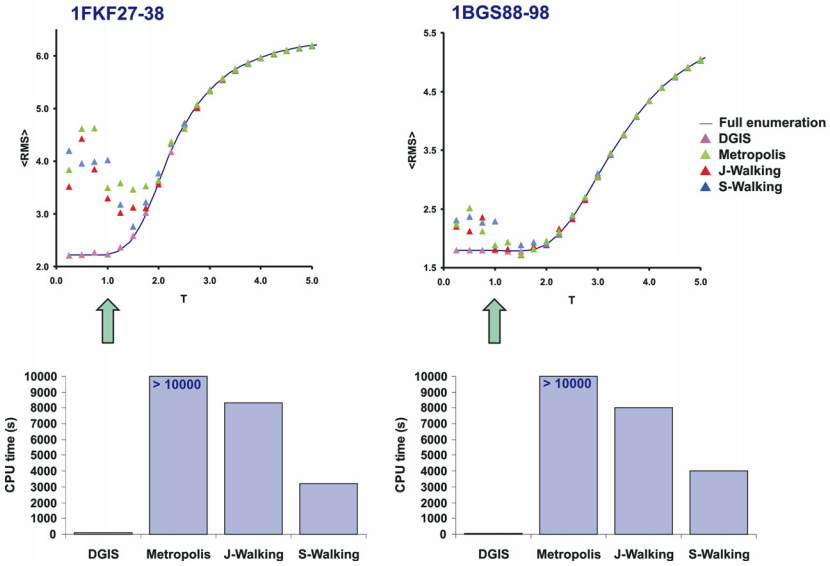


Fig. 1. Boltzmann-weighted RMS (<RMS>) versus model temperature (T) for the two model peptide systems, and CPU time required to reach equilibrium at T = 1.0.

A novel and efficient Monte Carlo method

We have developed a hybrid Monte Carlo method called Density Guided Importance Sampling (DGIS) and applied it to a reduced protein model developed previously called

'RAFT'. In the DGIS method a density of states histogram is calculated initially using a temperature independent algorithm, which is Bayesian in spirit. A non-redundant library of conformations for each energy bin is also constructed. At a given temperature, the density of states histogram is converted into a probability histogram. Standard MMC sampling is then performed at this temperature, during which an occupancy histogram is constructed. At regular intervals the occupancy and probability histograms are compared. Equilibrium is deemed to be reached if they agree. If they do not, a conformation is chosen at random from the library of the most under sampled bin and the walk continued. The method guides the simulation towards under sampled regions and recognises when equilibrium has been reached. Thus, not only is it more efficient, it also has its own internal quality control.

To test our method we chose two β-hairpin peptide systems: 1fkf27-38 and 1bgs88-98. Both are small enough to enable full enumeration using RAFT. The Boltzmann-weighted RMS obtained by full enumeration and the DGIS methods are plotted as a function of temperature in Fig. 1. As can be seen, the DGIS method yields accurate results over the entire temperature range. We have also performed standard MMC sampling along with two other extended Monte Carlo methods developed to overcome sampling inefficiencies: Jump-Walking and Smart-Walking. At each temperature, these were run for the length of time the DGIS method took to converge. For both systems, over the critical temperature range, the DGIS method outperforms all three methods. Also plotted is the time each method takes to reach equilibrium at a temperature of 1.0, where significant tunnelling problems are encountered. As can be seen, the DGIS method is significantly faster.

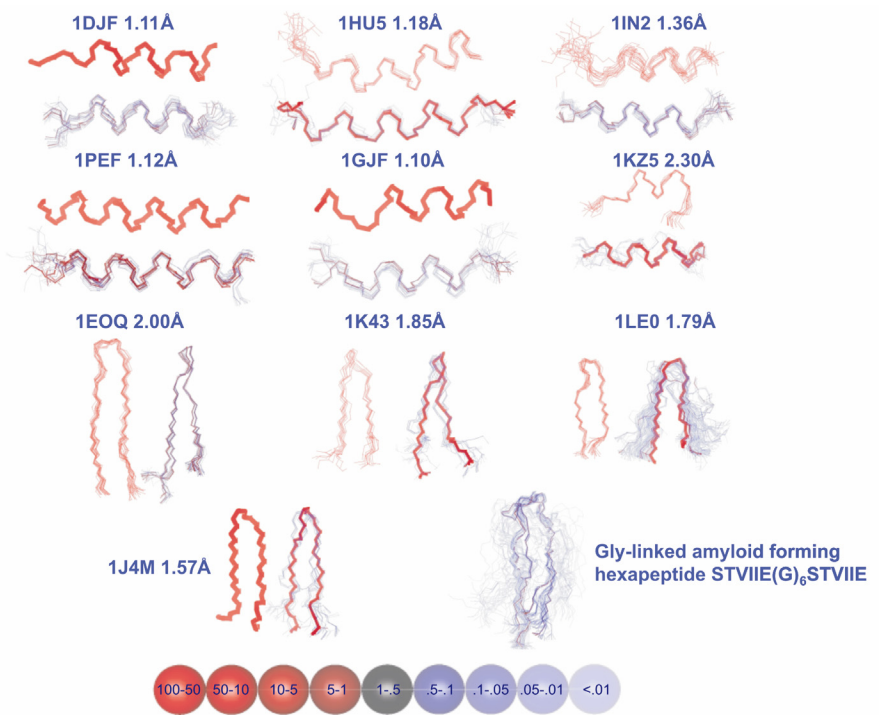


Fig. 2. Experimental peptide structures (upper and LHS images) along with predicted Boltzmann-weighted ensembles (lower and RHS images).

Accurate local structure predictions

We have applied RAFT and the DGIS method to a set of α-helical and β-hairpin peptides for which experimental structures are available. The experimental structures and predicted ensembles are shown in Fig. 2. Individual conformations constituting the ensembles have been superimposed, and are coloured and rendered according to their individual probabilities (see key at bottom of Fig. 2). The Boltzmann-weighted RMS values are shown. As can seen,

the ensemble-averaged structures predicted using RAFT and the DGIS algorithm agree well with the experimental structures in each case. We have also applied this to a hexapeptide designed by Luis Serrano's group. When we link two of these we obtain a β-hairpin. This is consistent with the fact that this hexapeptide forms β-amyloid fibres, and we are now applying this method to the study of amyloidic peptide association.

As the results in this study demonstrate, combining the DGIS method with a reduced but effective protein model serves as a useful tool for exploring and visualising the local conformational propensities of protein sequence fragments. This will have useful applications in protein structure prediction and for understanding protein mis-folding in disease.

Collaborators
Dr Richard Sessions, University of Bristol.

Publications
Sessions, R. B., Thomas, G. L. and Parker, M. J. (2004) Water as a conformational editor in protein folding. *J. Mol. Biol.* **343**, 1125-1133.

Thomas, G., Sessions, R. B. and Parker, M. J. (2005) Density Guided Importance Sampling: Application to a reduced model of protein folding. *Bioinformatics* (in press).

Funding
This work was funded by the BBSRC and The Wellcome Trust. M.J.P. is a BBSRC David Phillips and University Research Fellow.

Membrane sensory proteins in bacteria: how sensor kinases sense and respond to environmental signals in pathogenic and other bacteria.

Eun-Lee Jeong, Christopher A. Potter, Elodie Dupeux, Alison M. Day, Victor Blessie, Peter J. F. Henderson and Mary K. Phillips-Jones

Introduction

In bacteria, sensing and responding to changing environmental conditions is mainly performed by two-component signal transduction systems (TCSs). These systems are crucial for survival in natural environments, or hosts that fluctuate with respect to a wide range of factors, and not surprisingly they are almost ubiquitous amongst bacteria. In pathogenic species, TCSs are pivotal in regulating virulence factor gene expression, and it has been proposed that TCSs constitute excellent new targets for the design of novel antibacterial drugs. TCSs comprise a sensor kinase (SK, or histidine protein kinase) and a response regulator (RR). The SK is usually located in the membrane and is responsible for stimulus perception and signal transduction (via phosphorylation) to the partner RR which then effects an appropriate response. Structural and functional information about membrane SKs is generally lacking, yet this is an important area for elucidating mechanisms of signal 'sensing' by bacteria, and signal transduction across the membrane. No structures of intact membrane SKs have yet emerged. Moreover, knowledge of the signals (environmental cues) themselves is also often lacking. This lack of information arises, in part, because of previous technical challenges in isolating and purifying membrane proteins, including SKs. In collaboration with Peter Henderson's group, we have succeeded in routinely purifying large quantities of membrane SK proteins for structural and functional studies. Progress in the past year is detailed below.

Redox-responsive *in vitro* modulation of the signalling state of the isolated PrrB sensor kinase of *Rhodobacter sphaeroides*.

The global redox switch between aerobic and anaerobic growth in *Rhodobacter sphaeroides* is controlled by the PrrA/PrrB two-component system (also known as the RegA/RegB), in which PrrB is the integral membrane sensor kinase, and PrrA is the cytosolic response regulator. We were the first group to successfully heterologously overexpress the intact membrane protein component, PrrB (or indeed any membrane sensor kinase), using technologies successfully developed by the Henderson group for other groups of membrane proteins.

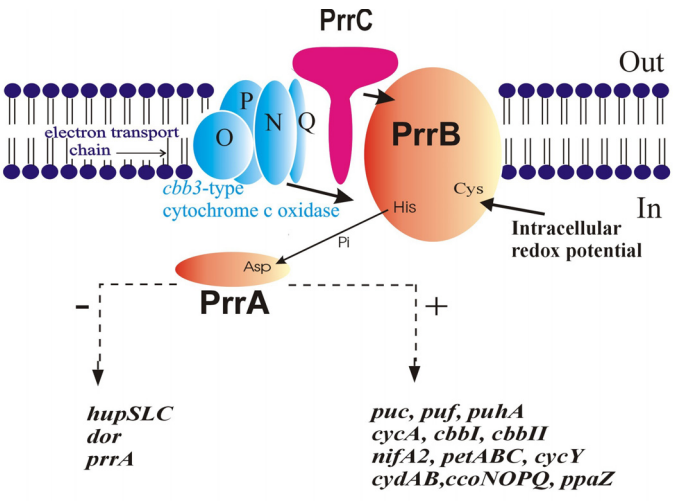


Fig.1. The Prr (Reg) pathway of *R. sphaeroides*

In the case of PrrB, the signal is redox potential, and this signal is generated from within the membrane itself (via the respiratory cbb_3-type cytochrome *c* oxidase and/or another membrane protein PrrC), and/or intracellularly. Previously, we showed that the overexpressed intact protein is functional both in *E. coli* inner membranes and as purified protein, as shown by its autophosphorylation, phosphotransfer and PrrA-dephosphorylation activities. Our kinetic data also revealed that the transmembrane region has important regulatory activity

.

Recently we showed that intact PrrB also retains its ability to sense and respond to its redox signal, paving the way for investigations of structural changes occurring during signal sensing. Fig 2 demonstrates that purified PrrB autophosphorylates in response to reducing conditions induced by reversible thiol exchanger dithiothreitol. This contrasts with a previous study of soluble PrrB (lacking its transmembrane region) which reported only weak increases in PrrB~P in response to DTT-induced reducing conditions. Fig 2 also shows that the higher levels of PrrB~P obtained in these experiments resulted in a concomitant increase in phosphotransfer to PrrA, resulting in higher levels of PrrA~P in our phosphotransfer assays. This therefore confirms that subsequent signal transduction from PrrB to PrrA occurs successfully. That the intact protein also phosphorylates as predicted in response to DTT-induced reducing conditions in the membrane environment is confirmed in Fig 3.

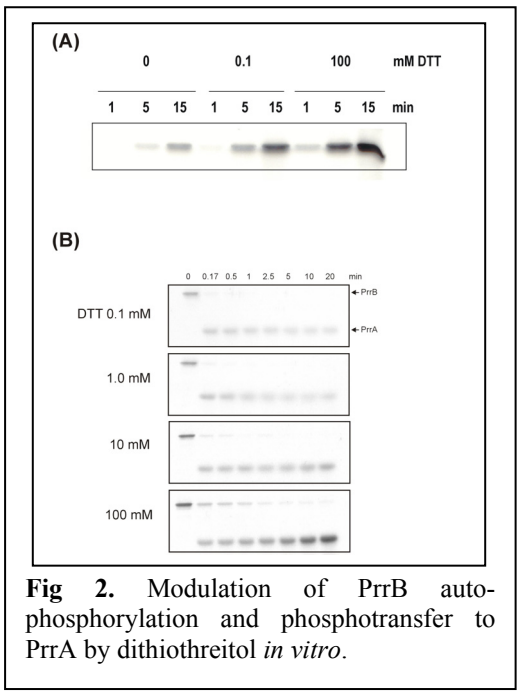


Fig 2. Modulation of PrrB auto-phosphorylation and phosphotransfer to PrrA by dithiothreitol *in vitro*.

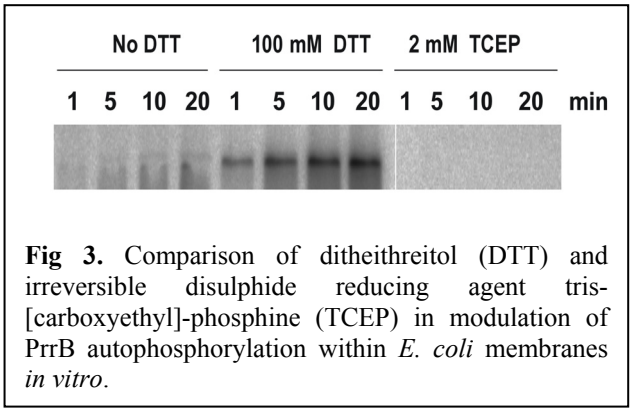


Fig 3. Comparison of ditheithreitol (DTT) and irreversible disulphide reducing agent tris-[carboxyethyl]-phosphine (TCEP) in modulation of PrrB autophosphorylation within *E. coli* membranes *in vitro*.

In the past year, we have identified at least one redox-responsive, membrane-bound protein of *R. sphaeroides* that further modulates PrrB phosphorylation kinetics, and that may therefore regulate levels of PrrB~P *in vivo*. This data has recently been submitted for publication. The system is therefore giving insights into the signal sensing mechanisms. Targetted mutagenesis will facilitate elucidation of the structure-activity relationships of the single PrrB, PrrA and regulatory proteins as well as their complexes. The ability to produce milligram quantities of highly purified PrrB protein is also enabling us to undertake 2D/3D crystallisation in order to elucidate the 3D structure of this sensor kinase by electron or X-ray diffraction.

Overexpression and/or purification of virulence-associated intact SKs from pathogenic bacteria.

This year, we have successfully overexpressed and/or purified a further four, full-length membrane sensor kinases, all of which are known or candidate regulators of virulence factor or antibiotic resistance genes in pathogenic species. In addition to PrrB above, these membrane SKs will also be included in structural studies to elucidate their 3D structure.

Publications

Saidijam M., Psakis, G., Clough, J.L., Meuller, J., Suzuki, S., Hoyle, C.J., Palmer, S.L., Morrison, S.M., Pos, M.K., Essenberg, R.C., Maiden, M.C.J., Abu-bakr, A., Baumberg, S.G., Stark, M.J., Ward, A., O'Reilly, J., Rutherford, N.J., Phillips-Jones M.K. and. Henderson, P.J.F. (2003) Collection and characterisation of bacterial membrane proteins. *FEBS Lett.* **555,** 170-175.

Jeong, E-L., Potter, C.A., Dupeux, E., Day, A.M., Williamson, M.P., Henderson, P.J.F. and Phillips-Jones, M.K. Redox-responsive *in vitro* modulation of the signalling state of the isolated PrrB sensor kinase of *Rhodobacter sphaeroides,* (submitted)

Laguri, C., Phillips-Jones, M.K. and Williamson, M.P. (2003) Solution structure and DNA binding of the effector domain from the global regulator PrrA (RegA) from *Rhodobacter sphaeroides*: Insights into DNA binding specificity. *Nucl. Acids Res*. **31**, 6778-6787.

Collaborators

Mike P. Williamson, Krebs Institute Structural Studies Group, University of Sheffield.
David I. Roper, University of Warwick.

Funding

We are grateful to the BBSRC for financial support, and to Lepjohn Ltd and the Faculty of Biological Sciences, University of Leeds for provision of a studentship to VB.

Convergent transcription studied using the atomic force microscope

Neal Crampton and Neil Thomson

The atomic force microscope (AFM), imaging in tapping-mode (TM), is very adept at determining the relative position of proteins bound to nucleic acid templates. This enables direct visualisation of the centre-of-mass movements of proteins along DNA templates, either in diffusive or energy-coupled processes. This makes the AFM suitable as a tool for studying the outcomes of protein-protein interactions between DNA dependent enzymes on single DNA molecules.

"Nested gene" is a term coined for a gene that lies completely within the sequence of another gene, and frequently in the opposite orientation. Several such genes are known to exist in the human genome and eukaryotes such as *Saccharamyces cerevisiae.* Whether these genes come about as an unavoidable consequence of their compressed genetic arrangement, or whether such genes have an intrinsic effect on transcriptional regulation is unclear. We are using AFM to study the consequences of the encounter between two RNA polymerases actively transcribing in opposite directions on model double-stranded DNA templates. This research aims to answer the question of whether two nested genes can be simultaneously expressed.

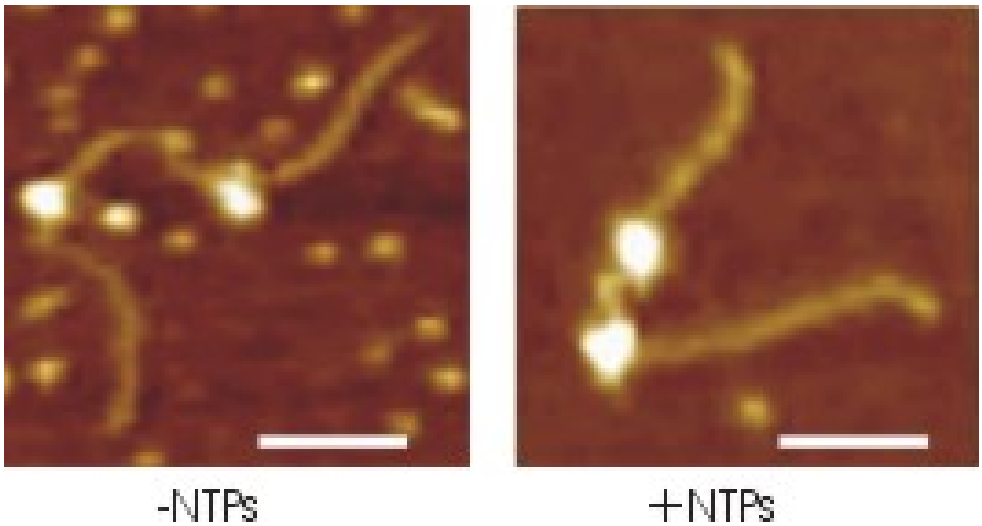


Figure 1: TM AFM images in air of an 1150bp DNA template with two *E. coli* RNA polymerases bound and deposited on mica before, and after addition, of ribonucleoside triphosphates (NTPs). In the absence of all 4 NTPs the polymerases remain at their promoters (left panel). Addition of NTPs cause the RNAP to initiate from their promoters, transcribe the DNA convergently and, in this particularly case, collide and stall while still physically separated (right panel). The scale bars are 75 nm.

AFM imaging can be conducted in air or under aqueous fluid environments, the only requirement being that the protein-DNA complexes must be bound to a flat support surface. Convergent transcription can then either be studied in a discontinuous fashion through imaging samples dried onto mica or potentially in a real-time, single molecule experiment under buffer conditions where the activity of the polymerases on surfaces is maintained. Preliminary results from the discontinuous assays, suggest that the polymerases may be unable to pass each other, with either one being displaced by the other or one stalling against the other, with the possibility that one is back-tracking along the DNA template. However, the time resolution in these "snap-shot" experiments is tens of seconds to minutes and problems of asynchronicity between the two polymerases, makes the interpretation of these data not unequivocal. Future experiments plan to pursue the single molecule experiment in attempt to address these shortcomings. While the scan speeds of most available AFM

technology remains rather slow for these experiments, there are potential ways of setting up the experiment to maximise the information gained (see the upcoming publication in Gene).

Collaborators
Jennifer Kirkham and William A. Bonass
(Oral Biology, Leeds Dental Institute)

Carolyn W. Gibson (University of Pennsylvania, USA)

Claudio Rivetti (University of Parma, Italy)

Publication
Gibson C.W., Thomson N.H., Abrams W.R., Kirkham J. (2005) "Nested genes: biological implications and use of AFM for analysis." Gene*, in press*.

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

Sub-molecular resolution of globular proteins using amplitude modulation atomic force microscopy

Neil Thomson

The atomic force microscope (AFM) is capable of imaging a wide variety of biological samples from whole cells down to through macromolecular assemblies to single molecules, such as nucleic acids and proteins. The success of the AFM in biological imaging comes principally from the development of a.c. imaging modes, where the cantilever is vibrated close to its resonant frequency and either the change in amplitude or frequency of the vibration, when the tip is in proximity to the sample, are used as the feedback signal to produce an image. Amplitude modulation (AM) techniques, such as tapping-mode, have been preferred for biological samples, mainly because these techniques are easier to implement in air and under aqueous fluids (the environments of choice for biological imaging) because the quality factor of the cantilevers used in these environments is relatively low. Frequency modulation (FM) techniques are typically undertaken with very stiff cantilevers in vacuum conditions, where the quality factor of the AFM cantilevers is high and the frequency shifts are readily detectable.

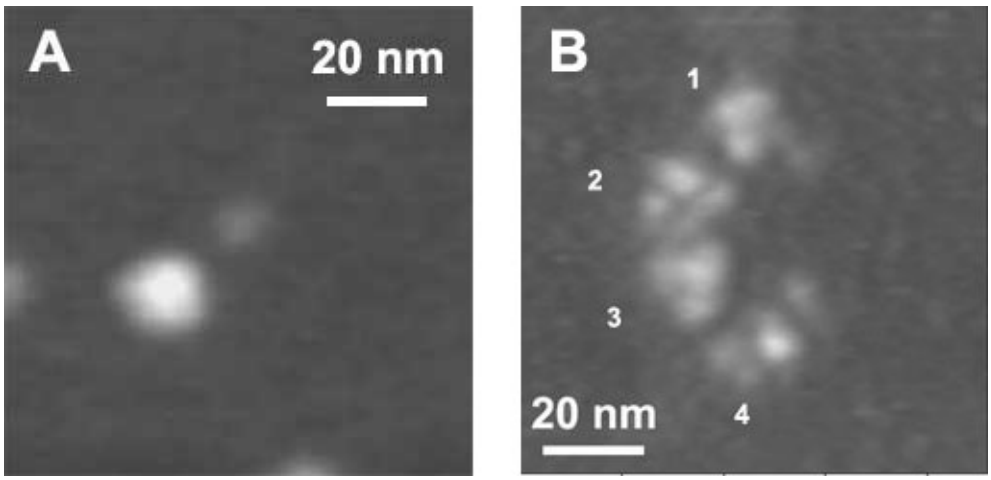


Fig. 1: Amplitude modulation (AM) AFM images of IgG antibody molecules physisorbed to mica taken under ambient conditions. A) Typical image of an antibody taken using repulsive force regime tapping-mode where there is no discernible sub-structure. B) Four IgG molecules on mica taken using AM AFM when the tip is predominantly in the attractive force regime. Each labelled molecule has 3 discernible subunits, one of which tends to be more prominent than the other two (depending upon the protein orientation). These features are attributed to the F_c and 2 F_{ab} domains. Both images are rendered at similar magnification and the height greyscales are 0 to 10 nm. These are monoclonal IgG against the N-terminus of the A-subunit of DNA gyrase and were kindly supplied by Tony Maxwell's group (John Innes Centre, Norwich).

Typically, in tapping-mode isolated globular proteins appear in the AFM images as featureless globules with no internal structure (see Fig. 1A). Relatively recently, however, Ricardo Garcia's group (Madrid, Spain) demonstrated that through understanding of the physics of the vibrating cantilever interacting at the surface and careful control of the microscope parameters, sub-molecular resolution imaging of, in this case, immunoglobulin G (IgG) antibodies was possible. The AFM is thought to be operating in the attractive force regime as opposed to the usual repulsive force regime accessed via tapping-mode. This enabled the 50 kDa F_{ab} and F_c domains of the antibodies to be resolved. The only previous attainment of this goal was achieved by Zhifeng Shao's group (Virginia, USA) who built a cryo-AFM and obtained comparable sub-molecular resolution using contact-mode AFM.

I have successfully repeated the Garcia group's imaging of IgG molecules (see Fig. 1B) and have also found that the amount of water present on the mica support surface, on which the antibodies are bound, can have a very strong influence on whether this sub-molecular resolution is achieved. This effect is related to the energy dissipation from the AFM cantilever and the exact mechanisms involved are still under investigation. Removing the surface bound water through desiccation of the prepared samples enabled reproducible imaging of both IgG and DNA gyrase, with sub-molecular details resolved. When the orientation of IgG on the mica surface was favourable, features consistent with the size of 25 kDa were resolved, where the two immunoglobulin folds in the light and heavy chains crossover. The size and separation of the features measured by AFM were comparable with those obtained from X-ray crystallography data from the protein data bank.

These studies show that although AFM is not a particularly high-resolution technique for structural biology, it can give information about the assembly and topology of multi-subunit proteins and more complex macromolecular systems. This work also serves as a platform for progress in obtaining chemical and biological contrast through AM AFM techniques which will enhance the power of the technique applied to biology.

Publications

Thomson N.H. (2005) "The domain sub-structure of Immunoglobulin G resolved to 25kDa using amplitude modulation AFM in air." *Ultramicroscopy,* in press.

Thomson N.H. (2005) "Imaging the sub-structure of antibodies with tapping-mode AFM in air: The importance of a water layer on mica " *Journal of Microscopy* **217**, 193-199.

Acknowledgements

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

Investigating the affinity of molecular interactions of the ϕ29 packaging motor

Mark A. Robinson, Arron Tolley and Nicola J. Stonehouse

Introduction

ϕ29 is a bacteriophage responsible for the infection of *Bacillus* species and its proposed mode of action of DNA packaging has stimulated much recent interest. Commonly, biologically endogenous molecular motors display either a rotary motion e.g. bacterial flagella, or a linear movement e.g. muscle contraction. The ϕ29 molecular motor consists of a protein/RNA complex and the action of this motor results in packaging of the double-stranded genomic DNA into preformed procapsids. A novel RNA-RNA multimerisation event is thought to be involved, resulting, in combination with ATPase activity, in possible rotation of the molecular motor and concomitant translation of genomic DNA through the connector complex.

Results

The rules governing this protein/RNA interaction, for example the possible existence of sequence specificity, are not known and the aim of this current work is directed towards developing an understanding of this system. Both RNA (pRNA) and connector protein components form multimeric complexes and the multimerisation of pRNA has been previously investigated in this laboratory by analytical ultracentrifugation (AUC) and light scanning experiments. These experiments indicated that pRNA undergoes Mg^{2+}- dependent formation of monomeric, dimeric and trimeric species, which may act as components of the ultimate higher order species.

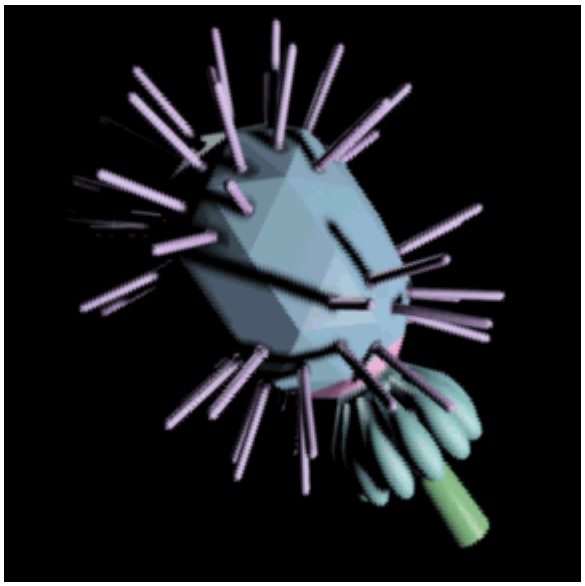


Fig. 1: Cartoon representation of ϕ29 produced by Katie Radcliffe
The phage procapsid is shown in blue, spikes in purple and the head-tail connector in pink. pRNA is not shown as it is not present in an assembled phage particle.

Current work has involved the chemical and enzymatic synthesis of novel pRNA species – hairpin loop domains and full length molecules, both wild-type and mutant. We have shown that the affinity of the pRNA-pRNA interaction is μM, whereas the K_D of pRNA-connector binding is nM by surface plasmon resonance studies. Connector binding affinity is increased in the presence of magnesium ions, however, this affinity does not seem to be dependent on pRNA multimerisation (although pRNA multimerization is essential for DNA packaging). Differences in the magnitude of affinity between components of the motor lead us to deduce that it is unlikely that there is a rotation between the connector and pRNA during DNA

packaging, and we propose that magnesium-induced conformational changes could drive the packaging event.

Collaborators
Neil Thomson
Peixuan Guo, Purdue University

Publications
Wood J.P.A, Capaldi, S.A, Robinson, M.A., Baron, A.J. and Stonehouse, N.J. (2005) RNA multimerisation in the DNA packaging motor of bacteriophage ϕ29. *Journal of Theoretical Medicine*. In press.

Funding
Funding from the BBSRC is gratefully acknowledged.

Using *in vitro* selection to investigate RNA - protein interactions in picornaviruses

Mark Ellingham, David Rowlands and Nicola J. Stonehouse

Introduction

Foot-and-mouth disease virus (FMDV) is a highly important animal pathogen. As a member of the picornavirus family, it consists of a single strand of RNA surrounded by a protein capsid. Clinically, foot-and-mouth disease (FMD) is characterised by painful vesicles that develop in the mouth, hoof and udders of an infected animal and results in chronic lameness and a permanent loss of performance. As well as the suffering endured by the animal itself, the disease can have a significant economic impact. For example, the cost of the 2001 FMD outbreak in the British Isles has been estimated at around £9 billion.

Results

SELEX is a powerful technique that works on the principle that nucleic acid sequences isolated on the basis of their binding affinity to a specific target can be preferentially amplified over competing sequences. Repetition of this process results in a population of sequences (aptamers) that specifically bind to the target with a very high affinity.

15 rounds of selection have been performed against the 3D protein (an RNA polymerase) of FMDV-C using a random sequence library of RNA as a starting pool, and with the intention of generating high affinity RNA ligands against the protein. Binding studies have been carried out, which indicated that the final RNA product of the process had a higher affinity for the 3D protein than the starting pool of random RNA sequences. In addition, a number of RNAs from the final product were sequenced and found to have significant areas of homology with the 5'-untranslated region of the viral genome. This is illustrated in Fig. 1.

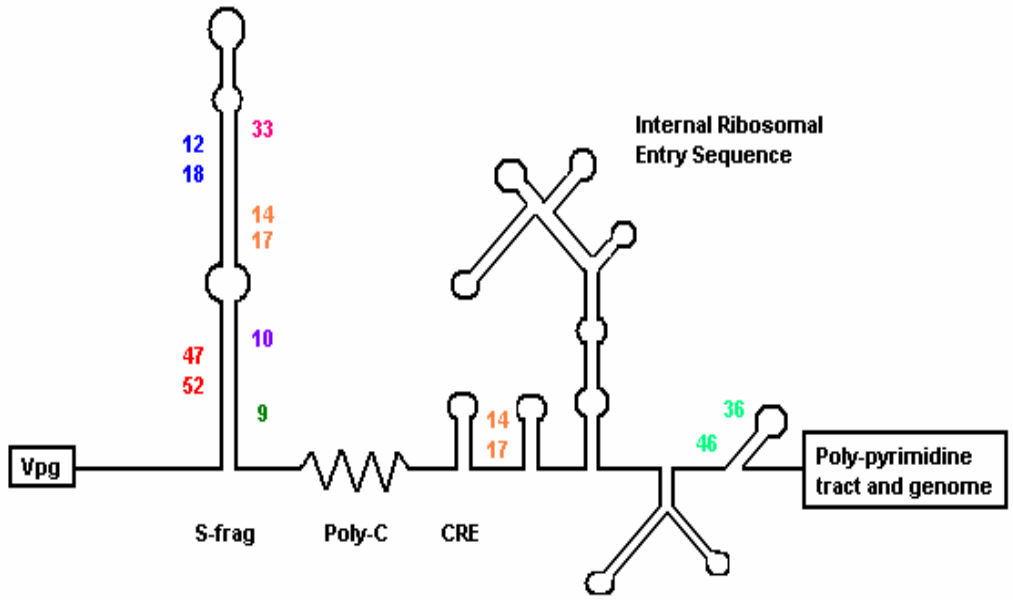


Fig. 1. Illustration of 5' UTR of FMDV, showing areas of homology with sequences of selected numbered aptamers. Coloured numbers represent related aptamer sequences.

Certain aptamer species were chosen from the pool to assess their ability to inhibit the *in vitro* catalytic activity of the 3D protein. Three aptamers were found to interact with the enzyme in such a way that its ability to transcribe RNA was significantly diminished. Further studies are to be undertaken to assay the ability of these aptamers to inhibit viral progression *in vivo*, as well as structural studies whose aim is to elucidate binding motifs on both the aptamers and 3D.

The inhibitory properties of these aptamers could potentially lead to applications such as diagnostics, molecular probes or even therapeutics.

Collaborators
Esteban Domingo (Madrid)
Nuria Verdaguer (Barcelona)

Funding
With thanks to the BBSRC for funding.

Single molecule spectroscopy of a mechano-chemical motor controlling transcriptional initiation.

Rob Leach, Chris Gell, Alastair Smith and Peter Stockley

Introduction

Controlling the rate of transcription initiation provides an organism with the means to alter gene expression in response to environmental and/or developmental cues. In prokaryotes, a single core RNA polymerase (RNAP) facilitates transcription from different promoter sites through the formation of distinct holoenzymes where a sigma factor subunit protein complements RNAP. The sequence-specific DNA binding of sigma factors then allows RNAP holoenzymes to selectively target a range of promoters.

Sigma54 endows the prokaryotic RNAP complex with a further level of control more commonly seen in eukaryotes, where upstream DNA enhancer elements are utilised to control transcription at promoter sites. When sigma54-RNAP binds to a promoter it blocks the DNA melting function of the holoenzyme, which rests in an inactive but responsive state. Activation of this quiescent complex occurs when an enhancer-binding ATPase, e.g. PspF, bound to a site distal to the promoter, is brought into contact with sigma54 through formation of a DNA loop. This transient interaction causes the sigma54 to undergo an ATP-dependent isomerisation that allows transcription to proceed. Incubation of PspF with the transition state analogue, ADP.AlF_4, traps it as a ternary complex with promoter bound sigma54-RNAP, enabling further analysis of protein:protein interactions in biochemical assays or structural studies.

Results

Using site-directed cysteine mutants of sigma54 labelled with Alexa Fluor 594, we have investigated sigma54-RNAP holoenzyme binding to *nif*H promoter DNA sequences, labelled with Alexa Fluor 488, in ensemble assays. Excitation of donor fluorescence at 488nm leads to emission at ~515nm. In the presence of acceptor, this diminishes and an acceptor peak appears at ~615nm, demonstrating fluorescence resonance energy transfer (FRET) (Fig. 1). Differences in donor quenching by differently labelled sigma factors has provided an approximation of which residues may be closer to the donor fluorophore (Fig. 1). Addition of PspF with or without ATP, or non-hydrolyzable analogues, reveals a pronounced bathochromic shift in the acceptor fluorescence, indicating the fluorophore is experiencing a more polar environment and providing further evidence of sigma54 isomerisation in response to activation by the mechano-chemical motor, PspF (Fig. 2).

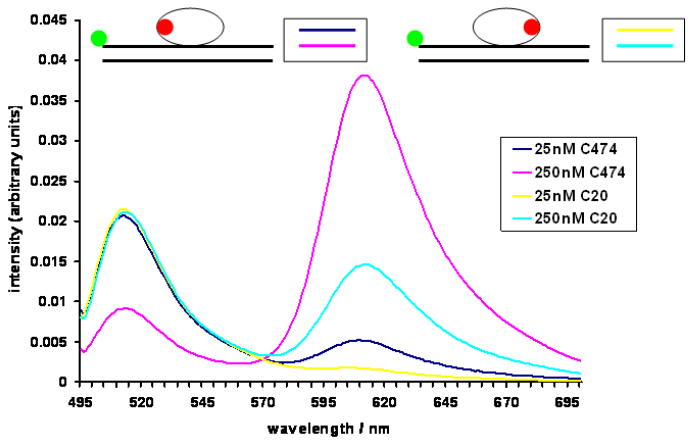


Fig. 1. Ensemble FRET assay of sigma54-RNAP binding to 25nM nifH promoter DNA. Donor fluorescence is quenched and acceptor fluorescence increases in the presence of 250nM sigma54 C474 but not 250nM sigma54 C20.

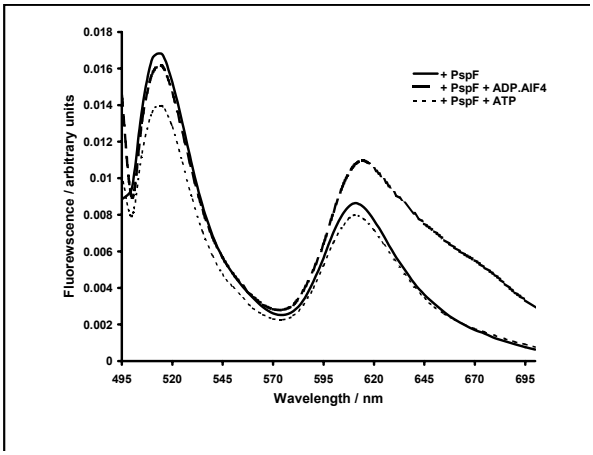


Fig. 2. Ensemble FRET assay of cysteine-mutant σ54 in the presence of PspF. Ensemble FRET assays were repeated in the presence of 500nM PspF activator protein. Emission spectra were then measured and a final concentration of either 200nM ADP + 5mM NaF + 200nM $AlCl_3$ or 2mM ATP added. After incubation at 30°C for 5 mins emission spectra were measured again. It is thought that PspF and sigma54 form an unstable complex that is stabilised by the presence of the transition state analog, $ADP.AlF_4$. The red-shifted emission spectrum for sigma54 in the presence of PspF ± $ADP.AlF_4$ implies a change in the polarity of the fluorophore's environment. In the presence of hydrolysable ATP this red-shift is not present, suggesting a change in structure as predicted for an isomerisation event.

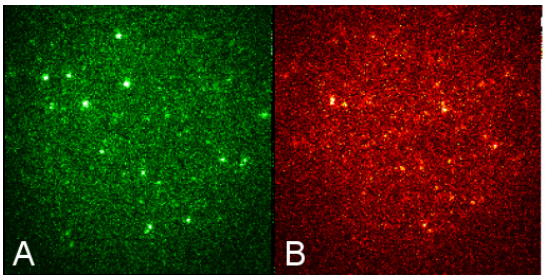


Fig. 3. TIRF image of single molecules. Single molecules of Alexa Fluor 488-labelled biotinylated nifH promoter (A) are immobilised on biotin-BSA coated glass coverslips via interaction with streptavidin. After incubation with ~30fM Alexa Fluor 594-labelled sigma54-RNAP complex, fluorescence due to FRET can be detected (B).

Our work has progressed to the stage where we can now start to dissect the multiple conformational changes occurring over time in the above ensemble assays at the single molecule level. Functionalised glass coverslips bearing immobilised donor-labelled *nif*H promoters are illuminated by total internal reflection fluorescence (TIRF) in a confocal microscope (Fig. 3A). Binding of acceptor-labelled sigma54-RNAP (Fig. 3B) will result in an anti-correlated FRET signal where donor fluorescence decreases as acceptor fluorescence increases. Resolution of these FRET signals over microseconds to seconds will allow us to investigate the conformational changes required for transcription as the complex proceeds from initial binding to ATP-dependent isomerisation and thence to transcriptional activation.

Collaborators
Prof. Martin Buck & Dr Sivaramesh Wigneshweraraj, Department of Biology, Imperial College of Science, Technology and Medicine, London, United Kingdom.

Funding
We acknowledge the support of the BBSRC.

Development of regulated RNA aptamers as tools for *in vivo* post-genomic analysis.

Tamara Belyaeva, Blandine Clique, Ben Whittaker, Adam Nelson,
Ian Hope and Peter Stockley.

There is a great deal of interest in naturally occurring RNA domains that can lead to direct genetic regulation by the small molecular weight products of biosynthetic pathways – "riboswitches". Our long term goal is to create a generic version of these regulatory systems that can be "switched" between off- and on-states by the addition to the cellular growth medium of small molecular weight effector molecules. This strategy will allow critical issues of the roles of specific macromolecular interactions throughout cellular growth and development to be addressed in our test model organism *C. elegans*.

As a proof of principle of this idea, we have used the *E. coli* methionine apo-repressor, MetJ, as a model system. We used a number of different approaches for selection of aptamers against MetJ, including robotic selection on a Biomek 2000 Automated Workstation, gel-retardation and filter-binding assays. Isolated RNA aptamers bind to MetJ with K_d's ~5 nM, and were able to interfere efficiently with MetJ binding to promoter DNA *in vitro.* In order to create an allosterically-regulated aptamer against MetJ, we tested a library of novel aminoglycoside derivatives for antimicrobial activity against an *E. coli* reference strain. Five such aminoglycoside compounds from the library that did not show antibiotic activity were chosen as the potential effector molecules, and we have begun the selection of aptamers against these compounds. Once tight binding aminoglycoside domains are to hand, the next step will be to combine the two aptamer sequences, anti-target aptamer, and effector-binding aptamer, into the same RNA molecule, followed by further selection to obtain allosteric coupling. Such RNA aptamers would have great potential in regulation of protein interactions in response to the addition of small molecular weight effector.

Funding
We acknowledge the support of the BBSRC for this research.

Fleximers of the dynein motor protein

Stan Burgess, Matt Walker and Peter Knight

Introduction

Flexible macromolecules pose special difficulties for structure determination by crystallography or NMR. Progress can be made by electron microscopy, because the shapes of individual molecules are recorded, but electron cryo-microscopy of unstained, hydrated specimens is limited to larger macromolecules because of the inherently low signal-to-noise ratio. For 3-dimensional structure determination, the single particles must be invariant in structure. We have used negative staining and single-particle image processing techniques to overcome some of these limitations and to explore the structure and flexibility of single molecules of the microtubule motor protein dynein.

The approach we have used is to align the images based on features of one part of the molecule (e.g. the head) and then group the images into classes based on features of another part (e.g. the position of the stalk or tail). The averaged image of each group then shows consistent features of the class, and the whole series of class averages shows the range of shapes (the 'fleximers') the molecule adopts.

We show two-dimensional projection images of negatively-stained dynein-c, a flagellar inner arm dynein from *Chlamydomonas reinhardtii,* revealing new details of its structure. The dynein heavy chain (~500 kDa) comprises three domains: tail, head, and stalk. The head contains six AAA+ modules and forms a ring-like globular domain ~13 nm in diameter. The stalk, which emanates from the head, is a 15nm-long, intramolecular, anti-parallel, coiled coil, which has at its distal end the microtubule-binding domain. The ~25nm-long tail domain contains dimeric p28 and monomeric actin light chains, responsible for attachment of this motor to its cargo (a doublet microtubule within the 9+2 axoneme). Both the proximal tail and the stalk are flexible, suggesting that they act as compliant elements within this motor. This is shown in three different views of the molecule (Fig. 1).

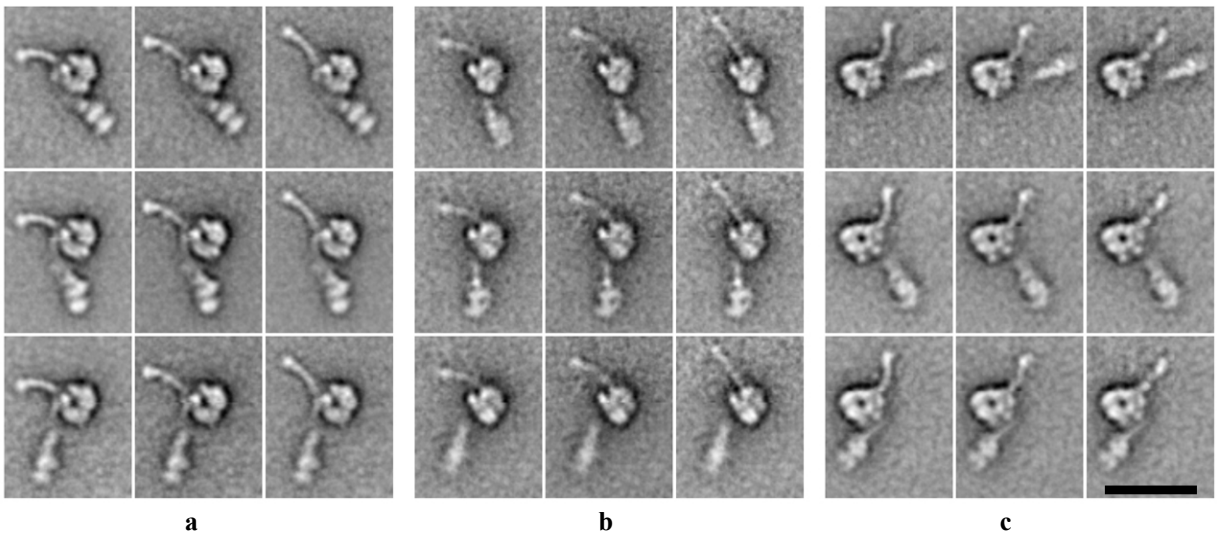


Fig. 1 Fleximers of dynein-c.
(a) Left, (b) side and (c) right views of apo-dynein-c, each showing 3 different stalk (columns) and tail (rows) conformations as illustrating the range of positions these domains adopt in relation to the head domain.
Scale bar 30nm.

Analysis of left views of ADP.Vi-dynein and apo-dynein (representing pre- and post-power stroke conformations of the motor) shows the combined effect of tail and stalk flexibility on the spatial distribution of the microtubule-binding domain relative to a fixed distal tail (Fig. 2).

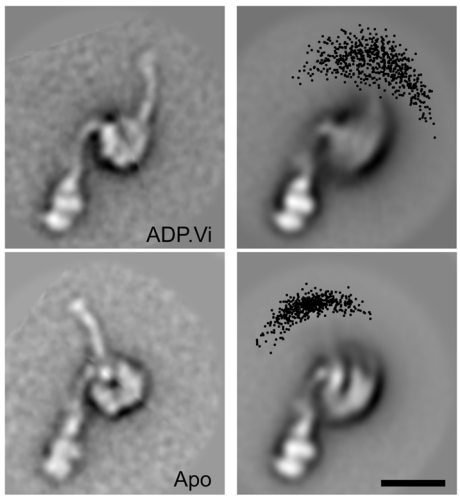


Fig. 2 Powerstroke of dynein-c.
(Left) Mean conformations of ADP.Vi (upper) and apo- (lower) molecules after alignment of their tail domains. (Right) Distribution of microtubule-binding domain positions from all fleximers. Scatter plots are superposed on the global averages which shows, as expected, smearing of the head and stalk. Scale bar: 15 nm.

The mean conformational change produces a ~15nm displacement of the microtubule-binding domain. However, the analysis of fleximers indicates the potential range of power stroke sizes about this mean which might underlie the inverse dependence of the step size on load seen by others in optical trap experiments.

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

Publications

Burgess, S.A., Walker, M.L., Sakakibara, H., Oiwa, K. and Knight, P.J. (2004) The structure of dynein-c by negative stain electron microscopy. *J. Struct. Biol.* **146**, 205-216.

Burgess, S.A. and Knight, P.J. (2004) Is the dynein motor a winch? *Curr. Opin. Struct. Biol.* **14**, 1-9.

Burgess, S.A., Walker, M.L., Thirumurugan, K., Trinick, J. and Knight, P.J. (2004) Use of negative stain and single-particle image processing to explore flexibility in macromolecules. *J. Struct. Biol.* **147**, 247-258.

Funding
This work was funded by the BBSRC.

Regulated conformation of myosin 5

Kavitha Thirumurugan and Peter Knight

We have found that myosin 5, an important actin-based vesicle transporter, has a folded conformation that is coupled to inhibition of its enzymatic activity in the absence of cargo and Ca^{2+}. In the absence of Ca^{2+}, where the actin-activated MgATPase activity is low, purified brain myosin 5 sediments in the analytical ultracentrifuge at 14 S as opposed to 11 S in the presence of Ca^{2+} where the activity is high. At high ionic strength, it sediments at 10 S independent of Ca^{2+} and its regulation is poor. These data are consistent with myosin 5 having a compact, inactive conformation in the absence of Ca^{2+} and an extended conformation in the presence of Ca^{2+}or high ionic strength. Electron microscopy reveals that in the absence of Ca^{2+}, the heads and tail are both folded to give a triangular shape (Fig. 1a), very different from the extended appearance of myosin 5 at high ionic strength (Fig. 1c). Single particle image processing of these folded molecules reveals an enigmatic but rather consistent structure (Fig. 1b). A recombinant myosin 5 heavy meromyosin fragment that is missing the distal portion of the tail domain is not regulated by Ca^{2+} and has only a small change in sedimentation coefficient, which is in the opposite direction to that seen with intact myosin 5. Electron microscopy indicates that its heads are extended even in the absence of Ca^{2+}. These data suggest that interaction between the motor and cargo binding domains may be a general mechanism for shutting down motor protein activity and thereby regulating the active movement of vesicles in cells.

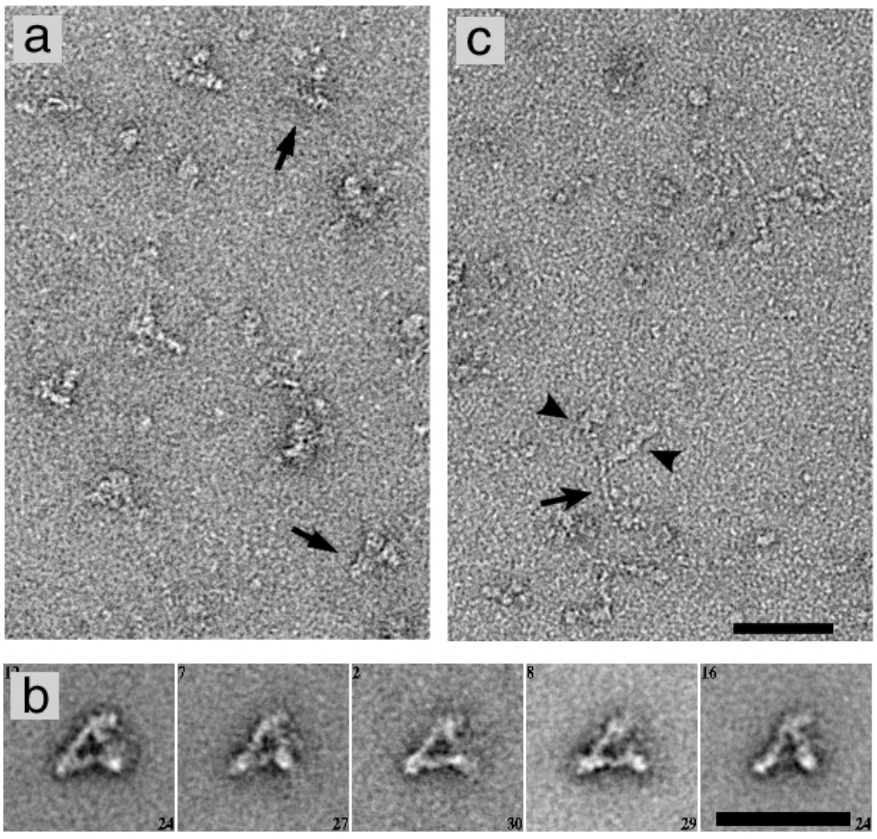


Fig. 1 Negative stain electron microscopy of myosin 5.
a, low Ca^{2+}, low ionic strength; arrows indicate two examples of the compact molecules visible in this field. **b**, typical averaged images from single particle image processing of compact molecules; class sizes are about 30 molecules. **c**, high ionic strength, low Ca^{2+}; molecules are not compact. Instead, the two heads (arrowheads) and tail (arrow) of molecules are visible. Scale bars, 50 nm.

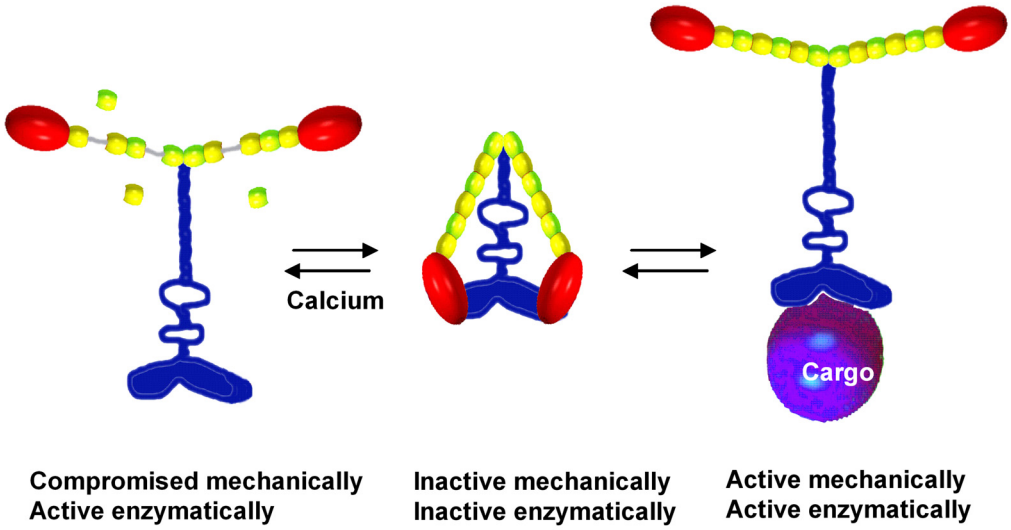


Fig. 2 Speculative model for conformation-dependent regulation in myosin 5.
The compact, shut down molecule (centre) could convert to the opened active state as a result of becoming attached to cargo (right). Elevated Ca^{2+} levels also activate the enzymatic activity, but if Ca^{2+}-binding by the calmodulin subunits of the myosin levers causes them to dissociate (left) it might compromise mechanical activity due to a consequent weakening of the lever.

An advantage of having a highly-regulated myosin 5 is that unregulated, non-cargo-bound myosin 5 in cells would needlessly hydrolyse ATP and would stay associated with actin filaments due to its processivity. We envisage that regulated myosin 5 would dissociate from its cargo on reaching its destination and then fold into the inactive form that could freely diffuse and effectively be recycled. These ideas are summarised schematically in Fig. 2.

Collaborators

James R Sellers, John A Hammer III and Fei Wang (NIH, MD, USA).
Walter F Stafford (Boston Biomedical Research Institute, MA, USA).

Publications

Wang, F., Thirumurugan, K., Stafford, W.F., Hammer, III, J.A., Knight, P.J. and Sellers, J.R. (2004) Regulated conformation of myosin V. *J. Biol. Chem.* **279**, 2333-2336.

Funding

This work was funded by the BBSRC.

Structural modelling of protein-DNA interactions

Richard Gamblin and Richard Jackson

A large number of *in silico* models of protein-DNA interactions reflect a simplistic view of DNA recognition, whereby proteins recognise a specific sequence of nucleotides. These models use aligned binding sites to generate a position specific scoring matrix (PSSM), in what is essentially a table giving the prevalence of nucleotides at each position in the section of bound DNA. This can then be used to characterise and identify other such binding sites in DNA sequences.

This representation of recognition is, however, far simpler than present *in vivo*. DNA is a negatively charged bi-polymer projecting into three dimensional space rather than a one dimensional series of letters, and it is this that proteins must negotiate to distinguish specific regions. The difference between the PSSM models of transcription factor binding and the *in vivo* reality provides an explanation as to why PSSM binding site predictions are plagued with erroneous predictions. Recent studies into protein-DNA complex structures have indicated the presence of trends in amino acid-DNA base interactions, and preliminary results from the latest models of protein-DNA recognition, which utilise this type of information, appear promising.

We have developed a novel method with the aim of quantifying structural features that confer specific binding properties not evident from sequence similarity alone. Using a catalogue of hydrogen bonding and non-bonded contact patterns from a non-redundant set of protein and DNA complex structures, an overall statistical knowledge based model was developed to represent specific amino acid-DNA base/ backbone interactions. This was applied to create new PSSM-type models, termed structurally derived matrices (SDMs).

Assessment of the functional differences between the SDM and PSSM models revealed that SDM predictions were significantly poorer than the equivalent PSSM predictions. The SDMs correctly predicted binding sites as the top 'hit' in only 2% of cases, compared with 58% of cases by the equivalent PSSM for a diverse set of experimentally characterised binding sites (see Fig. 1).

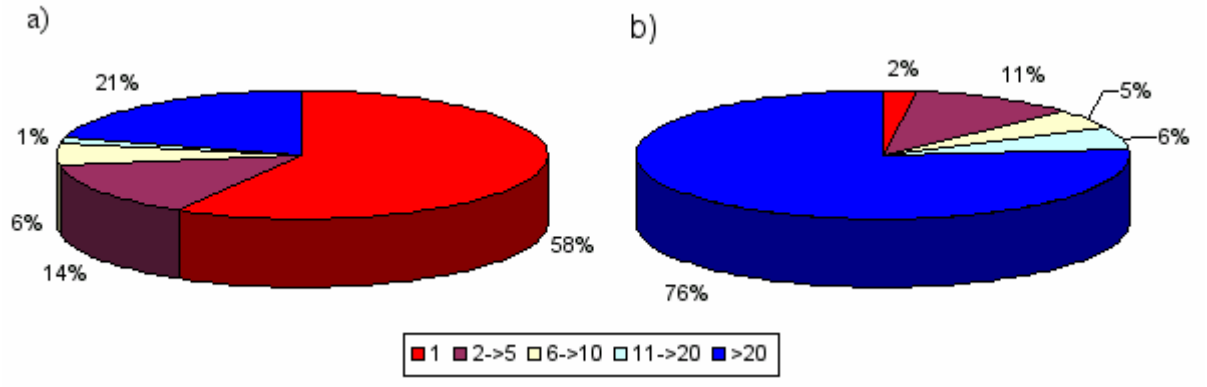


Fig. 1. Binding site predictions by a) the PSSM model and b) the SDM model. Correctly predicted binding sites as the top 'hit' appear in red, correctly predicted binding sites in the top 2 to 5 'hits' appear in maroon, those in the top 6 to 10 appear in yellow, those in the top 11 to 10 appear in light blue and finally binding sites predicted outside the top 20 'hits' appear in blue.

Our findings suggest that, while there is clearly some information to be obtained from analysis of these intermolecular interactions, application at the amino acid–DNA base level to a matrix-type model is much worse than PSSM models currently available. Indeed, although PSSM models have their limitations, they do perform very well on short sections of DNA sequence when representing a well defined binding site.

Recently we shifted our broad spectrum analysis and focussed our attention on the GATA family of transcription factors in *Arabidopsis thaliana*. The members of this class IV zinc finger protein are typified by the GATA motif that they selectively bind. PSSMs representative of this DNA recognition site are available for mammalian systems, however consideration of the length of sequence to be searched in *A. thaliana*, coupled with the abbreviated nature of the motif, mean that binding site predictions made with the PSSM model could never achieve statistical significance.

Consequently an alternative approach was taken, which involved investigation of the DNA binding domain of the *A. thaliana* GATA factors, using multiple sequence alignment and homology modelling techniques. Our findings suggested that these proteins interact with their cognate DNA in fundamentally the same way at the molecular level.

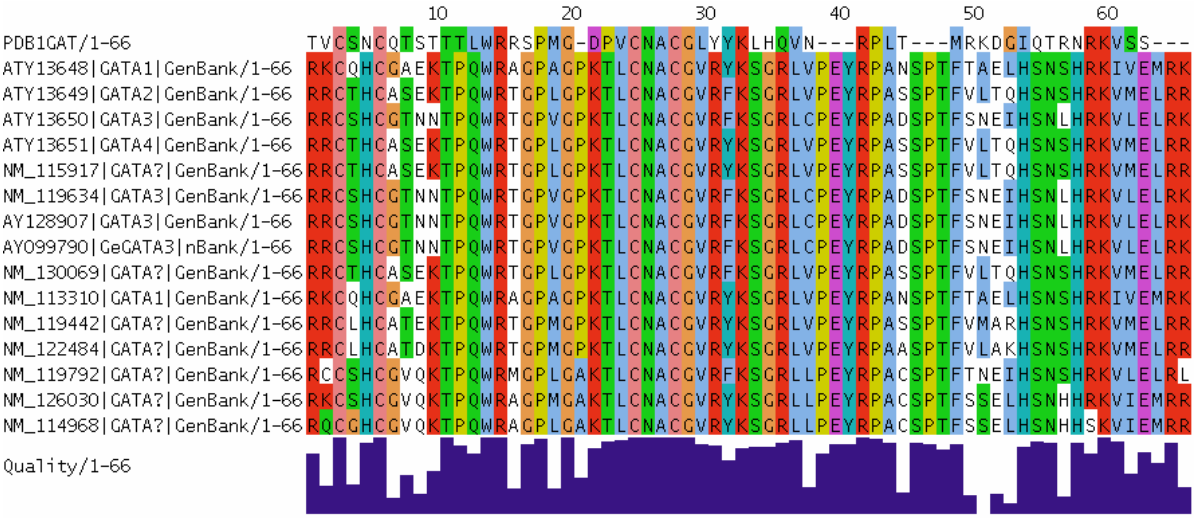


Fig. 2. Multiple sequence alignment of *Arabidopsis thaliana* sequences aligned against a mammalian template. The domain selected from the mammalian sequence is the zinc finger DNA binding domain as found in the 1gat PDB structure.

Our most recent efforts have involved developing a molecular mechanics model which will allow us to study the energy terms present in protein-DNA interactions. Using a modified version of the MultiDock interface refinement tool, we are currently investigating the components of these interactions to identify key features that are important to binding and recognition.

Collaborators
Professor P. M. Gilmartin, Centre for Plant Science, University of Leeds, UK.

Funding
This work was supported by the BBSRC.

Searchable database containing comparisons of ligand binding sites at the molecular level for the discovery of similarities in protein function

Nicola Gold and Richard Jackson

Structural genomics projects produce large amounts of data of which some are solved structures of hypothetical proteins of unknown function. The aim of this project is to aid the characterisation of these proteins by structure based prediction of protein function based on common modes of molecular recognition.

The current project extends previous work which demonstrated a method based on geometric hashing to compare the structures and properties of ligand binding sites and assess the extent of their similarity. In particular the binding site of the phosphate moiety of the large class of nucleotide ligands (ATP/ADP, GTP/GDP, FAD, NAD) was studied and is now extended to include the entire ligand binding sites of these proteins. The current project uses geometric hashing (described previously) to give a similarity score, a superposition, RMSD and equivalenced atoms for each pair of compared binding sites. These data are stored in a World Wide Web accessible database which is searchable with a PDB code and ligand information (such as ligand name, number and chain). Submission of these data rapidly returns a ranked list of similar ligand binding sites with the most similar at the top. Each hit is coloured according to its similarity to the query's overall fold and SCOP family. Interesting hits can then be selected and superimposed on the query allowing further examination and visualisation with molecular graphics packages (Fig. 1). A multiple alignment of structurally equivalenced atoms is also provided.

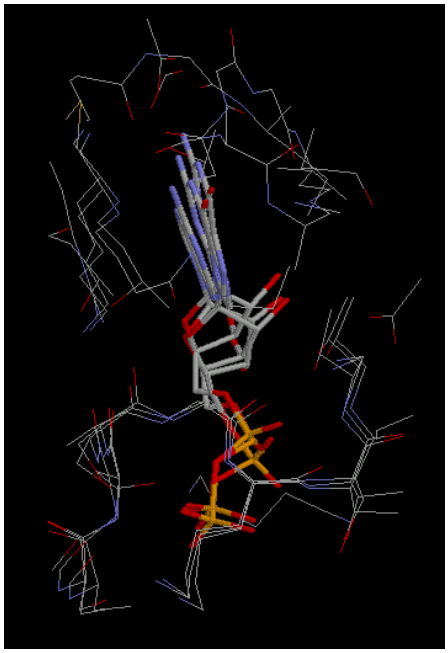


Fig. 1. Superposition of the GDP binding site of Elongation factor Tu protein (1dg1), the GDP binding site of signal sequence recognition protein Ffh (1ng1) and the ADP binding site of phosphor-enolpyruvate carboxykinase (1k3c). 1dg1 and 1ng1 are from different SCOP families sharing the same superfamily and fold, whereas 1k3c is classified in a different superfamily and fold group.

A frequency distribution of scores in the database gives two score thresholds which are used in a similarity matrix (Fig. 2). Here, strong similarity (score>39) is indicated by a black dot and weaker, but still significant similarity (score 25-39), is indicated with a red dot. Binding sites have been ordered along the axes by their evolutionary relationships *i.e.* close evolutionary relatives are adjacent; therefore similarity between close family members is displayed along the diagonal. This representation allows us to detect binding site similarity in the absence of sequence or fold similarity (off the diagonal). Similarity can, for example, be found between the binding sites of Elongation factor Tu and phosphoenolpyruvate carboxykinase (Fig. 1) which have different overall folds but share a common structural P-loop.

Future work will see this method and database extended to include ligand binding sites from other proteins. This database and comparison method can then be used to discover new similarities indicating potential functional relationships between proteins and may uncover binding site similarities in proteins previously thought to be unrelated.

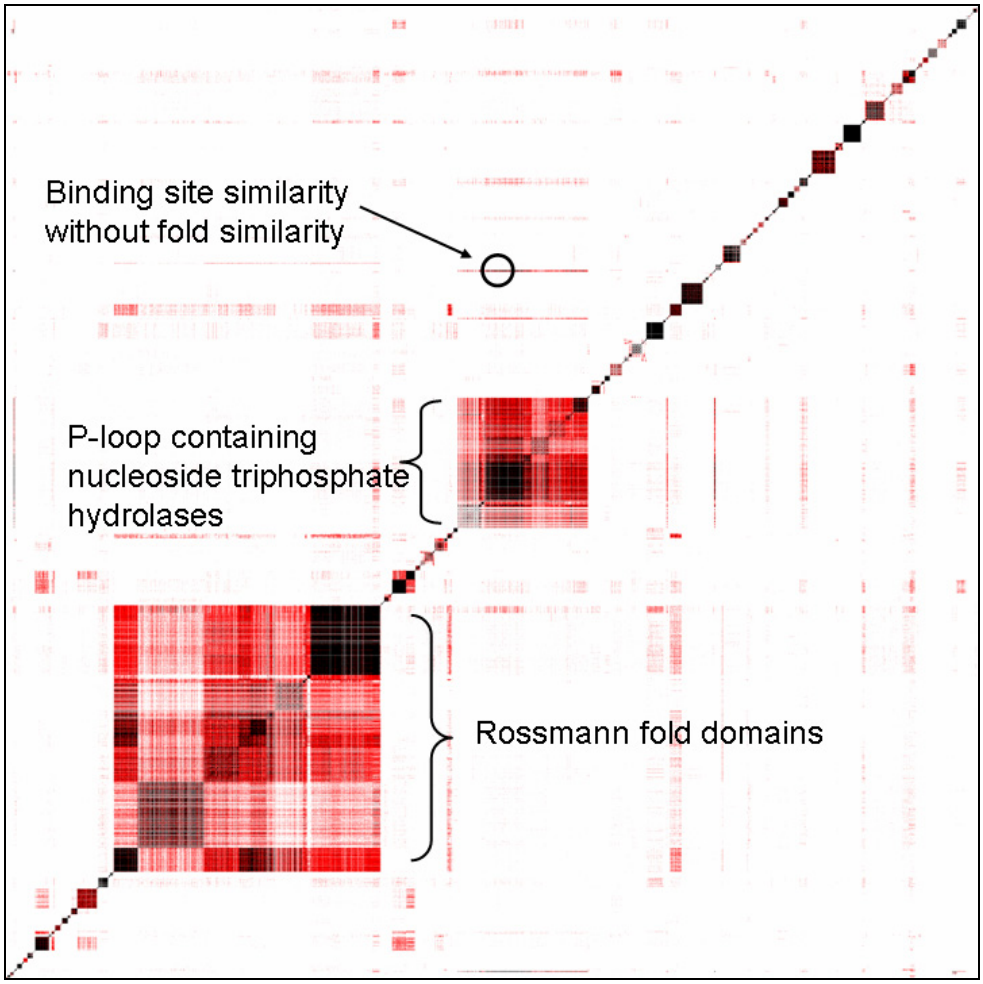


Fig. 2. Similarity matrix from the database of nucleotide binding sites. Strong similarity (score>39) is indicated by a black dot and weaker but still significant similarity (score 25-39) is indicated with a red dot. Binding sites have been ordered along the axes by their evolutionary relationships *i.e.* close evolutionary relatives are adjacent in the matrix. Evolutionary similarity is therefore displayed along the diagonal. Similarity away from the diagonal identifies similarity between proteins which share little or no fold similarity.

Publication

Brakoulias, A. and Jackson, R.M. (2004) Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching. *Proteins; Structure function and bioinformatics,* **56,** 250-260

Funding

We wish to acknowledge the support of the BBSRC.

Flexligdock: A flexible ligand – protein docking tool

Peter R. Oledzki and Richard M. Jackson

The small molecule docking algorithm Flexligdock is being developed from QFIT into a comprehensive flexible ligand docking tool for lead compound identification in virtual screening. The program can predict experimental ligand binding modes (Fig. 1) through biomolecular-ligand interactions using a probabilistic sampling method in conjunction with a molecular mechanics force field.

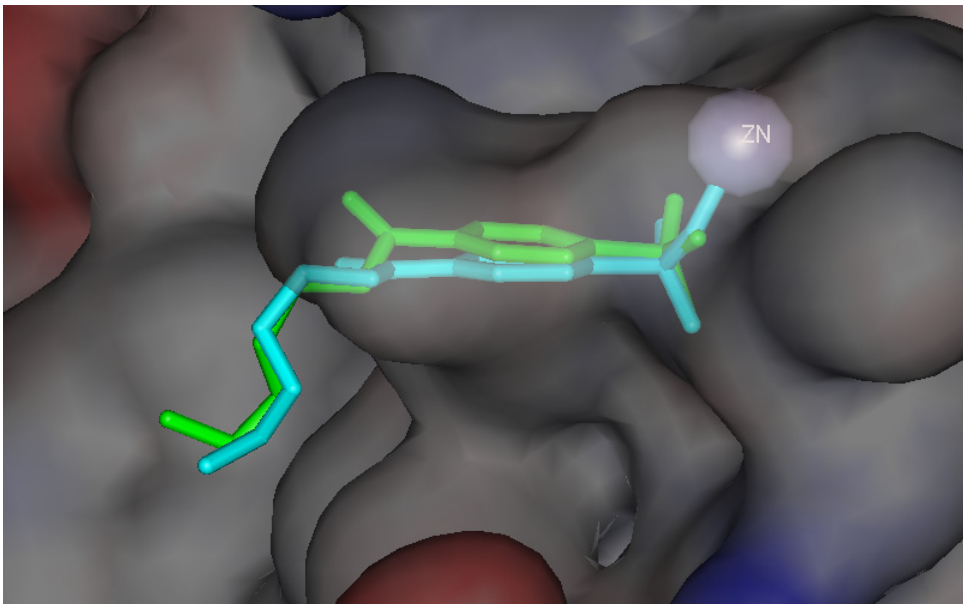


Fig. 1: Ligand conformations of 4-sulphonamide-[1-(4-aminobutane)]benzamide in complex with carbonic anhydrase II. The experimental crystal conformation is shown in light blue and the successfully docked solution is shown in green.

The method fragments a potential ligand along each torsion bond and utilises an interaction point methodology to map the ligand onto an interaction energy grid map of a protein target (Fig. 2). The method proceeds to reassemble/dock the ligand within the protein binding site with an incremental construction method. Predefined torsion angles are used to sample vast areas of chemical space for low energy ligand conformations.

The algorithm has been parameterised on a data set of 46 protein-ligand complexes obtained from a recently released docking data set. The parameterisation data set contained a structurally diverse set of proteins and a variety of ligands that contained 0-23 torsion bonds.

The FlexX validation data set of 200 protein-ligand complexes has been docked with Flexligdock to permit comparison against other existing protein-ligand docking algorithms. The FlexX docking algorithm docks 46% of the data set <2Å RMSD as the top ranked solution, whereas Flexligdock docks 50%. When the entire ranked solution set is considered FlexX docks 70% of the dataset <2Å RMSD, whereas Flexligdock successfully docks 74%. Currently, further improvements to the docking algorithm are being undertaken to increase the accuracy of Flexligdock.

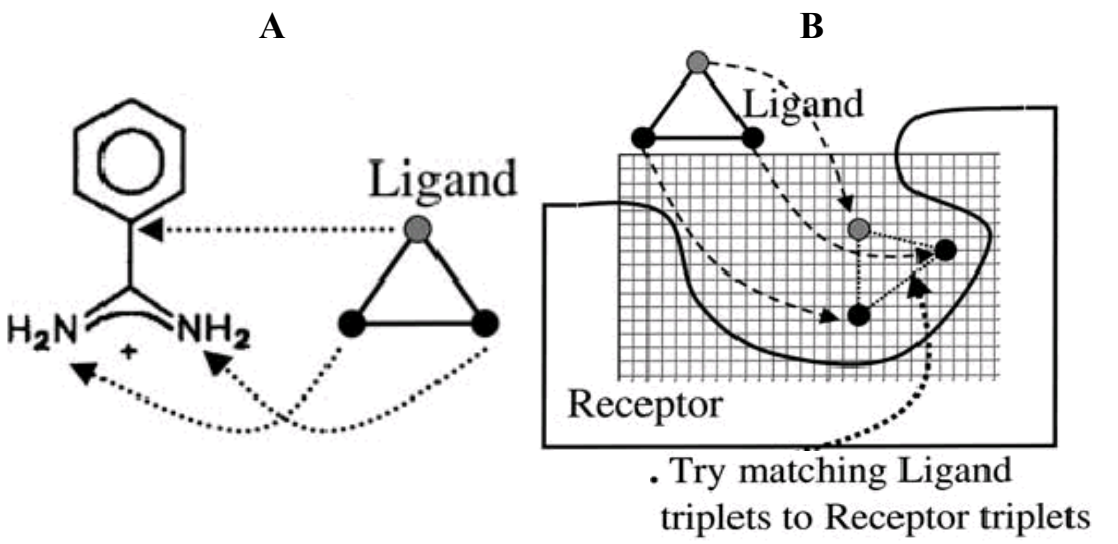


Fig. 2: A) Ligands can be represented as a triplet by three functional groups. B) Ligand triplets are matched to the receptor triplets and ranked to find the best conformations.

Docking studies on a set of protein kinase structures was undertaken in collaboration with Dr Adam Nelson from the Chemistry department. A set of 20 low energy ligand conformers were obtained from Dr Nelson and docked to PDK1 and GSK-3β protein structures. The ligands were a set of conformationally diverse macrocyclic bisindolylmaleimides, containing enantiomeric ligands that varied in their size of tether linker.

The results obtained provided supplementary computational evidence that a specific type of enantiomeric conformation of the ligands was more favourable than the other. This had been originally concluded from Dr Nelson's binding affinity data, which was unknown at the time of the docking.

Funding

We wish to acknowledge support of BBSRC.

Predicting protein-protein interactions

Nicholas Burgoyne and Richard Jackson

Now that both the individual protein structures and certain aspects of their molecular biology can be determined at a genomic scale there is a growing knowledge gap emerging in biochemistry. Although we often know the structures of two interacting proteins, the structure of them in complex is still difficult to determine. Predictions of the interface can be very instructive for a molecular biologist, by guiding experimental analysis of the complex. Predictions of the structure of a complex can also be useful in the structure determination processes of protein complexes by NMR, X-ray crystallography and electron microscopy.

Predictions of protein-protein interfaces are very useful, but the process is difficult. This is primarily due to the small set of structural examples that are currently known. Just as it is assumed that there are a given number of protein folds, there is likely to be a defined number of interactions between them. The known protein interfaces show great diversity in terms of the size of the buried interface and the chemical composition of their binding surfaces. Despite this fact, there are sufficient similarities to predict protein interfaces by taking the mean interface-like properties from known complexes.

It may also be possible to concentrate the prediction on a certain subset of the entire protein-protein interface. It is known that most protein-interfaces contain clefts, into which are placed the sidechains of its interacting protein partner. Alanine-scanning mutagenesis and evolutionary conservation of residues around the clefts suggests that some residues are more important to the interaction than others. It is interesting to note that most interfaces have these important clefts, and that their properties are largely consistent across the known interfaces.

It is our intention to attempt to predict these interface-clefts from a protein surface by relying on the known properties that give the clefts the ability to bind small regions of their interaction partners. This would indicate which regions of the surface are most-likely to interact with the sidechains of the binding protein. Experimental evidence (mentioned previously) indicates that there can often be more than one of these functionally important sites in a given complex, it is also known that these sites often do not alter conformation on formation of the protein complex. Therefore the prediction of the relevant clefts could act as a basis for the generation of possible protein-protein interactions.

This work is ongoing.

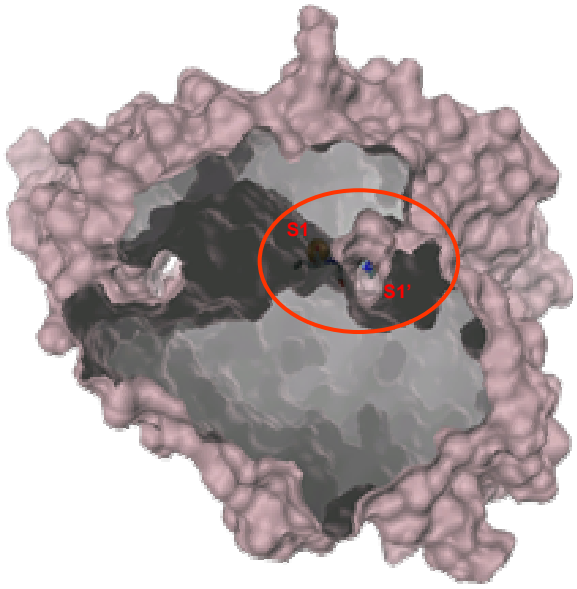
Funding
This work is funded by the MRC.

Strategies for ACE2 structure-based inhibitor design

Monika Rella and Richard Jackson

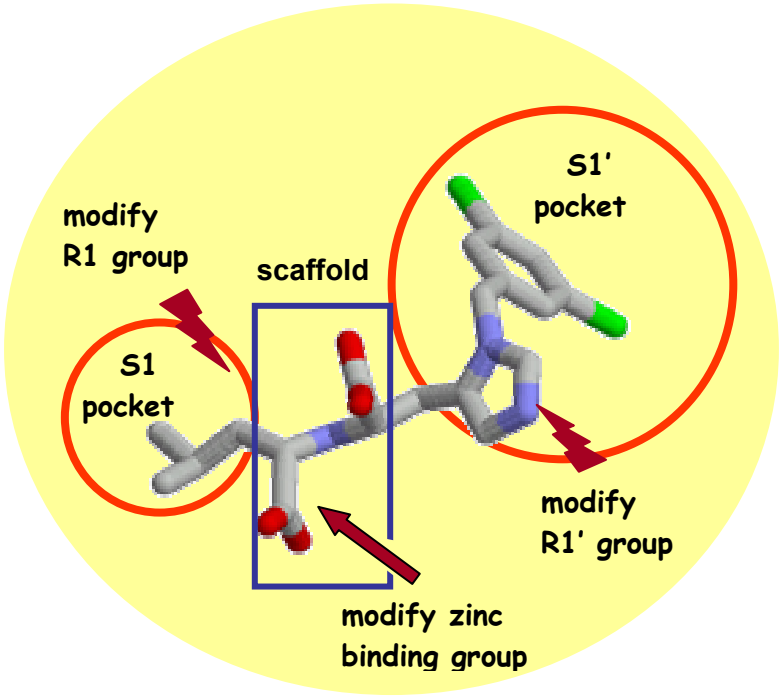
Angiotensin-Converting Enzyme (ACE) is an important drug target for hypertension and heart disease. Recently, a unique human ACE homologue termed ACE2 has been identified in the School of Biochemistry and Microbiology, which has been linked to hypertension, heart and kidney disease. In addition, ACE2 was shown to function as SARS-Coronavirus receptor. This surprising role and its assumed counter-regulatory function to ACE make ACE2 an interesting new cardio-renal disease target.

With the recently resolved ACE2 structure in complex with an inhibitor available, a structure-based drug design project has been undertaken to identify novel potent and selective inhibitors. The general long-term strategy comprises computational structure-based drug design approaches, as well as chemical synthesis of promising candidates, or purchase of existing compounds and bioassay-based potency evaluation by our collaborators. Computational approaches involve combinatorial library design and docking as well as pharmacophore-based virtual screening of large compound databases.



Cross-sectional view of the ACE2 substrate-binding site in form of a channel that stretches the whole protein. The protein is displayed as sliced surface to exhibit the interior and the translucent channel which is formed upon inhibitor binding. The inhibitor is sitting in the narrow bottleneck connecting the left (S1) and right (S1') active site pockets. The size of S1 is limited to the side chain of the inhibitor, whereas S1' is more spacious. The zinc metal ion is shown in spacefill.

A small number of synthetically accessible fragments were selected and individually evaluated for improved interaction energy using our in-house rigid-body docking tool Q-fit. In a second step, the *de novo* design tool, SynSPROUT, was applied for replacing the original side chain of the core structure with each fragment, and docking the whole molecule. Currently, we are preparing a combinatorial library on a larger scale by mimicking a synthetic reaction that involves synthetic starting material such as boronic acid derivatives and any naturally occurring amino acid. R-groups suitable for chemical synthesis will be assessed and prioritised via docking as described above.



Scaffold-based design strategy for new ACE2 inhibitors using the bioactive conformation of the ACE2 inhibitor. A combinatorial library is designed by replacing both R-groups on a common scaffold with suitable synthetic starting material that should satisfy both active site pockets (S1, S1'). In addition, alternative zinc binding groups can be screened for.

In a complementary approach, a protein-based pharmacophore model was created manually comprising several chemical features such as hydrogen bonding, electrostatic and hydrophobic interactions aligned in 3D, resembling specific drug-receptor interactions. Selectivity of the model was ensured by initial screening for ACE inhibitors and enrichment enhanced through repeated optimisation cycles. The final model is currently used to search 2.5 million unique compounds for matching features. Hits will further be evaluated and prioritised via docking and the most promising candidates proposed for purchase and biological testing.

Collaborators

R. Grigg, School of Chemistry, University of Leeds.
A.P. Johnson, School of Chemistry, University of Leeds.
T. Langer, Department of Pharmaceutical Chemistry, University of Innsbruck, Austria
A.J. Turner, School of Biochemistry and Microbiology, University of Leeds.

Funding

The author wishes to acknowledge support for this work by the University of Leeds Research Scholarship.

NMR facility

Arnout Kalverda and Steve Homans

Overview of facility
The NMR facility is equipped with a 750 MHz, a 600 MHz and two 500 MHz Varian Inova NMR spectrometers. All instruments are setup to use ^{1}H, ^{13}C, ^{15}N and ^{2}H during normal operation. In the spring of 2005, a cryoprobe will be added to the 750 MHz NMR spectrometer that will enhance the sensitivity of the system by up to a factor of 3.

Dynamics and thermodynamics of ligand binding
The binding of a ligand to a protein often results in changes in the dynamics of the protein upon binding. These dynamic changes are reflected in the thermodynamics of binding through their contribution to the entropy of the system. Using Mouse Urinary Protein, changes in backbone and side-chain dynamics have been probed with ^{15}N and ^{2}H relaxation methods. Comparisons were made between the apo-protein and two closely related ligands. The changes in dynamics upon ligand binding indicate a complex network of dynamic compensation. Whereas methyl groups that are located in the binding pocket become more rigid, other side-chains located in a shell further away from the ligand binding site become more mobile. This hints that there is a shift in the most prevalent dynamic modes upon ligand binding. Measurements of the changes in dynamics are being continued with site directed mutants of the protein.

The interplay between dynamics and thermodynamics in ligand binding is also being studied in a second system: arabinose binding protein which binds to galactose and deoxy-derivatives of galactose. As a step towards using NMR measurements with this protein, the sequential assignments have been completed and NMR relaxation studies are underway.

Identification of a methyl-thio sugar substituent in the *M. tuberculosis* cellwall
Lipoarabinomannan is a major component of the *Mycobacterium tuberculosis* cell wall. An unusual sugar substituent has been found as part of the mannosyl capping structure on Lipoarabinomannan. Using a combination of NMR, mass spectrometry and chemical synthesis, this sugar substituent was determined to be a 5-deoxy-5-methylthio-α-xylofuranosyl structure. It is reasonable to assume that it may be biosynthetically derived from 5'-methyl-thioadenosine, which is a byproduct of polyamine biosynthesis. The presence of this unusual sugar may provide an angle on potential new targets for the development of antituberculosis drugs.

NMR of protein folding
In collaboration with the group of Prof. S.E. Radford, native state hydrogen exchange has been used with the Colicin Immunity protein Im7 to show that the hydrogen exchange data provide information on the secondary structure of an intermediate state. To demonstrate whether hydrogen exchange occurred from the intermediate state, local fluctuations or global unfolding, data were compared between the native protein and a mutant (I72V). The mutation significantly destabilises the intermediate state relative to the unfolded state, but only slightly destabilises the native state. The hydrogen exchange rates reflect the free energy difference with the state from which exchange occurs, and the hydrogen exchange patterns shift with the changes in free energy upon mutation. Thus residues that exchange from the intermediate have decreased hydrogen exchange rates, while residues that exchange through global unfolding have increased hydrogen exchange rates in this case. This method, using site directed mutagenesis to identify the state from which hydrogen exchange occurs,

could have more widespread use in identifying the presence of secondary structure in intermediate states.

Publications

Homans, S.W. (2004) NMR spectroscopy tools for structure aided drug design. *Angewandte Chemie* **43**, 290-300.

Bingham, R.J., Findlay, J.B.C., Hsieh, S.Y., Kalverda, A.P., Kjellberg, A., Perazzolo, C., Philips, S.E.V., Seshadri, K., Trinh, C.H., Turnbull, W.B., Bodenhausen, G. and Homans, S.W. (2004) Thermodynamics of binding of 2-methoxy-3-isopropyl pyrazine and 2-methoxy-3-isobutyl pyrazine to the major urinary protein. *J. Am. Chem. Soc.*. **126**, 1675-1681.

Daranas, A.H., Kalverda, A.P., Chiovitti, A. and Homans, S.W. (2004) Backbone resonance assignments of the L-arabinose binding protein in complex with D-galactose. *J. Biomol. NMR* **28**, 191-192.

Gorski, S.A., Leduff, C.S., Capaldi, A.P., Kalverda, A.P., Beddard, G.S., Moore, G.R. and Radford, S.E. (2004) Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7. *J. Mol. Biol.* **337**, 183-193.

Turnbull, W.B., Shimizu, K.H., Chatterjee, D., Homans, S.W. and Treumann, A. (2004) Identification of the 5-methylthiopentosyl substituent in *Mycobacterium tuberculosis* lipoarabinomannan. *Angewandte Chemie* **43**, 3918-3922.

Funding

We thank the Wellcome Trust and BBSRC for funding.

Dynamics in the unfolded state of β_2-microglobulin

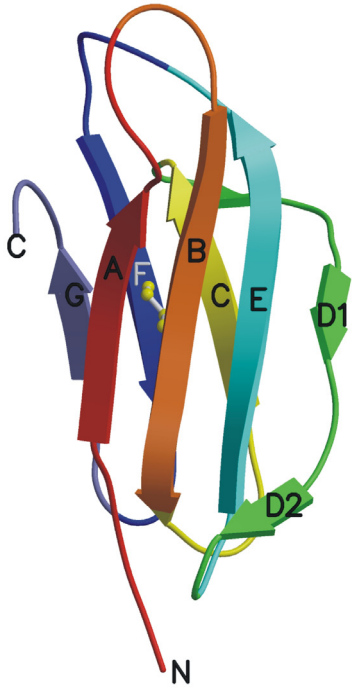
Geoffrey Platt, Susan Jones, Clemens Stilling, Thomas Jahn, Katy Routledge, Arnout Kalverda, Steve Homans and Sheena Radford

Introduction

There are approximately twenty proteins with unrelated amino acid sequences and native structures that aggregate to form highly ordered amyloid fibrils *in vivo*, and result in specific disease states. More recently, several of these proteins, and some that are not disease-related, have been shown to form amyloid *in vitro* by manipulation of solution conditions. In all cases, the normally soluble, monomeric proteins deposit as insoluble fibres with characteristic cross-β structure. Our research aims to elucidate the mechanism of fibril formation for human β_2-microglobulin (β_2m) (Fig. 1), which causes haemodialysis related amyloidosis in all patients with renal failure.

Dynamic properties of acid unfolded β_2m

In acidic conditions (below pH 4.0) *in vitro* β_2m spontaneously self-assembles into amyloid-like fibrils with different morphologies that are dependent on the pH and ionic strength of the solution. At pH 2.5 monomeric β_2m is highly unfolded and the protein assembles with nucleation-dependent kinetics to form fibrils that are long and straight. Structural characterisation of amyloid-ogenic precursor conformations is vital to further our understanding of the mechanisms of fibril formation. We have used NMR relaxation methods (Fig. 2) to explore the structural and dynamic properties of acid-unfolded, wild-type β_2m, as well as two β-strand E variants (Y66E and F62AY63AY67A) engineered to investigate the role of aromatic ring stacking in determining the properties of the unfolded state. The results show that the acid denatured ensemble of β_2m contains significant and substantial non-native structure, stabilised by cooperative interactions between aromatic residues, as well as the presence of a single disulphide bond. Furthermore, relaxation-dispersion NMR experiments reveal that the acid unfolded ensemble of β_2m involves two or more distinct species in dynamic equilibrium. Replacement of aromatic residues in the variants greatly disrupts residual structure as seen in acid denatured F62AY63AY67A β_2m, which resembles a highly unfolded state. The results support the view that the clustering of aromatic residues is a common feature of the unfolded states of many proteins.



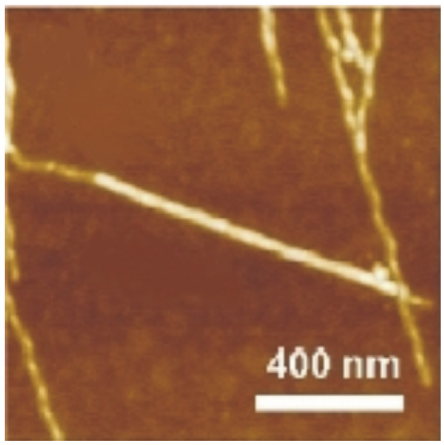


Fig. 1. Upper panel: Ribbon diagram of β_2m. Lower panel: Atomic force microscopy image of β_2m fibrils formed at pH 2.5.

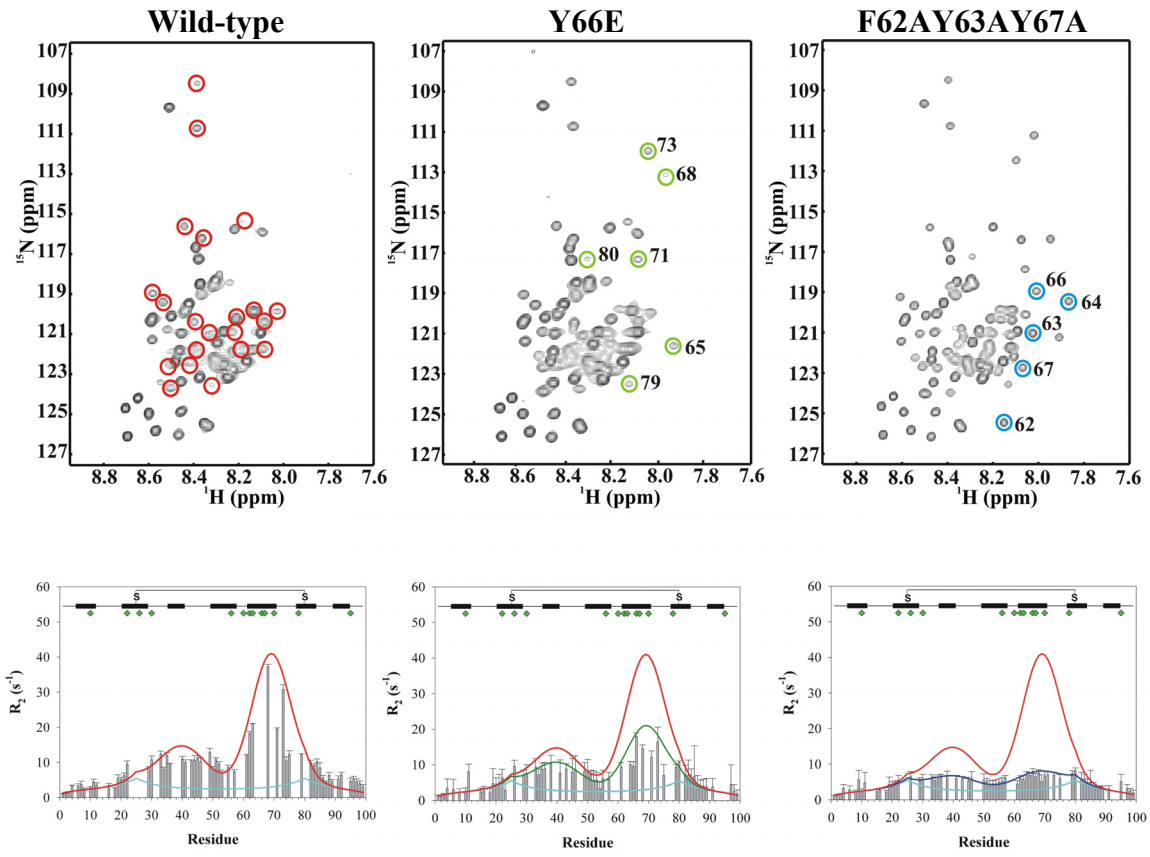


Fig. 2. ^{1}H-^{15}N HSQC NMR spectra of wild-type β_2m and variants at pH 2.5 (Top panels). Relaxation rates for β_2m and variants at pH 2.5 on a per-residue basis that indicates increased dynamics on replacement of aromatic residues, especially in the region corresponding to native β-strand E (Lower panels).

Fibril formation by variant β_2m proteins

Correlating structure and dynamics information with the amyloidogenicity of β_2m is an important step in understanding the mechanism of fibril formation. We are currently analysing the role of the β-strand E of β_2m, a region that is rich in aromatic residues and has previously been implicated in fibril formation. We are using a variety of spectroscopic methods such as Thioflavin-T fluorescence, circular dichroism and electron microscopy in conjunction with NMR techniques to study the rate of fibril formation and morphological features of β-strand E mutants at a range of pH values.

Collaborators

We thank Victoria McParland, currently working at the National Cancer Institute, NIH (USA), for her contribution to this work.

Publications

Platt, G.W., McParland V.J., Kalverda, A.P., Homans, S.W. and Radford, S.E. (2005) Dynamics in the unfolded state of β_2-microglobulin studied by NMR. *J. Mol. Biol.* **346**, 279-294.

Funding

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

Ultra-rapid folding of the B domain of *Staphylococcal* protein A

George Dimitriadis, Graham Spence, Jennifer Clark, Daniel Lund,
Sheena Radford and Alastair Smith

The three-helix bundle B domain of protein A (BdpA, Fig. 1) has been the protein of choice for a number of theoretical, and more recently, experimental protein folding studies because of its small size and simple topology. Although theoretical studies suggest the fast folding of BdpA follows a diffusion-collision-like mechanism, in which pre-existing structural elements that persist in the denatured state coalesce, developing cooperative interactions that stabilise the native state, differences exist in different models in the importance of individual structural elements in the folding process.

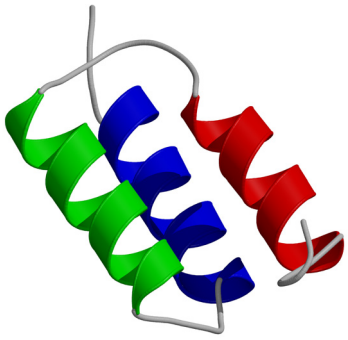


Fig. 1. The structure of the B domain of *Staphylococcal* protein A. Helix 1 is shown in red, helix 2 in green and helix 3 in blue. Drawn using Molscript and Raster3D.

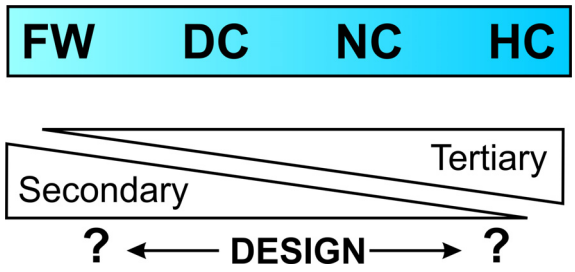


Fig. 2. A continuum of folding mechanisms, from framework (FW), through diffusion-collision (DC) and nucleation-condensation (NC), to hydrophobic collapse (HC), and the relative importance of secondary and tertiary interactions for these models.

The diffusion-collision model is supported by work from the Oas and Weaver groups, which has shown that the experimentally determined folding rate constant is consistent with that predicted by the diffusion-collision model (Myers, 2001; Islam, 2002). However, a further study of BdpA using phi-value analysis by the Fersht group (Sato, 2004) suggests that folding occurs *via* a nucleation-condensation-like mechanism, capturing further aspects of the theoretical studies, but leaving open important questions about how BdpA folds to its native state. These findings highlight the fact that proteins do not fold by discrete mechanisms, but rather through an infinite variety of mechanisms determined by the relative importance of secondary and tertiary interactions (Fig. 2).

Development of a state-of-the-art temperature jump apparatus has led to a fully computer controlled instrument capable of producing temperature jumps of up to 25°C within 8 ns, able to detect changes in the sample fluorescence with a dead time of better than 130 ns, and capable of directly measuring rate constants on the order of 150000 s^{-1}. In combination with the measurement of equilibrium free energy surfaces, which follow the population of species with respect to denaturant concentration and temperature, the microscopic rate constants for folding and unfolding can be determined over a wide range of final conditions. The combination of this instrument and protein model allow exciting and fundamental questions about protein folding to be addressed.

Current Work

Using a combination of the laser induced temperature jump apparatus and 10 selected single amino acid variants, the predictive ability of a general diffusion-collision model has been investigated. Seven of the ten mutants were designed to alter the propensity of the three

helices of BdpA by either mutating residues within the helices to alanine (increasing helical propensity, but decreasing the burial of hydrophobic surface area) or glycine (decreasing helical propensity and burial of hydrophobic surface area), or by replacing residues that form an N-terminal hydrophobic helix cap. The three remaining mutants each removed specific tertiary contacts from the hydrophobic core of the protein. The results showed that unless the change in helical propensity is dramatic, *i.e.* mutations to glycine residues, diffusion-collision is unable to predict the observed rate constants, necessitating the development of a refined model.

Analysis of theoretical and experimental evidence suggests that BdpA folds *via* a mechanism that that lies somewhere between diffusion-collision and nucleation-condensation. We are currently investigating whether it is possible to use rational design to alter the relative importance of secondary and tertiary interactions during BdpA folding, therefore altering the mechanism by which this protein folds (Fig. 2). Two variants have been designed with the aid of AGADIR (Lacroix, 1998), an algorithm that predicts the helical propensity of a sequence. The first of these has been designed to have as little intrinsic helical propensity as possible, whilst not disrupting the native helical interfaces, the second having been designed to have a high degree of secondary structure present in the in the denatured ensemble. Characterisation of these variants is underway.

BdpA is also being developed as a system to study protein folding and dynamics at the single molecule level, using fluorescence resonance energy transfer (FRET) between the dyes AlexaFluor 488 and AlexaFluor 594. Variants have been designed and purified that contain a pair of cysteine residues at sites predicted to give a large change in FRET as the protein folds and unfolds. Currently, the labelling strategy is being optimised to give the largest yield of protein labelled with both dyes.

Future Work

Perhaps the most exciting possibility is to try and engineer BdpA to fold in the "downhill" regime, where the rate of folding is under diffusion control. This would, for the first time, provide an experimental measure of the pre-exponential factor, which defines the fastest rate at which a protein can fold. This could be achieved by using knowledge from phi-value analysis to specifically stabilise the rate-limiting transition state. Other projects include using a diffusion-collision model to predict the changes in rate constants caused by increasing linker lengths between helices, or by making changes to the propensity of individual helices.

Publications

Dimitridis, G., Drysdale, A., Myers, J. K., Arora, P., Radford, S. E., Oas, T. G. and Smith, D. A. (2004) Microsecond folding dynamics of the F13W G29A mutant of the B domain of Staphylococcal protein A by laser induced temperature jump. *Proc. Nat. Acad. Sciences USA*. **101**, 3809 – 3814.

Collaboration

The study of the B domain mutants is in collaboration with Terrance Oas and Pooja Arora at the Department of Biochemistry, Duke University, North Carolina.

Funding

This work was funded by the BBSRC, EPSRC and the University of Leeds. Sheena Radford is a BBSRC Professorial Fellow.

Tailoring the folding mechanisms of the four helix bundle proteins Im7 and Im9

Stuart Knowling, Claire Friel, Victoria Morton, Graham Spence, Eva Sanchez-Cobos, Susanne Cranz-Mileva and Sheena Radford

Introduction

The explosion in available genomics data over the last decade has highlighted the great value of the ability to determine, directly from its primary sequence, the native structure to which a protein folds. Also, the increase in diseases known to result from protein misfolding has made it imperative that a more detailed understanding of the mechanisms by which a protein folds to its native-state is developed and key steps identified.

Im7 and Im9 present a powerful opportunity to study the mechanisms of protein folding. These proteins have the same four helical fold and 60% sequence homology. Despite their similarity in structure and sequence, however, Im7 and Im9 fold with different kinetic mechanisms *in vitro* (at 10 °C, pH 7.0); Im7 folding with a three-state mechanism, in which an on-pathway intermediate is populated early during folding, whilst Im9 folds with a two-state transition without populating an intermediate.

Switching the folding mechanism of Im9 from a two-state to a three-state mechanism

Since Im7 and Im9 have such similar structures and sequences, the question of why Im9 does not populate an intermediate during folding is a fascinating one. Previous structural characterisation of the mechanism of Im7 folding using Φ-value analysis has shown that the intermediate populated early during folding of this protein contains helices I, II and IV, but lacks helix III. Importantly, this analysis also highlighted several residues which stabilise the intermediate in Im7 folding through the formation of non-native contacts. Interestingly, several of the residues at these key positions are substituted with less hydrophobic amino acids in Im9, raising the possibility that a three helical intermediate may be formed during Im9 folding, but may be too unstable to detect kinetically. To test this hypothesis, several mutations were made in the Im9 sequence, designed using available knowledge of the structure of the intermediate formed during Im7 folding. In each case the residue at each chosen site in Im9 was substituted with the equivalent residue in Im7.

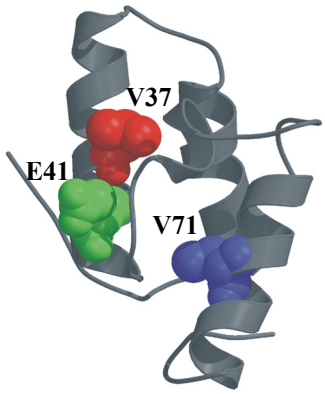


Fig. 1. Ribbon diagram of the structure of native Im9 showing the mutations made to switch the folding mechanism of this protein from two- to three-state.

Excitingly, combining three of these mutations in the triple mutant V37L/E41V/V71I resulted in an immunity protein with a native conformation destabilised only slightly relative to wild-type Im9 ($\Delta\Delta G_{un}$ = +4 kJ/mol), but which folded through a stable and highly populated intermediate ($\Delta\Delta G_{ui}$ > -11.5 kJ/mol) (at pH 7.0, 10 ºC). A brief Φ-value analysis confirmed that the intermediate formed during Im9 folding is structurally similar to the three helical species formed during Im7 folding. These data indicate, therefore, that misfolding on the pathway to the native-state is a common feature of immunity protein folding and, importantly, demonstrates that Im7 and Im9 fold with a similar structural mechanism, despite the differences in their kinetic mechanisms of folding.

Trapping the folding Intermediate of Im7 at equilibrium

A complete understanding of folding requires that all species on the reaction coordinate are identified and their structure, stability and folding/unfolding kinetics determined. A key step in our quest to understand the folding of the immunity proteins, therefore, is to define the structures of the unfolded, intermediate and native states of these proteins in as much detail as possible. The most powerful approach towards this goal is to use mutagenesis to trap non-native states at equilibrium, such that their structural properties can be determined directly using NMR. Commencing with Im7, we have taken this approach and have redesigned the sequence of the protein such that the kinetic intermediate is significantly populated at equilibrium, the end product of the folding reaction. This was achieved by making use of the fact that helix III plays little role in folding until after the rate-limiting transition state has been traversed. Altering side-chains in helix III thus destabilises the native protein significantly, but has little effect on the stability of the intermediate. As a consequence, the intermediate should become the ground state of the folding reaction (named here I^{eqm}). Several variants of Im7 were created by residue-specific mutation of helix III, or deletion of the 6 residues that comprise helix III in native Im7 and replacing these residues with either three or six glycines (in the variants H3G3 and H3G6, respectively) (Fig. 2). All four variants created trapped a species at equilibrium that resembles the kinetic intermediate. Perhaps most excitingly, these proteins give rise to well resolved NMR spectra, opening the door to analysis of their structural properties in detail using NMR methods.

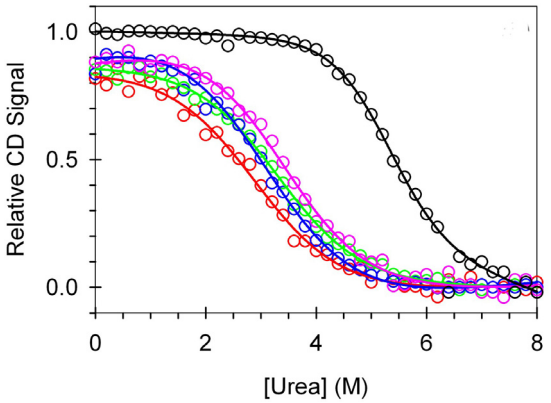


Fig. 2. Equilibrium urea denaturation of Im7* (black) and the variants L53AI54A (green); I54A (red); H3G3 (blue) and H3G6 (pink). Each dataset is fitted to a two-state transition (solid lines).

Future Work

The next year will be an exciting period with the culmination of a large amount of work currently underway. Studies on the role of helical propensity in the folding mechanism of Im9 have recently been completed and the role of Helix III in Im7 folding is now being studied by rational redesign of the helical sequence TDLIYY. The helical propensity of Im7 and Im9 are also being studied using peptide synthesis, as a prelude to analysis of the folding of these proteins using diffusion-collision models. Further work is also being conducted on the mechanisms of the early steps in folding using temperature jump and ultra-rapid mixing experiments, combined with an analysis of the role of the intermediate in tailoring the landscape for folding using the variants described above.

Collaborators

Martin Karplus, Harvard and Strasbourg Universities; Colin Kleanthous, University of York; Geoff Moore and Sara Whittaker, University of East Anglia; Emanuele Paci, University of Leeds; Michele Vendruscolo, Cambridge University

Publications

Paci, E., Friel, C.T., Lindorff-Larsen, K., Radford, S.E., Karplus, M. and Vendruscolo, M. (2004) Comparison of the transition state ensembles for folding of Im7 and Im9 determined using all-atom molecular dynamics simulations with Φ value restraints. *Proteins: Structure, Function and Genetics,* **54**, 513-525.

Gorski, S.A., Le Duff, C.S., Capaldi, A.P., Kalverda, A.P., Beddard, G.S., Moore, G.R. and Radford, S.E. (2004) Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of 1m7. *J. Mol. Biol.*, **337**, 183-193.

Friel, C.T., Beddard, G.S. and Radford, S.E. (2004) Switching two-state to three-state kinetics in the helical protein Im9 *via* the optimisation of stabilising non-native interactions by design. *J. Mol. Biol.* **342**, 261-273.

Spence, G.R., Capaldi, A.P. and Radford, S.E. (2004) Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J. Mol. Biol.* **341**, 215-226.

Funding
This work was funded by the EPSRC, BBSRC, FEBS, the Marie Curie Host Fellowship for Early Stage Training and The Wellcome Trust. We thank Keith Ainley for technical support.

Investigating the structure of monomeric and fibrillar β_2m by mass spectrometry

Andrew M. Smith, Toni J.H. Borysik, Sarah L. Myers, Rebecca J. Rose,
Sheena Radford and Alison E. Ashcroft.

Introduction.

β_2-microglobulin (β_2m) is a small protein of 11,860 Da, 99 amino acids in length, which readily forms amyloid-like fibrils in acidic conditions *in vitro*. *In vivo* β_2m is found within the plasma at low concentrations. However, in patients suffering from chronic kidney disease the concentration of β_2m can increase up to 60 fold, and this can lead to formation and deposition of amyloid fibrils within collagen rich areas of the body, such as the joints. At this time it is not known how β_2m forms amyloid fibrils *in vivo* at a more neutral pH where the protein is in its native state fold and is known to be stable. NMR and CD indicate that the protein is partially folded at ~ pH 3.6 and is largely unfolded at ~ pH 2.5 where the protein forms amyloid fibrils *in vitro*. Our group is investigating the structure of monomeric and fibrillar β_2m by mass spectrometry.

Charge state distributions of β_2m.

Electrospray ionization mass spectrometry (ESI-MS) can be used to identify different conformers of a protein, as unfolded species tend to carry on average more charges. By assigning Gaussian distributions to the charge state distribution (CSD) at different pH values it is possible to identify three different conformations of β_2m. The ratios of the three conformers vary depending on the pH with a more expanded conformer of the protein present at very low pH (pH 1.0-2.6), a partially expanded conformer present between pH 2.6 and 5.4, and a compact conformer at pH 5.4-6.0. These represent the acid unfolded, partially folded and native states of the protein, respectively. Two mutants that are known by other techniques to be globally destabilised were compared with wild-type β_2m using the charge state distributions one of which is shown in Fig. 1. The CSDs of all three proteins can be fitted to the same Gaussian distributions as the wild-type protein, although the pH at which different species become populated during pH denaturation was protein-specific. The mutants populate proportionally higher concentrations of partially and acid unfolded conformers at higher pH than the wild-type protein. The data provide a unique opportunity to delineate and quantify species co-populated in solution, one or more of which may be important in amyloid formation.

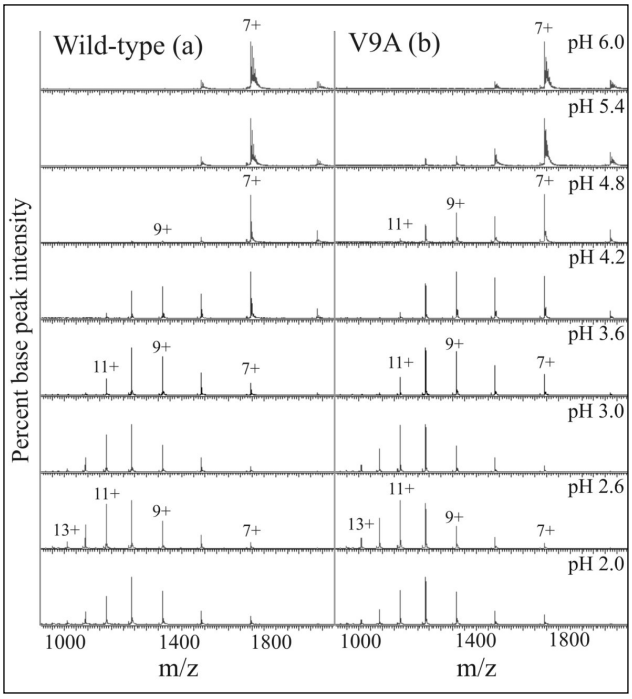


Fig. 1: Mass spectra taken at different pH values showing the shift in the CSD from a compact distribution centred on the 7 charge state at pH 6 to multiple CSD at lower pH. The transition between different protein conformers occurs at different pH for the two proteins. a) Wild-type β_2m, b) the mutant V9A β_2m[1].

Monitoring β_2m assembly into amyloid fibrils using mass spectrometry.
Although ESI-MS provides data with high mass accuracy, the technique is not usually used to quantitate protein concentrations, since it is not possible to relate the observed peak heights for a specific protein to its concentration in solution. However, by using an internal standard of known concentration, it is possible to create a calibration across a concentration range for a protein of interest, hence making the method semi-quantitative. In this study we have used a constant concentration of bradykinin to follow the concentration of monomeric β_2m under *in vitro* amyloid forming conditions at pH 3.6. This reveals an exponential decrease in monomer concentration over an 8 hour period until there is less than 20% of the original concentration remaining (Fig. 2a). It is possible to follow the appearance of β_2m oligomers as the monomer concentration decreases. The molecular weight of the oligomers detected suggests that the process by which amyloid forms under these conditions is by monomer addition, as a series of oligomers from dimer to dodecamer are detected (Fig. 2b). The population of these oligomers decreases at long times of assembly, as higher order oligomers and fibrils are formed. The data provide an exciting framework with which to determine the mechanism of amyloid formation under different conditions and to determine the mechanism of action of potential inhibitors of amyloidosis.

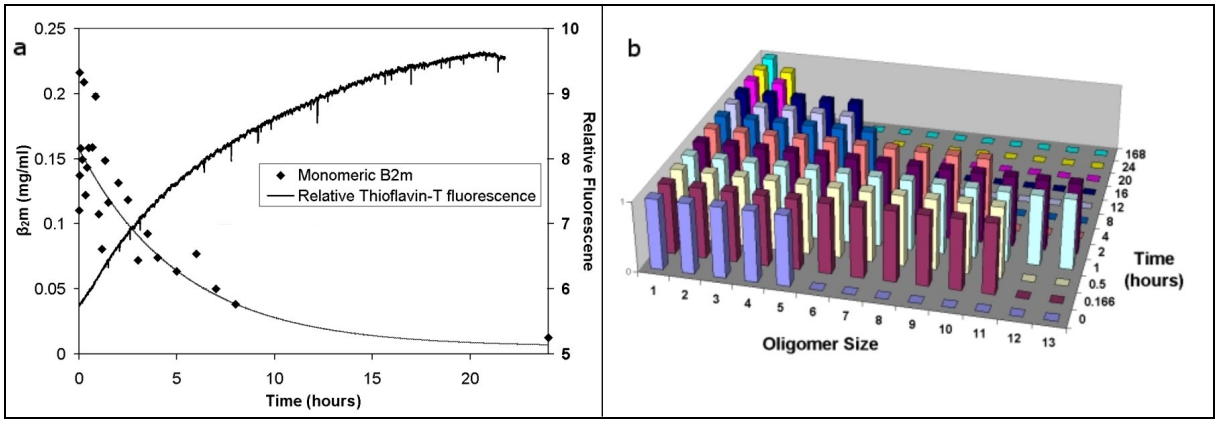


Fig. 2: a) A time course following the concentration of monomeric β_2m determined by ESI-MS at pH 3.6, 150mM ammonium formate, as amyloid forms. Amyloid fibril formation was also monitored using Thioflavin-T fluorescence. b) A time course following the presence or absence of oligomers of β_2m at pH 3.6, 150mM ammonium formate at different time points during amyloid assembly.

Funding.
We thank the BBSRC, The Wellcome Trust, Waters MS Technologies Center, and the University of Leeds for financial support.

Collaborators.
Bob Bateman, Iain Campuzano, Paul Read, Ashley Sage, Waters MS Technologies Center.

Publications.
Borysik, A.J.H., Radford, S.E. and Ashcroft, A.E. (2004) Co-populated conformational ensembles of beta(2)-microglobulin uncovered quantitatively by electrospray ionization mass spectrometry. *J Biol Chem* **279**, 27069-27077

Borysik A.J.H., Read P., Little D.R., Bateman R.H., Radford S.E. and Ashcroft A.E. (2004) Separation of beta(2)-microglobulin conformers by high-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled to electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom* **18**, 2229-2234

Amyloid under the atomic force microscope

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

Introduction

Amyloid fibrils are the underlying physical feature in a subset of protein misfolding diseases known collectively as the amyloidoses, and amyloid-like fibrils are also found in other areas of biology. The precise mechanism by which proteins assemble into amyloid is unknown, although it is beginning to emerge that the assembly of proteins into amyloid is heterogeneous, a property which has been used to explain the existence of prion-strains in prion-related biology. Furthermore, the self-assembly pathway, both *in vitro* and *in vivo*, is often associated with the existence of small 'prefibrillar' assemblies (*i.e.* either pore-like, small globular oligomers, and/or worm-like (flexible) fibrils), which form prior to the long, straight fibrils classically observed in amyloid disorders. In cell-viability assays, 'prefibrillar' assemblies have been shown to be toxic, and thus determining their role in the mechanism of amyloid formation is imperative. Utilising atomic force microscopy (AFM), we have studied the self-association mechanism of β_2-microglobulin (β_2m) into amyloid-like fibrils *in vitro*.

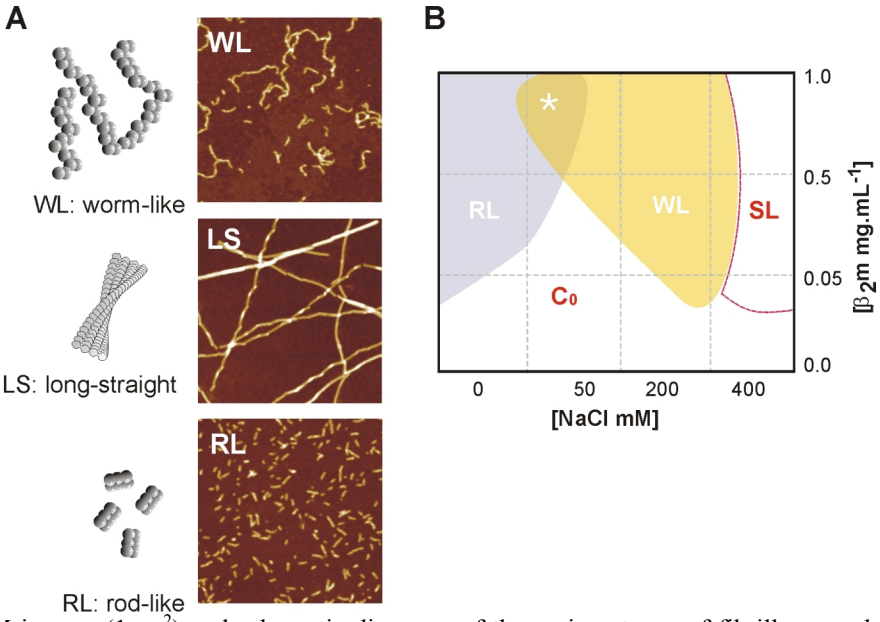


Fig. 1. A. AFM images (1μm^2) and schematic diagrams of the various types of fibrillar morphologies formed when β_2m assembles into amyloid-like fibrils under different conditions *in vitro*. **B.** Example state-diagram describing fribil morphology *versus* solution conditions, determined using AFM. In this case the effect of NaCl and protein concentration, in pH 3.5 buffer, on the end-point morphology after 2-4 weeks of incubation at 37°C is shown. Here (*) denoted heterogeneous regions, while SL is the solubility limit of monomeric and/or polymeric β_2m, and C_0 the apparent critical concentration for assembly.

The heterogeneous nature of β_2m assembly

The deposition of β_2m, a 99 residue all-β-sheeted protein, into amyloid fibrils is associated with the condition 'haemodialysis-related amyloidosis', which affects all patients with renal failure on long-term dialysis. *In vitro* at low pH, β_2m forms distinct classes of amyloid fibrils, as defined by differences in their morphology and persistence-length. The appearance of these fibrils types (worm-like (WL); rod-like (RL) and long, straight (LS)) is highly dependent on the solution conditions (Fig. 1A). In order to determine the factors governing these differences, we used AFM to map systematically the final end-point fibril morphology after long periods of incubation (~4 weeks), varying the conditions such as pH, salt and protein concentration. These data, encompassing over 500 AFM images, were then used to

construct state diagrams (Fig. 1B), and used to define regions where different fibril morphologies are favoured over others. To determine the relationships between various fibril types a series of experiments are currently being performed in which fibrils of one morphological type are formed and the solution conditions are then rapidly changed such that they cross a 'state-boundary'. Current data suggest that fibrils of different morphology form on distinct and competitive pathways of assembly, defining an energy landscape that rationalises the sensitivity of fibril morphology on the solution conditions.

Publications

Gosal, W.S., Myers, S.L., Radford, S.E. and Thomson, N.H. (2005) Amyloid under the atomic force microscope. *Protein & Peptide Letters.* **In press**.

Gosal, W.S., Morten, I.J., Hewitt, E.W., Smith, D.A., Thomson, N.H. and Radford, S.E. (2005) Competing pathways determine fibril morphology in the self-assembly of β_2-microglobulin into amyloid. **Submitted**.

Funding

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

Crystal structures of bacteriophage MS2 coat protein mutants complexed with Qβ RNA stemloop operators – a proposed discrimination mechanism for the binding of RNA operators by related phages.

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 dimmers, 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 has, for many years, been the paradigm for studying RNA/protein interactions at the atomic level.

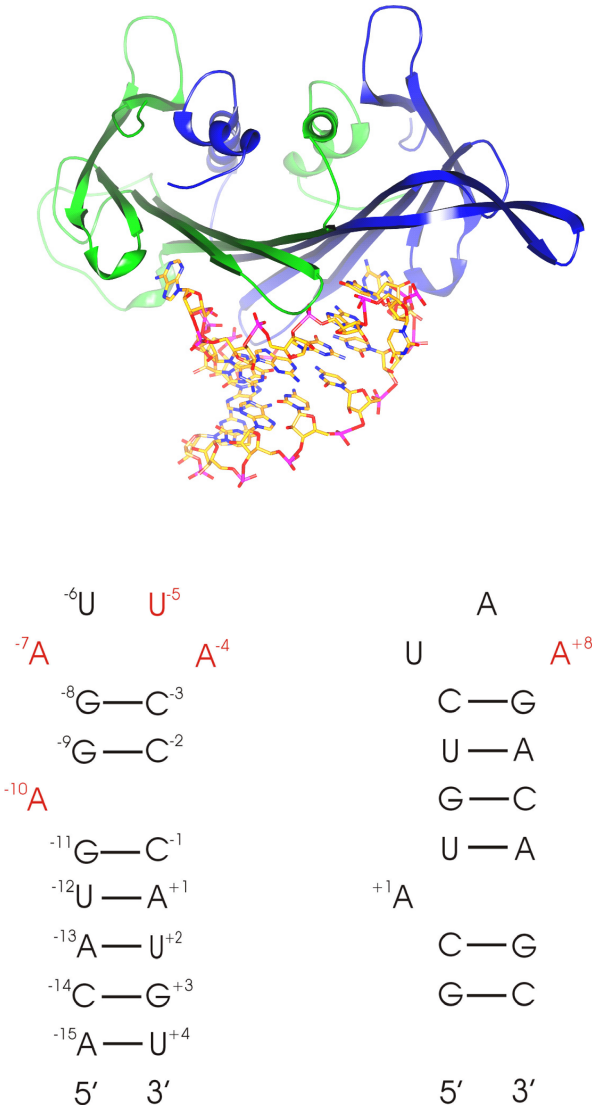


Fig. 1
(Top) Ribbon diagram of an AB coat protein dimer from the MS2 phage (subunit A in blue, B in green) complexed with the wild-type MS2 RNA stem-loop operator (shown in stick format).
(Bottom) Diagram of the secondary structure of the MS2 (left) and Qβ (right) stemloop operators. Residues that have been shown to be important for high affinity binding to capsid protein are highlighted in red.

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. 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. 1). *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 at residues 87 and 89 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 with 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.

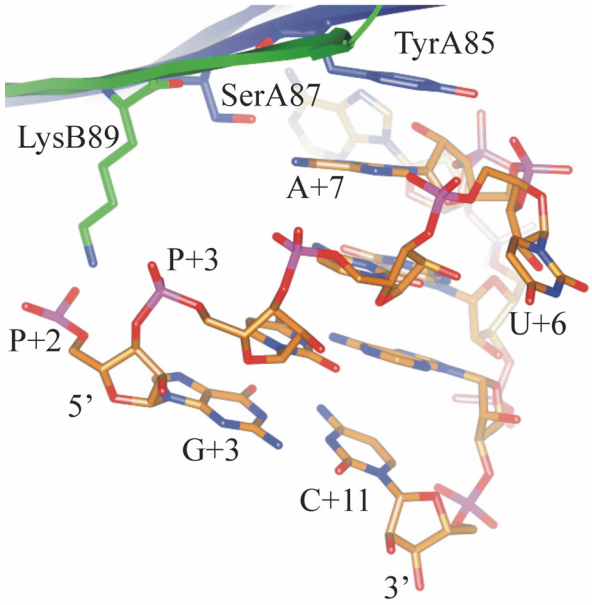


Fig. 2. Diagram showing the structure of the complex between the Qβ RNA at the AB binding site and the N87S, E89K MS2 coat protein, subunit A is in blue, subunit B in green. RNA is shown in stick format.

Results

Diffraction data were collected for three different MS2 mutants (N87S, E89K and N87S, E89K) complexed with Qβ RNA stemloop operators at the SRS, Daresbury, UK. Although the electron density for the lower stem region of the RNA is weak in each of the complexes, unambiguous modelling of the loop and upper stem region of the RNA is possible (Fig. 2). In contrast to the four base loop observed in the MS2 stemloop-coat protein complex, the Qβ RNA maintains its three base loop topology on complex formation. The N87S mutation mediates the binding of Qβ RNA by its shorter sidechain allowing the stacking of A+7 onto the underside of TyrA85, an interaction that is not favoured by the bulky asparagine residue. The E89K mutation leads to the LysB89 sidechain being located between the phosphate groups at the P+2 and P+3 positions thus giving potential for the formation of hydrogen bond interactions which explains the increase in affinity for Qβ RNA displayed by capsids with this mutation. Thus the effects of these two mutations on binding affinity can be explained.

No crystal structure of the Qβ RNA stem-loop complexed to Qβ coat protein has yet been determined. However, using the crystallographic structure of the Qβ RNA complexed to the MS2 coat protein mutants as a model, and given the structural similarities between the MS2 and Qβ coat proteins, it has been possible to propose a model of the complex. A paper describing these crystallographic studies is currently in preparation.

Collaborators

Lars Liljas and Kaspars Tars, Uppsala University, Sweden

Funding

This work was funded by the BBSRC.

Towards the structure of a bacteriophage T7 endonuclease I / Holliday junction complex

Jonathan M. Hadden and Simon E.V. Phillips

Homologous genetic recombination is important in the repair of double-strand breaks in DNA, in the rescue of stalled replication forks, and in the creation of genetic diversity. The central intermediate in this process is the four-way (Holliday) DNA junction. This must be ultimately resolved by nucleases that are selective for the structure of the junction.

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 and accumulate branched DNA intermediates. We recently presented the crystal structures of a catalytically impaired mutant of endonuclease I (E65K) without metals bound (Hadden, *et al.* 2001) and the wild-type protein with metals bound (Hadden, *et al.* 2002) both in the absence of DNA. We are currently trying to crystallise an endonuclease I / Holliday junction complex. By studying the structure of this complex we hope to be able to study the mechanism of Holliday junction cleavage in more detail.

A number of different synthetic Holliday junctions have been synthesized and these have been incubated with different inactive mutants of endonuclease I. After incubation the complex was subjected to gel-filtration chromatography using a Superdex75 column. The peak corresponding to an endonuclease I Holliday junction complex was isolated and concentrated prior to setting up crystallisation trials. Each complex was subjected to a 576 condition crystallisation trial using robotically based high-throughput procedures. To date 36 potential crystallisation conditions have been identified. Some of the crystals obtained are shown in Fig. 1.

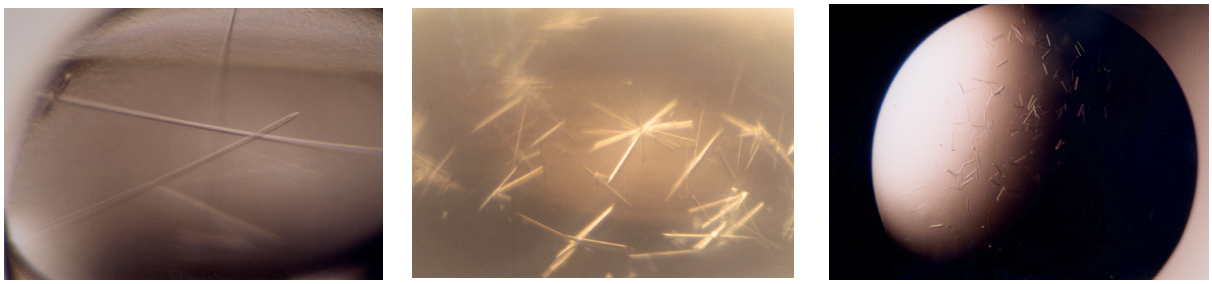


Fig. 1: Crystals of an Endonuclease I / Holliday junction complex

The best crystals obtained diffract X-rays to 5Å and we are currently investigating a number of different methods to improve upon this.

Publications

Déclais, A.-C, Fogg, J.M., Freeman, A., Coste, F., Hadden, J.M., Phillips, S.E.V. and Lilley, D.M.J. (2003) The complex between a four-way DNA junction and T7 endonuclease I. *EMBO J.*, **22**, 1398-1409.

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 Cancer Research UK for financial support.

Crystal structure of human keto-hexokinase and its complexes with fructose and a nucleotide analogue.

Chi H. Trinh, Aruna Asipu, Bruce E. Hayward, David T. Bonthron and Simon E.V. Phillips

Introduction

Essential fructosuria is a benign inborn error of metabolism. It is a condition which is believed to result from deficiency of hepatic fructokinase (ketohexokinase, KHK). KHK catalyses the phosphorylation of the ketose sugar fructose, to yield fructose-1-phosphate. Alternative splicing of the KHK gene generates a "central", predominantly hepatic isoform ketohexokinase-C (KHK-C) and a more widely distributed ketohexokinase-A (KHK-A). Both KHK isoforms are active. KHK-A has much poorer affinity for fructose than KHK-C but is considerably more thermostable. Mutations that cause essential fructosuria consequently result in significant loss of KHK-C activity but not of KHK-A. Affected individuals therefore probably have a selective deficiency of hepatic KHK, with peripheral KHK-A being preserved. Here, we report the structure of the human KHK-C and its comparison with the previously solved structure of human KHK-A in complex with fructose and an ATP analogue.

Crystallographic Studies

KHK-C was crystallised using vapour diffusion techniques. Data to a resolution of 2.9Å was collected using synchrotron radiation at Daresbury Laboratory at a temperature of 100 K. The KHK-C structure was determined by the method of molecular replacement using the structure of KHK-A as the starting model. Model building and refinement was accomplished using the programs O and CNS, respectively.

Structures of ketohexokinase-C and ketohexokinase-A

There are two dimers of KHK-C in the asymmetric unit. Each KHK-C subunit consists of 298 residues. The alternative splicing of the gene results in a difference of a single region between the two isoforms of the protein (Fig. 1). Like the structure of KHK-A the subunit of KHK-C has two distinct secondary structural elements; a central α/β fold and a 4-stranded β-sheet. A notable difference between the two isoform structures is the orientation of the two subunits forming the KHK-C dimer when compared to that for KHK-A. When one of the subunits of KHK-C is superimposed onto the corresponding one from KHK-A, the central α/β fold of the molecules fit well. However the relative orientation of the second subunits is different (Fig. 1).

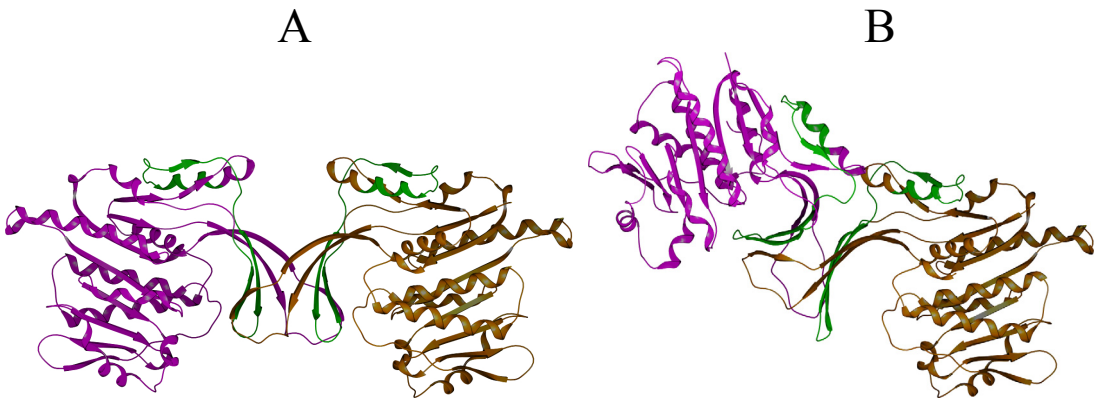


Fig. 1. Ribbon diagram of KHK-A dimer (A) and KHK-C dimer (B). For both structures the gold subunit is in the same orientation. The alternative splicing of the KHK gene resulting in a difference of a single region between the two isoforms of the protein (residues 72 to 115) is shown here in green.

The overall structure of KHK-C is similar to KHK-A. There is one active site per KHK monomer and this is located between the central α/β fold and the 4 stranded β-sheet forming the dimer interface (Fig. 2). For the structure of the KHK-A complex electron density is observed for both the fructose molecule and the nucleotide ATP analogue. Superposition of the two different isoforms monomer subunits revealed the residues forming the binding site to be conserved. The fructose molecule and the ATP analogue from the structure of KHK-A can be modelled into the binding cleft of KHK-C.

Fig. 2. Structure of the ketohexokinase-A monomer binding to a fructose molecule and an ATP analogue, AMP-PNP

Collaborators
Aruna Asipu, Bruce E. Hayward and David T. Bonthron, Molecular Medicine Unit, University of Leeds, St. James's University Hospital, Leeds, UK

Funding
This work was funded by the Wellcome Trust.

Identification of a new monosaccharide in *Mycobacterium tuberculosis* lipoarabinomannan

Bruce Turnbull, Kazumi Shimizu, Achim Treumann, Susanne Hesketh and Steve Homans

Introduction

Tuberculosis (TB) remains a major threat to world health, with approximately eight million cases of TB resulting in two million deaths per annum. Ninety-five percent of TB cases arise in the developing world, where the HIV epidemic has dramatically increased the number of people who succumb to the disease. The WHO has highlighted the urgent need for more effective anti-TB drugs and for rapid, specific and sensitive diagnostic tests for the causative agent, *Mycobacterium tuberculosis* (Mtb). The ability of the Mtb to enter and colonise macrophage cells can, in part, be attributed to a carbohydrate structure called lipoarabinomannan (LAM, Fig. 1a), that is present in the Mtb cell wall. Mannose residues that cap the dendritic arabinan chains of LAM, facilitate entry of Mtb into alveolar macrophages, following interaction with the macrophage mannose receptor. LAM then promotes the intracellular survival of Mtb by down-regulating the immune response and providing anti-oxidative protection for the bacterium. The mechanisms by which LAM exerts is biological activity inside macrophage cells remain to be elucidated.

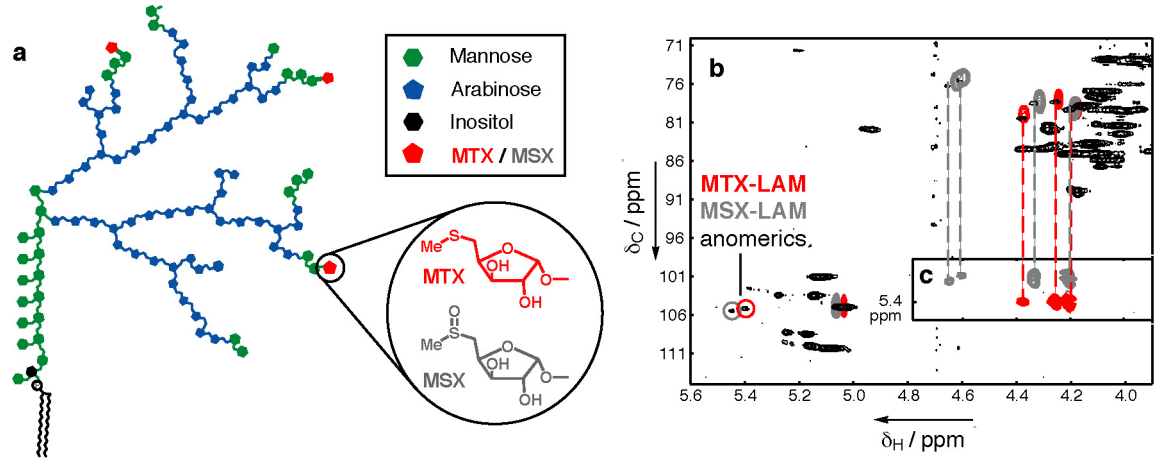


Fig. 1. (a) Cartoon representation of LAM showing the location of MTX/MSX (red) attached to the mannosyl caps (green) at the ends of the arabinan (blue) chains. (b) An overlay of HSQC spectra for the α-methyl glycosides of MTX (red), and MSX (grey) and LAM (black). (c) Inset is a strip from an HCCH-TOCSY spectrum of LAM which highlights the ^{1}H chemical shifts for MTX and MSX that correlate with the anomeric signals at ca. 5.4 ppm

Recently, we discovered an unusual sugar residue attached to the mannosyl caps of LAM. High resolution NMR spectroscopy and mass spectrometry studies revealed that this novel monosaccharide was a 5-methylthio-pentofuranoside, which was partially oxidised to the sulfoxide. However, it was not possible to determine the stereochemistry of the pentose using NMR methods. Consequently, we chemically synthesised all eight possible isomers of methylthio-pentose in the form of their methyl glycosides. On comparing the HSQC spectra for these compounds with that for LAM, it became clear that only the α-*xylo*-configured sugar (methylthio-xylofuranoside, MTX) gave signals that correlated with the natural product (Fig. 1b and c). Similarly, the chemical shifts for the sulfoxide form of the natural sugar were in generally good agreement with those for the oxidised α-xylofuranoside (MSX).

The discovery of MTX constitutes the first example of a methylthio-sugar residue incorporated into a polysaccharide, and only the fifth example of a xylo-configured sugar

outside the plant kingdom. Presumably, the biosynthetic precursor of MTX is methylthio-adenosine – a ubiquitous by-product of polyamine biosynthesis that is derived from SAM. Nevertheless, Mtb invests significant biosynthetic effort into incorporating MTX into its cell wall, which implies that this sugar may provide some advantage to the bacterium. The occurrence of MTX at the non-reducing termini of the polysaccharide, locates it at the bacterium's surface, where it could easily act as a ligand for receptors or regulatory proteins, either on the surface of the macrophage, or inside the host cell, following endocytosis. It is possible that conversion of MTX to its oxidised form (MSX) may even play a role in LAM-mediated oxidative protection for Mtb. Further structural and biochemical studies of MTX are in progress.

Collaborators
Delphi Chatterjee, Department of Microbiology, Immunology and Pathology, Colorado State University, USA.

Publications
Turnbull, W.B., Shimizu, K.H., Chatterjee, D., Homans, S.W. and Treumann, A. (2004) Identification of the 5-methylthiopentosyl substituent in *Mycobacterium tuberculosis* lipoarabinomannan. *Angew. Chem. Int. Ed.* **43**, 3918-3922.

Funding
This work has been funded by the Wellcome Trust and the University of Leeds.

Thermodynamics of binding of small hydrophobic ligands to the major urinary protein

Elizabeth A. Barratt, Richard J. Bingham, Richard W. Malham,
Simon E. V. Phillips and Steve W. Homans

Introduction

An understanding of protein ligand interactions is vital to our understanding of many biological processes. Isothermal titration microcalorimetry has long been used to study these interactions. However, this technique gives values for the overall changes in the energetics of a system upon binding of a ligand to a protein molecule, but can tell us nothing of the processes involved in causing these energetic changes. The Major Urinary Protein (MUP) (Fig. 1) is used as a model protein to study the processes involved in the binding of small, hydrophobic ligands to proteins. A great deal is already known about this protein, including the crystal structures of the apo form, as well as in the presence of many natural and synthetic ligands. NMR relaxation data has also been acquired to give an idea as to what happens in terms of molecular motions upon ligand binding. A combination of this data has allowed inferences to be made as to the role of individual residues. By using a combination of site-directed mutagenesis, isothermal titration microcalorimetry, X-ray crystallography, NMR spectroscopy and computational molecular modelling, it is hoped that we can further understand the contribution to the overall binding energetics of particular amino-acid residues and ligand protein interactions in this model system.

Hydrogen bonding

The binding of ligand to MUP is an enthalpically driven process. Factors which affect the enthalpy of binding include intrinsic factors, and solvation effects of both the protein and ligand. The MUP Y120F mutant was produced to probe the contribution to the binding energetics of the single hydrogen bond present in the MUP binding site. This hydrogen bond is produced by the interaction between the hydroxyl group of residue Tyr 120 and the ring nitrogen of the pyrazine ligands of MUP. Removal of this hydrogen bond would be expected to make the ligand binding reaction less favourable.

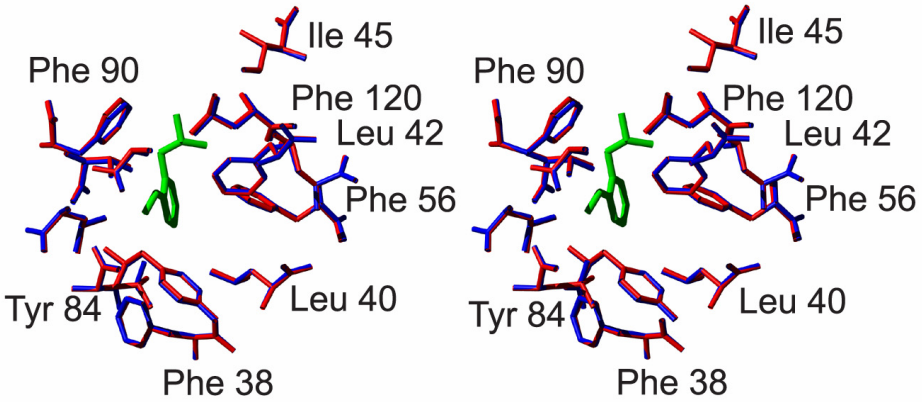


Fig. 1. Stereo view of the superimposition of the ligand binding sites observed in crystal structures of uncomplexed Y120F MUP (blue) and Y120F MUP (red) in complex with 2-isobutyl-3-methoxypyrazine (green). No ordered water molecules are observable within the binding site in either structure.

The ITC data recorded for this mutant showed the enthalpic contribution of binding to be less favourable in the Y120F mutant than in wild-type MUP, however overall ligand binding is still an enthalpically driven process. X-ray crystallography, along with molecular modelling showed there to be very little, if any, water in the MUP binding site in the absence of ligand. As a result of this, the only solvent re-organisation occuring when ligands bind to MUP is the

re-organisation of the water surrounding the ligand molecule. Therefore solvent reorganisation plays only a minor role in binding of ligand to MUP. This has been confirmed by the measurement of ITC data in deuterated solvent. This leaves the dominant interaction driving the binding of pyrazines to MUP as van der Waals interactions.

Van der Waals interactions

In order to examine the effect of van der Waals interactions on ligand binding, a number of mutants have been made where leucine and isoleucine residues in the MUP binding site have been substituted for alanines. A combination of all the techniques previously mentioned should give an insight into the role of van der Waals interaction on the thermodynamics of binding of pyrazines to MUP.

Collaborators

We gratefully acknowledge a fruitful collaboration with Daniel J. Warner and Charles A. Laughton at the University of Nottingham

Publications

Bingham, R., Bodenhausen, G., Findlay, J. B.C., Hsieh, S-Y., Kalverda, A.P., Kjellberg, A., Perazzolo, C., Phillips, S.E.V., Seshadri, K., Turnbull, W.B. and Homans, S.W. (2004) Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the Major Urinary Protein. *J. Am. Chem. Soc.* **126**, 1675-1681

Funding

We thank the BBSRC and The Wellcome Trust for financial support

Relating protein dynamics to the thermodynamics of ligand binding in arabinose binding protein

Christopher MacRaild, Agnieszka Bronowska, Antonio Hernández-Daranas, Arnout Kalverda and Steve Homans

Introduction

The interactions of proteins with small molecule ligands are of central importance to much of biology. The ability to predict and manipulate such interactions will open a number of future avenues in research, biotechnology and therapeutics. In particular, an understanding of the molecular basis of protein-small molecule interactions is crucial to attempts to design novel drug technologies.

Extensive research efforts have been aimed at relating the known structure of protein-ligand complexes to the thermodynamics of that interaction. Currently, such attempts have resulted in only limited improvements in our ability to predict or design novel protein-ligand interactions. One reason for the limited success of these approaches is their neglect of the role of protein and ligand conformational dynamics in determining ligand binding thermodynamics.

Our goal is to address this problem by means of nuclear magnetic resonance (NMR) relaxation analysis, which will provide specific insight into protein dynamics in the presence and the absence of ligand. These results will be integrated with thermodynamic data obtained from isothermal titration calorimetry (ITC), and from computational simulations, in the hope of building a more thorough understanding of the role of conformational dynamics in protein-ligand interactions. The model system chosen for this work is the interaction between the arabinose binding protein (ABP) with its natural ligands, the mono-saccharides L-arabinose and D-galactose, and derivatives thereof. This serves as an ideal model, in part because of the large unfavourable entropic contribution to binding, which is difficult to account for in terms of current understandings. This large entropy change on binding is a common characteristic of protein-carbohydrate interactions, and contributes to the fact that these interactions have proved particularly challenging for prediction and design.

NMR relaxation studies of ABP

An essential prerequisite to NMR studies of protein dynamics is the establishment of the ^{1}H, ^{15}N and ^{13}C resonances which contribute to the NMR spectra under consideration. Such assignments have been obtained for the complex of ABP with the ligand D-galactose using conventional triple-resonance assignment strategies (Fig. 1). Similar strategies have yielded near-complete assignments for the protein in the absence of ligand. Work is ongoing to complete these assignments using residue-specific labelling techniques. At over 33 kDa, the size of ABP and the consequent complexity of the NMR spectra makes assignment somewhat challenging.

Full sets of ^{15}N relaxation measurements (T1, T2 and NOE at two magnetic fields) have been made for ABP in its apo form and in complex with D-galactose, L-arabinose and D-fucose. The analysis of these measurements is currently underway, to extract information on the extent and timescales of ps-ns motion of the protein backbone and ^{15}N-containing sidechains. Similar measurements of relaxation parameters for ^{13}C and ^{2}H labelled methyl-containing sidechains are also planned.

Any differences in NMR relaxation parameters between apo ABP and the protein in complex with ligands will be interpreted in terms of changes in dynamics which occur when ABP

binds its ligand. In turn, the extent of observed dynamic changes can be related to binding thermodynamics by means of the recently established relationship between changes in order parameters derived from NMR relaxation and changes in conformational entropy. Because the NMR measurements are specific to individual bond vectors, this approach offers an unprecedented degree of structural resolution in thermodynamic analysis of protein function.

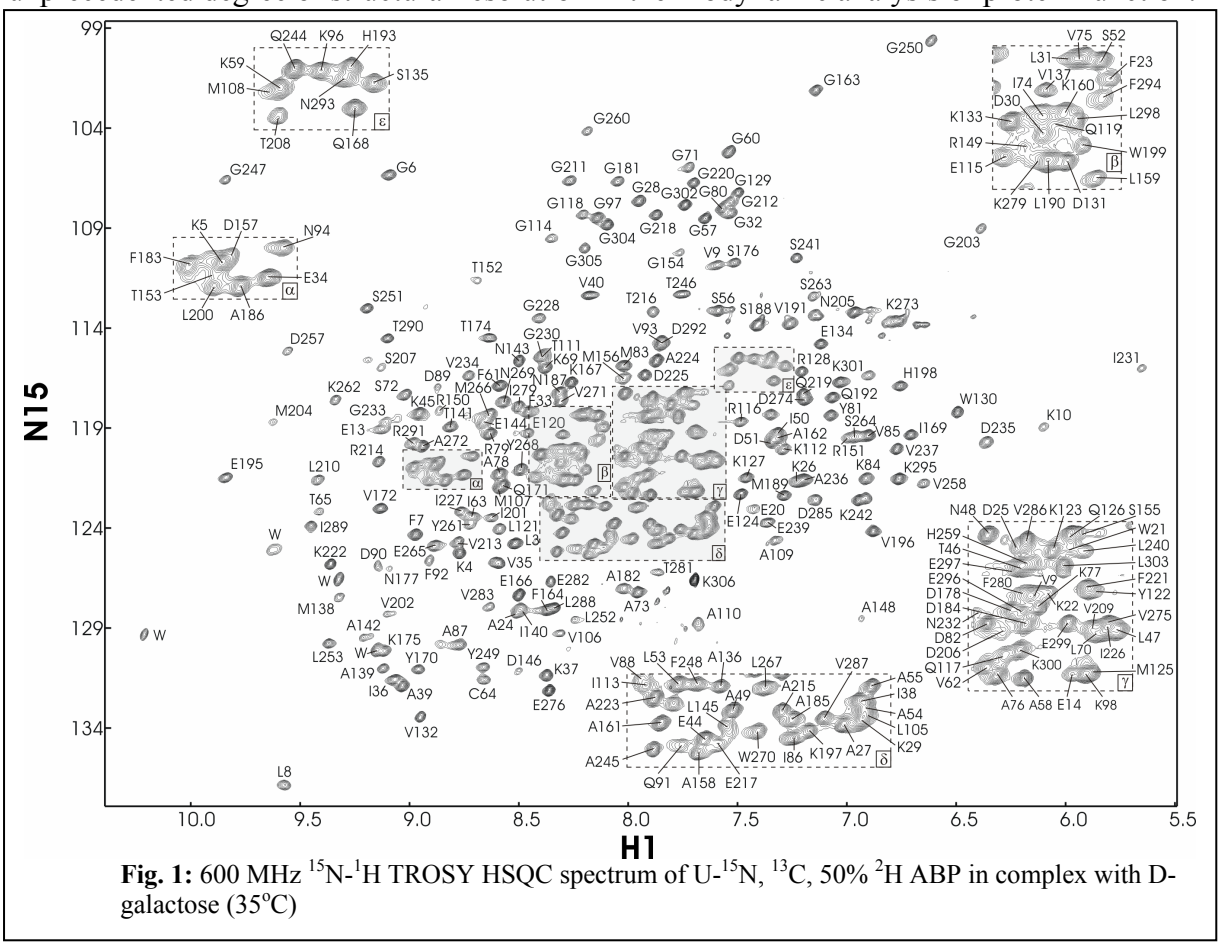


Fig. 1: 600 MHz ^{15}N-^{1}H TROSY HSQC spectrum of U-^{15}N, ^{13}C, 50% ^{2}H ABP in complex with D-galactose (35^{o}C)

ITC studies of the binding of ABP to deoxy derivatives of D-galactose

Isothermal titration calorimetry is a powerful tool for the thermodynamic analysis of protein-ligand interactions, providing from a single experiment three key thermodynamic parameters: change in free energy (ΔG°), change in enthalpy (ΔH°), and change in entropy (ΔS°). We have used ITC to examine the thermodynamics of binding of ABP to a panel of ligands comprising D-galactose and each of its deoxy derivatives. A number of high quality crystal structures exist for the corresponding complexes, and the high level of chemical similarity between these ligands means that changes in observed thermodynamics can be attributed to specific structural features within the complex.

Each of the deoxygalactose derivatives binds ABP with significantly reduced affinity, relative to D-galactose, highlighting the significant degree of specificity of this interaction. When effects arising due to solvation are taken into account, the change in the intrinsic binding free energy for each derivative relative to galactose is ~-30 kJ/mol. The fact that this value is similar for each derivative studied is remarkable, because each of the hydroxyl groups of galactose subtends a different number of hydrogen bonds to the protein. Furthermore, we find that the enthalpy of binding for 1-, 2- and 3-deoxygalactose is ~30 kJ/mol less favourable than that of galactose. This cannot be readily accounted for in terms of ligand solvation effects, leading to the unexpected conclusion that hydrogen bonds between protein and ligand are significantly more favourable in enthalpy than are hydrogen bonds between ligand and solvent.

Computational studies of solvation free energies

The process by which a ligand is bound within a protein binding site involves the removal of water molecules which prior to binding acted to solvate the ligand. Likewise, water molecules are displaced from the binding site by the ligand. The consequent loss of solvent-solute interactions makes an additional contribution to the thermodynamics of ligand binding which must be accounted for in any attempt to understand molecular basis of protein-ligand interactions. The magnitude of this contribution is equal to the solvation free energy of the de-solvated groups.

Solvation free energies can be measured for volatile or sparingly soluble solutes, but for highly soluble solutes such as D-galactose and its derivatives, experimental approaches are not feasible. Instead, we are currently investigating a number of computational approaches to address this problem. Specifically, an approach involving thermodynamic integration of alchemical transformations in solution and in vacuum is currently underway. Complementary approaches involving *ab initio* and semi-empirical quantum chemical calculations are also being investigated.

Publications

Hernández-Daranas, A., Shimizu, H. and Homans, S.W. (2004). Thermodynamics of binding of D-galactose and deoxy-derivatives thereof to the L-arabinose-binding protein. *J. Am. Chem. Soc.* **128,** 11870-11876.

Hernández-Daranas, A., Kalverda, A.P., Chiovitti, A. and Homans, S.W. (2004). Backbone resonance assignment of the L-arabinose binding protein in complex with D-galactose. *J. Biomol. NMR* **28,** 191-192.

Funding

This project is funded by the Wellcome Trust

Dynamics of the MS2 coat protein

Gary Thompson, Nicola Stonehouse, Peter Stockley and Steve Homans

Introduction

The MS2 bacteriophage presents a rich system for understanding the processes involved in assembly of large macromolecular systems and especially viruses with icosohedral geometry. Icosohedral viruses are a particularly fascinating system, as due to steric and symmetry constraints, identical coat proteins have to take up multiple conformations (quasi-equivalence). In the case of the MS2 protein, this involves forming protein dimers where the FG loop [residues 66 -83] take-up one of three conformations A, B or C, with each dimer containing either AB or CC pairs. It has previously been shown that addition of the 19mer MS2 translational repressor RNA (TR) to MS2 coat protein (MS2CP) initializes assembly of the complete virus. It has been postulated that the response to the TR RNA may be due to a templating reaction where the binding of the RNA affects the dynamics or conformation of the FG loop.

The aim of this project is to analyse the dynamics and structure of the FG loop in solution using NMR spectroscopy. To carry out this study, we have used a non-assembling mutant of the MS2 coat protein, MS2W82R, which still forms dimers. This system allows us to measure conformational and dynamic (entropic) changes within the protein, with and without RNA present, by NMR. So far a near complete assignment of the MS2W82R protein has been made using a triple labeled NMR sample [100% ^{15}N,100% ^{13}C, 50%D]. Analysis of the dynamics of the molecule has been made using ^{1}H-^{15}N heteronuclear relaxation experiments. ^{1}H-^{15}N heteronuclear nOe measurements of *apo*-MS2W82R (Fig. 1) clearly show significant flexibility within the FG loop without RNA present.

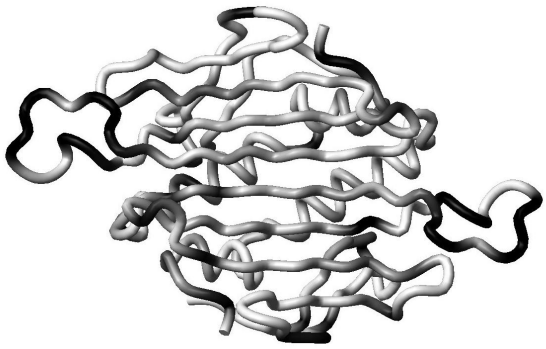


Fig. 1 Dynamics on ps-ns timescales for the backbone of *apo*-MS2W82R as detected ed by {^{1}H} -^{15}N Heteronculear nOe experiments. The image shows a tubes representation of the backbone of MS2W82R [PDB code 1MSC] with {^{1}H}-^{15}N heteronuclear nOe intensity shown by shading from white to black. White indicates little mobility [nOe ~ 0.8-0.9] and black indicates significant mobility [minimum nOe = 0.13]

Publications

Kalverda, A.P., Thompson, G.S. and Turnbull W.B. (2005) The importance of protein structural dynamics and the contribution of NMR spectroscopy in 'The Encylopedia of Genetics, Genomics, Proteomics and Bioinformatics', Wiley Interscience. In Press.

Stockley, P.G., Ashcroft, A.E., Francese, S., Thompson, G.S., Ransom, N., Smith, A., Homans, S.W. and Stonehouse, N.J. (2005) Dissecting the details of assembly of a T=3 Phage Capsid. *Journal of Theoretical Medicine.* In Press.

Funding

We wish to thank the BBSRC for funding.

Structure and function of mammalian scavenger receptors

Jane Murphy, Ravinder Vohra, John Hadden, Simon Phillips,
Shervanthi Homer-Vanniasinkam, John Walker and Sreenivasan Ponnambalam

Introduction

Atherosclerosis and vascular disease is responsible for 50% of the mortality in the UK and industrialised nations. Scavenger receptors are integral membrane proteins that bind modified low-density lipoproteins that are implicated in atherosclerotic plaque formation. A model protein that is the current focus of our efforts is the LOX-1 scavenger receptor that is found on vascular tissues especially endothelial cells. Increasingly, this receptor has also been reported to be expressed on other cell types including monocytes and macrophages that play a key role in the inflammatory response leading to atherosclerosis.

We are testing various bacterial, insect and mammalian expression systems to generate recombinant LOX-1 for biophysical, biochemical and cell biological studies. Structural approaches including NMR and X-ray crystallography are being investigated. Purified proteins are being tested for their ability to bind different ligands. Cell biological studies on the localisation and trafficking of LOX-1 allow us to reconcile how LOX-1 structure and function is linked to its temporal and spatial distribution within a cell. How such properties are altered in vascular disease states are an obvious area of interest.

Publications

Murphy, J.E., Tedbury, P., Homer-Vanniasinkam, S., Walker, J.H. and Ponnambalam, S. (2005) Biochemistry and cell biology of mammalian scavenger receptors. *Atherosclerosis,* In press.

Cobbold, C., Monaco, A.P., Sivaprasadarao, A. and Ponnambalam, S. (2003) Aberrant traficking of transmembrane proteins in human diseases. *Trends Cell Biol* **13**, 639-647.

Funding

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

Collaborations

We collaborate with Dr. John Walker (BMB, Leeds) Prof. Simon Phillips (BMB, Leeds) and Prof. Shervanthi Homer-Vanniasinkam (Vascular Surgery, Leeds General Infirmary) on this project.

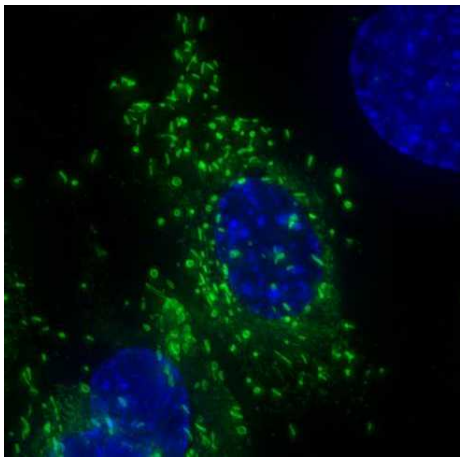
Membrane dynamics and remodelling by endothelial cells

Lorna Ewan, Shweta Mittar, Helen Jopling, Mudassir Mohammed, Gareth Howell, Shane Herbert, John Walker and Sreenivasan Ponnambalam

Introduction

Endothelial cells line all blood vessels and regulate vasculogenesis, angiogenesis and many aspects of vascular physiology. Angiogenesis is the sprouting of new blood vessels from pre-existing blood vessels. This phenomenon is regulated by the interaction of soluble growth factors (VEGFs) with cell surface receptors on the endothelium which can trigger protein secretion, gene expression and cellular proliferation. An overall aim is to understand how different biochemical interactions are linked to membrane remodelling by human endothelial cells during angiogenesis.

Other model systems include receptors that bind lipoproteins, lipid-modifying enzymes and factors that regulate blood clotting and blood pressure. Endothelial cell dysfunction is implicated in diseases such as atherosclerosis, cancer and diabetes. We are also studying atypical protein kinase C enzymes in regulating membrane dynamics at diverse intracellular locations such as the plasma membrane, Golgi apparatus and nucleus. These hormone-activated enzymes may function as molecular switches at different cellular locations.



The figure shows localisation of the pro-atherogenic hormone, Von Willebrand Factor, to cylindrical granules called Weibel-Palade bodies (Howell *et al*., 2004)

Publications

Grewal, S., Herbert, S.P., Ponnambalam, S., Walker, J.H. (2005) Cytosolic phospholipase A2α localises to different endothelial membrane compartments in an agonist and calcium-dependent manner. *FEBS Journal* **272,** 1278-1290.

Howell, G., Herbert, S., Smith, J.M., Ewan, L.C., Mittar, S., Mohammed, M., Hunter, A., Simpson, N., Turner, A., Zachary, I., Walker, J.H. and Ponnambalam, S. (2004) Endothelial cell confluence regulates Weibel-Palade body formation. *Mol Membr Biol* **21,** 413-421.

Grewal, S., Smith, J.M., Ponnambalam, S. and Walker, J.H. (2004) Membrane-dependent but calcium independent binding of cytosolic phospholipase A2α in endothelial cells. *Eur J Biochem* **271,** 69-77.

Cobbold, C., Coventry, J., Ponnambalam, S. and Monaco, A.P. (2004) Actin and microtubule regulation of *trans*-Golgi network architecture, and copper-dependent protein transport to the cell surface. *Mol Membr Biol* **21,** 59-66.

Collaborations

We collaborate with Dr. John Walker (BMB, Leeds), and Dr. Ian Zachary (Centre for Cardiovascular Biology and Medicine, University College London) on this project.

Funding

This work was supported by the British Heart Foundation, The White Rose Consortium, BBSRC and The Wellcome Trust.

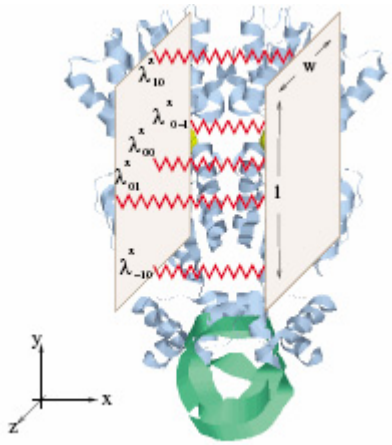
Brownian allostery in proteins

Rhoda Hawkins, Tom McLeish, Steve Homans, Peter Stockley,
Peter Knight and Stan Burgess

Introduction

Allosteric signalling is at the heart of protein networks: it is the effect by which the binding of a protein to a substrate molecule may be affected by the simultaneous binding of a second substrate, often at a site distant from the first. It permits a protein to act as a logical "gate", but poses a puzzle: how do the two binding sites communicate with each other? The canonical explanation invokes binding-induced conformational change that affects the distant site, but increasing evidence suggests that in many cases the information is transmitted not structurally, but dynamically. In such a picture of "allostery without conformational change", the Brownian motion of vibrational modes within the protein itself (some of them globally correlated over large distances) may be modified by the substrate binding. The binding free energy contains terms that arise from moderation of the amplitude of these fluctuations, so, although the motion itself is thermal and random, it may act as a channel for information flow across the protein.

We have applied these ideas to the DNA-binding *lac* repressor, using a coarse-grained model for the mutual vibrations of the dimeric protein. Harmonic potentials local to 5 patches on the dimer interface were themselves parameterised by a fully atomistic potential calculation (see Fig. 1). The full free-energy change on the three different binding states to DNA and co-repressor (lactose) could be calculated analytically within this model. We found that half of the allosteric free energy change can be attributed to the dynamic mechanism in the case of *lac*. Additionally, the model produces "design rules" for generating both co-operative and anti-cooperative binding that we are currently using to address other repressor systems.



More widely, this dynamic mechanism for allostery appears in many other systems, especially those involving coiled-coils. Current work is addressing the allosteric binding of the molecular motor dynein to microtubules, and the chemotaxis receptor cluster in *E. coli*. Both approaches work with models for the proteins that contain the long-wavelength modes (large scale), but not the local information, since small-scale dynamics is also generally unable to signal large distances. Initial results indicate that coiled-coils may transmit information along their length by bending, twisting and sliding motions of the two alpha-helices.

Publications

Hawkins, R.J. and McLeish, T.C.B. (2004) Coarse-grained model of entropic allostery, *Phys. Rev. Letts*, **93**, 098104.

Funding

This work was funded by the EPSRC.

Engineering substrate specificity in galactose oxidase

Sarah Deacon, Khaled Mahmoud, Kate Spooner, Susan Firbank, Kanjula Seneviratne,
Peter Knowles, Simon Phillips, and Michael McPherson

Galactose oxidase (GO; E.C. 1.1.3.9), is a 68kDa mononuclear copper-containing enzyme that oxidises primary alcohols to the corresponding aldehyde with coupled reduction of molecular oxygen to hydrogen peroxide according to the reaction scheme:

$$RCH_2OH + O_2 \rightarrow RCHO + H_2O_2$$

A copper ligand, Tyr 272, is covalently bonded through C_ε to the sulphur of Cys 228 and is the site of the catalytic free radical. The enzyme shows broad substrate specificity and displays a high K_M for substrates, including D-galactose (~70 mM). There is no crystallographic structure of substrate bound enzyme, but molecular modeling suggests a binding site consistent with known substrate specificity. For example, the enzyme is essentially inactive towards D –glucose, and the model indicates this would be due to a steric clash between the axial ligand Tyr495 and the O4 of glucose. We are interested in testing the substrate binding model and in altering substrate specificity of this enzyme, and one potential target substrate is fructose, against which the enzyme displays a very low activity.

The proposed site comprises Arg330, Gln406, Trp290, Phe 194 and Phe 464 (Fig. 1). We have introduced mutations at some of these residues to investigate the properties of the resulting proteins. To facilitate these studies we also developed an expression system based on the methylotrophic yeast, *Pichia pastoris,* to overcome some of the issues with the existing system based on the filamentous fungus *Aspergillus nidulans.* The crystal structure, spectroscopic properties and kinetic measurements show essentially no difference between the enzymes isolated from different expression hosts.

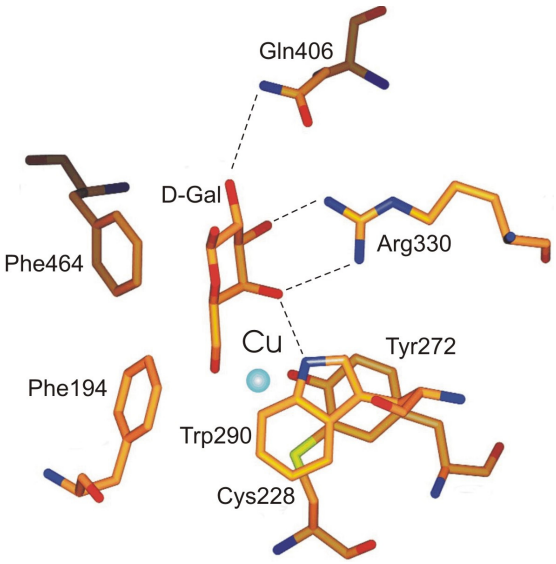


Fig. 1. Model of D-galactose binding to galactose oxidase. H-bonds are shown as dotted lines and the copper as a blue sphere.

Analysis of mutational variants revealed effects on kinetic parameters with substantial increases in K_M, indicating the importance in substrate binding of the residues tested (R330A and K, F464A).

The kinetic parameters for fructose oxidation by wild-type GO indicate a very high K_M ~2.5M and low k_{cat} (~9 $M^{-1}s^{-1}$). All the variants exhibited lower K_M values for fructose with the lowest (R330K) at ~1M. R330K was also the most interesting variant displaying an 8.2-fold increase in k_{cat}/K_M for fructose compared with wild-type GO, and a reduced level of discrimination between the substrates. Further studies are required to improve the kinetic parameters exhibited by this variant, but even at this stage the enzyme represents a better fructose oxidase than wild-type GO.

Publication

Deacon, S.E., Mahmoud, K., Spooner R.K., Firbank S J., Seneviratne, K., Knowles, P.F., Phillips, S.E.V. and McPherson, M.J. (2004) Enhanced fructose oxidase activity in a galactose oxidase variant. *ChemBioChem*, **5**, 972-979.

Funding

We acknowledge support from BBSRC.

Inhibitor studies with copper amine oxdiases

Christian Kurtis, Carrie Wilmot, Colin Saysell, Winston Tambyrajah, Simon Phillips,
Peter Knowles and Michael McPherson

Copper amine oxidases are homodimeric and contain a single cupric ion and a post-translationally modified tyrosine cofactor; 2,4,5-trihydroxyphenylalanine quinone (TPQ). TPQ biogenesis is an autocatalytic event requiring copper and oxygen. Together with international collaborators, we have engaged in a series of studies related to inhibitor binding to amine oxidases. For some enzymes, such as the human form, the identification of inhibitors that are selective for the copper and not flavin-containing amine oxidases is likely to prove an important advance. Crystallographic studies with the *E. coli* enzyme (ECAO), as a model for this class of enzymes, has revealed the formation of a reversible inhibitory complex with one enantiomer of tranylcypramine (TCP), an anti-depressant drug (Fig. 1). The binding of this drug to human copper amine oxidase may result in unwanted side-effects of the use of the inhibitor against the flavin-containing enzymes.

Fig. 1. The cover of FEBS letters issue showing the TCP complex

This TCP structural model is similar to the structure of the irreversible inhibitor 2-hydrazinopyridine (2-HP) bound to ECAO, which is widely used to investigate the formation of the TPQ cofactor and as a suicide inhibitor to study aspects of the enzyme mechanism.

We have undertaken a wide range of biochemical, chemical spectroscopic, mutational and structural studies to characterise the inhibition of ECAO by 2-HP. The interaction of this inhibitor with the enzyme results in two spectroscopically distinct species (adduct I and adduct II) that were considered to represent the equivalent to the substrate Schiff base and product Schiff base intermediates. We have now shown that in fact Adduct I represents both of these two intermediate states and that these are spectroscopically distinguishable. The shift to Adduct II involves a substantial rearrangement in the enzyme active site to allow the inhibitor TPQ complex to co-ordinate with the active site copper, a situation that is more readily achieved in the active site

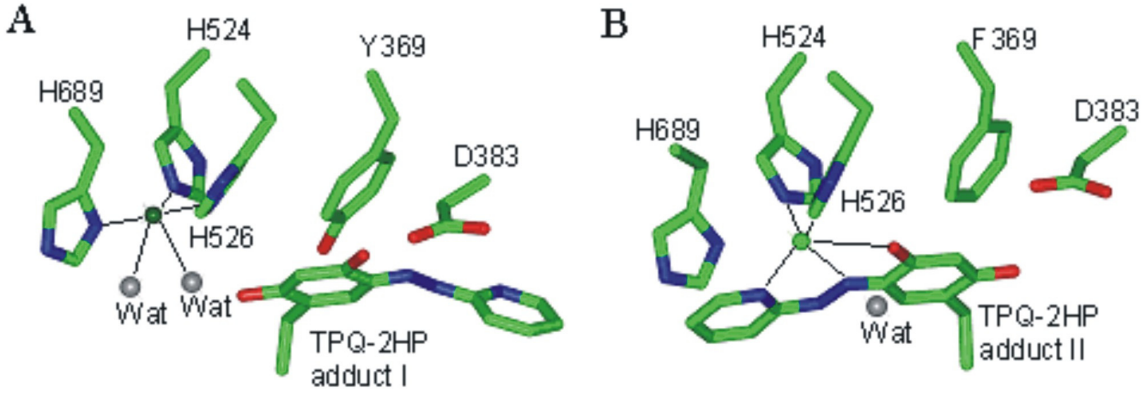


Fig. 2. Structure of 2-HP complexed with the Y369F variant of E. coli copper amine oxidase. (A) Adduct I with the 2-HP is the position normally occupied by substrate (B) Adduct II showing a dramatic rotation of the cofactor inhibitor complex to allow co-ordination with the copper centre.

variant Tyr369Phe which lacks a stabilizing H-bond interaction to the O4 of TPQ, thus facilitating rotation of the TPQ (Fig. 2).

We have also undertaken some work related to the human copper amine oxidase that functions as an important vascular adhesion protein. The mechanism by which this enzyme acts as a cell adhesion molecule is postulated to involve the formation of a covalent intermediate with peptide amine side chains on cell surface proteins. To test this, a series of peptides were examined for their ability to inhibit the enzyme activity. One peptide was shown to be an effective inhibitor supporting the prospect of a catalytic mechanism-based cell adhesion interaction. Such inhibitors may prove therapeutically valuable for control of inflammatory states.

Publications

Wilmot, C.M., Saysell, C.G., Kurtis, C.R.P., Blessington, A., Conn, D.A., McPherson, M.J., Knowles, P.F. and Phillips, S.E.V. (2004) Medical implications from the crystal structure of a copper-containing amine oxidase complexed with the antidepressant drug tranylcypromine, *FEBS Letters*, **576**, 301-305.

Yegutkin, G.G., Salminen, T., Koskinen, K., Kurtis, C.R.P, McPherson, M.J., Jalkanen, S. and Salmi, M. (2004) A peptide inhibitor of vascular adhesion protein-1 (VAP-1) blocks leukocyte-endothelial interactions under shear stress. *European Journal of Immunology* **34,** 2276-2285.

Mure, M., Brown, D.E., Saysell, C., Rogers, M.S., Wilmot, C.M., Kurtis, C.R.P., McPherson, M.J., Phillips, S.E.V., Knowles, P.F. and Dooley, D.M. (2005) Role of the interactions between the active site base and the substrate Schiff base in amine oxidase catalysis. Evidence from structural and spectroscopic studies of the 2-hydrazinopyridine adduct of *Escherichia coli* amine oxidase *Biochemistry*, **44**, 1568-1582.

Mure, M., Kurtis, C.R.P., Brown, D.E., Rogers, M.S., Tambyrajah, W.S., Saysell, C., Wilmot, C.M., Phillips, S.E.V., Knowles, P.F., Dooley, D.M. and McPherson, M.J. (2005) Active site rearrangement of the 2-hydrazinopyridine adduct in *Escherichia coli* amine oxidase to an azo copper(II) chelate form: A key role for tyrosine 369 in controlling the mobility of the TPQ-2HP adduct *Biochemistry*, **44**, 1583-1594.

Acknowledgements and Funding

We acknowledge the valuable collaborations with Prof Dave Dooley, Doreen Brown and Melanie Rogers at Montana State University, Prof Minae Mure at University of Kansas and Prof Sipra Jalkanen and Marko Salmiin Turku, Finland. This work was supported by BBSRC.

Astbury Seminars 2004

Thursday 8th January, 2004
Prof. Jim Naismith, Centre for Biomolecular Sciences, University of St. Andrews
"Biological halogenation"

Thursday 15th January, 2004
Dr Kathryn Chapman, Farfield Sensors, Salford University Business Park
"Make light work of protein characterisation"

Thursday 29th January, 2004
Prof. Felix Rey, Virologie Moleculaire & Structurale, CNRS, France
"Structure of class II enveloped virus membrane-fusion proteins"

Thursday 12th February, 2004
Prof. David Barford, Section of Structural Biology, The Institute of Cancer Research
"Mechanism of redox regulation of protein tyrosine phosphatases involves a novel sulphenyl-amide intermediate"

Thursday 11th March, 2004
Prof. Pete Artymiuk, Molecular Biology and Biotechnology, University of Sheffield
"What can structure tell us about informatics?"

Thursday 1st April, 2004
Dr. Dinshaw Patel
"Structural biology of RNA interference"

Tuesday 6th April, 2004
Dr Walter Stafford, Boston Biomedical Research Institute
"Analysis of Heteromeric Interacting Systems by Sedimentation"

Thursday 22nd April, 2004
Dr. Elizabeth Hewat, Institute de Biologie Structurale, Grenoble, France
"A cryo-elecron microscopy view of human rhinovirus infection"

Thursday 29th April 2004
Dr Bob MacCallum, Stockholm Bioinformatics Center, Sweden
"Predicting the nuclear proteome "

Thursday 6th May, 2004
Dr Nuria Verdaguer, Institut de Biologia Molecular de Barcelona
"Crystal structure of foot-and-mouth disease virus RNA-dependent RNA polymerase and its complex with a template-primer RNA"

Thursday 13th May, 2004
Dr. Bernadette Byrne, Department of Biochemistry, Imperial College
"Structural studies on nitrate metabolising enzymes and ABC-transport proteins"

Thursday 20th May, 2004
Dr Mitchell Guss, Molecular & Microbial Biosciences, University of Sydney
"New structures provide insights into the specificity and mechanism of copper amine oxidases"

Thursday 27th May, 2004
Prof. Hue Sun Chan, University of Toronto
"Cooperativity principles in protein folding: Calorimetric criteria, chevron plots and native-state hydrogen exchange"

Monday 7th June 2004
Luca Pellegrini, University of Cambridge
"Regulation of DNA recombination by the breast cancer susceptibility protein BRCA2"

Thursday 17th June 2004
Prof. Dagmar Ringe, Brandeis University
"Enzyme evolution: recognition vs mechanism"

Thursday 1st July 2004
Dr Gilad Haran, Chemical Physics Department, Weizmann Institute of Science, Israel
"The energy landscape of folding proteins - a single-molecule view"

Thursday 15 July 2004
Prof Jim Hogle, Harvard Medical School, Boston
"How does a nonenveloped virus get into a cell? Biochemical and structural characterization of poliovirus cell entry"

Thursday 28 October 2004
Dr Andrew J. Wilson, School of Chemistry, University of Leeds
"Adventures Assessing Architecture, Affinity and Activity"

Thursday 25 November 2004
Prof Timothy A. Keiderling, Department of Chemistry, University of Illinois, Chicago
"Taking Optical Spectroscopy to the (Structural) Limit for Peptide and Protein Applications with Isotopic Substitutions"

Thursday 9 December 2004
Dr Dale B. Wigley, Cancer Research UK, London Research Institute
"Crystal Structure of RecBCD; A Machine for Processing DNA Breaks"

Publications by Astbury Centre Members 2004

Appleford, P.J., Griffiths, M., Yao, S.Y.M., Ng, A.M.L., Chomey, E.G., Isaac, R.E., Coates, D., Hope, I.A., Cass, C.E., Young, J.D. and Baldwin, S.A. (2004). "Functional redundancy of two nucleoside transporters of the ENT family (CeENT1, CeENT2) required for development of *Caenorhabditis elegans*." *Molecular Membrane Biology* **21**, 247-259.

Baldwin, S.A., Beal, P.R., Yao, S.Y.M., King, A.E., Cass, C.E. and Young, J.D. (2004). "The equilibrative nucleoside transporter family, SLC29." *European Journal of Physiology* **447**, 735-743.

Baldwin, S.A. and Lucy, J.A. (2004). "The structural biology of membrane proteins." *Molecular Membrane Biology* **21**, 127-128.

Barnes, K., Ingram, J.C., Bennett, M.D.M., Stewart, G.W. and Baldwin, S.A. (2004). "Methyl-beta-cyclodextrin stimulates glucose uptake in Clone 9 cells: a possible role for lipid rafts." *Biochemical Journal* **378**, 343-351.

Bartlett, S. and Nelson, A. (2004). "Evaluation of alternative approaches for the synthesis of macrocyclic bisindolylmaleimides." *Organic & Biomolecular Chemistry* **2**, 2874-2883.

Bartlett, S. and Nelson, A. (2004). "Towards configurationally stable bisindolylmaleimide cyclophanes: potential tools for investigating protein kinase function." *Chemical Communications*, 1112-1113.

Bhogal, N., Blaney, F.E., Ingley, P.M., Rees, J. and Findlay, J.B.C. (2004). "Evidence for the proximity of the extreme N-terminus of the neurokinin-2 (NK2) tachykinin receptor to Cys(167) in the putative fourth transmembrane helix." *Biochemistry* **43**, 3027-3038.

Billington, N., Burgess, S.A., Patel, H., Chantler, P.D., White, H.D., Trinick, J. and Knight, P.J. (2004). "Conformational changes in myosin heads during ATPase." *Biophysical Journal* **86**, 403A-404A.

Bingham, R.J., Findlay, J.B.C., Hsieh, S.Y., Kalverda, A.P., Kjeliberg, A., Perazzolo, C., Phillips, S.E.V., Seshadri, K., Trinh, C.H., Turnbull, W.B., Bodenhausen, G. and Homans, S.W. (2004). "Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein." *Journal of the American Chemical Society* **126**, 1675-1681.

Borysik, A.J.H., Radford, S.E. and Ashcroft, A.E. (2004). "Co-populated conformational ensembles of beta(2)-microglobulin uncovered quantitatively by electrospray ionization mass spectrometry." *Journal of Biological Chemistry* **279**, 27069-27077.

Borysik, A.J.H., Read, P., Little, D.R., Bateman, R.H., Radford, S.E. and Ashcroft, A.E. (2004). "Separation of beta(2)-microglobulin conformers by high-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled to electrospray ionisation mass spectrometry." *Rapid Communications in Mass Spectrometry* **18**, 2229-2234.

Brakoulias, A. and Jackson, R.M. (2004). "Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: An automated all-against-all structural comparison using geometric matching." *Proteins-Structure Function and Bioinformatics* **56**, 250-260.

Burgess, S., Walker, M., Knight, P.J., Sparrow, J., Schmitz, S., Offer, G., Bullard, B., Leonard, K., Holt, J. and Trinick, J. (2004). "Structural studies of arthrin: Monoubiquitinated actin." *Journal of Molecular Biology* **341**, 1161-1173.

Burgess, S.A. and Knight, P.J. (2004). "Is the dynein motor a winch?" *Current Opinion in Structural Biology* **14**, 138-146.

Burgess, S.A., Walker, M.L., Sakakibara, H., Oiwa, K. and Knight, P.J. (2004). "The structure of dynein-c by negative stain electron microscopy." *Journal of Structural Biology* **146**, 205-216.

Burgess, S.A., Walker, M.L., Thirumurugan, K., Trinick, J. and Knight, P.J. (2004). "Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules." *Journal of Structural Biology* **147**, 247-258.

Burt, J., Dean, T. and Warriner, S. (2004). "The parallel synthesis of a disaccharide library using a solid phase, peptide-templated strategy." *Chemical Communications*, 454-455.

Calderwood, M.A., Hall, K.T., Matthews, D.A. and Whitehouse, A. (2004). "The herpesvirus saimiri ORF73 gene product interacts with host- cell mitotic chromosomes and self-associates via its C terminus." *Journal of General Virology* **85**, 147-153.

Calderwood, N.A., White, R.E. and Whitehouse, A. (2004). "Development of herpesvirus-based episomally maintained gene delivery vectors." *Expert Opinion on Biological Therapy* **4**, 493-505.

Caryl, J.A., Smith, M.C.A. and Thomas, C.D. (2004). "Reconstitution of a staphylococcal plasmid-protein relaxation complex in vitro." *Journal of Bacteriology* **186**, 3374-3383.

Cobbold, C., Coventry, J., Ponnambalam, S. and Monaco, A.P. (2004). "Actin and microtubule regulation of trans-Golgi network architecture, and copper-dependent protein transport to the cell surface." *Mol Membr Biol* **21**, 59-66.

Cooksey, J., Gunn, A., Kocienski, P.J., Kuhl, A., Uppal, S., Christopher, J.A. and Bell, R. (2004). "The nucleophilic addition of alpha-metallated 1,3-dioxanes to planar chiral cationic eta(3)-allylmolybdenum complexes. Synthesis of (2E,5S,6R,7E)-6-methyl-8-phenylocta-2,7-dienoic acid methyl ester, a key component of the Cryptophycins." *Organic & Biomolecular Chemistry* **2**, 1719-1731.

Daranas, A.H., Kalverda, A.P., Chiovitti, A. and Homans, S.W. (2004). "Backbone resonance assignment of the L- arabinose binding protein in complex with D-galactose." *Journal of Biomolecular NMR* **28**, 191-192.

Daranas, A.H., Shimizu, H. and Homans, S.W. (2004). "Thermodynamics of binding of D-galactose and deoxy derivatives thereof to the L-arabinose-binding protein." *Journal of the American Chemical Society* **126**, 11870-11876.

de Maturana, R.L., Treece-Birch, J., Abidi, F., Findlay, J.B.C. and Donnelly, D. (2004). "Met-204 and Tyr-205 are together important for binding GLP-1 receptor agonists but not their N-terminally truncated analogues." *Protein and Peptide Letters* **11**, 15-22.

Deacon, S.E., Mahmoud, K., Spooner, R.K., Firbank, S.J., Knowles, P.F., Phillips, S.E.V. and McPherson, M.J. (2004). "Enhanced fructose oxidase activity in a galactose oxidase variant." *Chembiochem* **5**, 972-979.

Dimitriadis, G., Drysdale, A., Myers, J.K., Arora, P., Radford, S.E., Oas, T.G. and Smith, D.A. (2004). "Microsecond folding dynamics of the F13W G29A mutant of the B domain of staphylococcal protein A by laser-induced temperature jump." *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3809-3814.

Emmett, S.R., Dove, B.K., Mahoney, L., Wurm, T. and Hiscox, J.A. (2004). "The cell cycle and virus infection." *Methods in Molecular Biology.* **296**, 197-218.

Fadouloglou, V.E., Tampakaki, A.P., Glykos, N.M., Bastaki, M.N., Hadden, J.M., Phillips, S.E., Panopoulos, N.J. and Kokkinidis, M. (2004). "Structure of HrcQ(B)-C, a conserved component of the bacterial type III secretion systems." *Proceedings of the National Academy of Sciences of the United States of America* **101**, 70-75.

Firbank, S., Rogers, M., Guerrero, R.H., Dooley, D.M., Halcrow, M.A., Phillips, S.E., Knowles, P.F. and McPherson, M.J. (2004). Cofactor processing galactose oxidase. *Free Radicals: Enzymology, Signalling and Disease***:** 15-25.

Fournier, L., Kocienski, P. and Pons, J.M. (2004). "The beta-lactone route to alpha,beta-unsaturated delta- lactones. Total syntheses of (+/-)-goniothalamin and (-)-massoialactone." *Tetrahedron* **60**, 1659-1663.

Friel, C.T., Beddard, G.S. and Radford, S.E. (2004). "Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design." *Journal of Molecular Biology* **342**, 261-273.

Gorski, S.A., Le Duff, C.S., Capaldi, A.P., Kalverda, A.P., Beddard, G.S., Moore, G.R. and Radford, S.E. (2004). "Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7." *Journal of Molecular Biology* **337**, 183-193.

Grewal, S., Smith, J., Ponnambalam, S. and Walker, J. (2004). "Stimulation-dependent recruitment of cytosolic phospholipase A(2)-alpha to EA.hy.926 endothelial cell membranes leads to calcium-independent association." *European Journal of Biochemistry* **271**, 69-77.

Griffin, S.D.C., Harvey, R., Clarke, D.S., Barclay, W.S., Harris, M. and Rowlands, D.J. (2004). "A conserved basic loop in hepatitis C virus p7 protein is required for amantadine-sensitive ion channel activity in mammalian cells but is dispensable for localization to mitochondria." *Journal of General Virology* **85**, 451-461.

Hao, J. and Berry, A. (2004). "A thermostable variant of fructose bisphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents." *Protein Eng. Des. Sel.* **17**, 689-697.

Hawkins, R.J. and McLeish, T.C.B. (2004). "Coarse-grained model of entropic allostery." *Physical Review Letters* **93**, art. no.-098104.

Heddle, J.G., Mitelheiser, S., Maxwell, A. and Thomson, N.H. (2004). "Nucleotide binding to DNA gyrase causes loss of DNA wrap." *Journal of Molecular Biology* **337**, 597-610.

Heinrich, M., Pyckhout-Hintzen, W., Allgaier, J., Richter, D., Straube, E., McLeish, T.C.B., Wiedenmann, A., Blackwell, R.J. and Read, D.J. (2004). "Small-angle neutron scattering study of the relaxation of a melt of polybutadiene H-polymers following a large step strain." *Macromolecules* **37**, 5054-5064.

Hernández-Daranas, A., Kalverda, A.P., Chiovitti, A. and Homans, S.W. (2004). "Backbone resonance assignment of the L-arabinose binding protein in complex with D-galactose." *J. Biomol. NMR* **28**, 191-192.

Hernández-Daranas, A., Shimizu, H. and Homans, S.W. (2004). "Thermodynamics of binding of D-galactose and deoxy-derivatives thereof to the L-arabinose-binding protein." *J. Am. Chem. Soc.* **128**, 11870-11876.

Hodgson, R. and Nelson, A. (2004). "A two-directional synthesis of the C-58-C-71 fragment of palytoxin." *Organic & Biomolecular Chemistry* **2**, 373-386.

Homans, S.W. (2004). "NMR spectroscopy tools for structure-aided drug design." *Angewandte Chemie-International Edition* **43**, 290-300.

Horn, W.T., Convery, M.A., Stonehouse, N.J., Adams, C.J., Liljas, L., Phillips, S.E.V. and Stockley, P.G. (2004). "The crystal structure of a high affinity RNA stem-loop complexed with the bacteriophage MS2 capsid: Further challenges in the modeling of ligand-RNA interactions." *RNA* **10**, 1776-1782.

Houmeida, A., Thompson, B., Tskhovrebova, L., Knight, P.J., Thirumurugan, K., Stafford, W.F. and Trinick, J. (2004). "Characterisation of the region of titin that forms end-filaments in the I-band of muscle sarcomeres." *Biophysical Journal* **86**, 212A-212A.

Howell, G.J., Herbert, S.P., Smith, J.M., Mittar, S., Ewan, L.C., Mohammed, M., Hunter, A.R., Simpson, N., Turner, A.J., Zachary, I., Walker, J.H. and Ponnambalam, S. (2004). "Endothelial cell confluence regulates Weibel-Palade body formation." *Molecular Membrane Biology* **21**, 413-421.

Jung, H.S., Burgess, S.A., Colegrave, M., Patel, H., Chantler, P.D., Chalovich, J.M., Trinick, J. and Knight, P.J. (2004). "Comparative studies of the folded structures of scallop striated and vertebrate smooth muscle myosins." *Biophysical Journal* **86**, 403A-403A.

Jung, H.S., Park, H.H., Mendieta, I.R. and Smith, D.A. (2004). "Determination of bonding structure of Si, Ge, and N incorporated amorphous carbon films by near-edge X-ray absorption fine structure and ultraviolet Raman spectroscopy." *Journal of Applied Physics* **96**, 1013-1018.

Kawakami, M., Byrne, K., Khatri, B., McLeish, T.C.B., Radford, S.E. and Smith, D.A. (2004). "Viscoelastic properties of single polysaccharide molecules determined by analysis of thermally driven oscillations of an atomic force microscope cantilever." *Langmuir* **20**, 9299-9303.

Kennedy, A., Nelson, A. and Perry, A. (2004). "Highly diastereoselective addition of ketone enolates to N- sulfinyl imines: Asymmetric synthesis of syn- and anti-1,3- amino alcohol derivatives." *Synlett*, 967-970.

Khan, A., Ashcroft, A.E., Korchazhkina, O.V. and Exley, C. (2004). "Metal-mediated formation of fibrillar ABri amyloid." *J. Inorg. Biochem.* **98**, 2006 - 2010.

Kocienski, P., Bell, R., Bourque, E., Christopher, J.A., Cooksley, J., Gunn, A., Kuhl, A., Li, Y.F., Uppal, S. and Yuen, J. (2004). "Synthetic applications of planar chiral cationic eta(3)- allylmolybdenum complexes." *Pure and Applied Chemistry* **76**, 477-494.

Lawrenson, I.D. and Stockley, P.G. (2004). "Kinetic analysis of operator binding by the *E. coli* methionine repressor highlights the role(s) of electrostatic interactions." *Febs Letters* **564**, 136-142.

Lister, I., Roberts, R., Schmitz, S., Walker, M., Trinick, J., Veigel, C., Buss, F. and Kendrick-Jones, J. (2004). "Myosin VI: a multifunctional motor." *Biochemical Society Transactions* **32**, 685-688.

Lister, I., Schmitz, S., Walker, M., Trinick, J., Buss, F., Veigel, C. and Kendrick-Jones, J. (2004). "A monomeric myosin VI with a large working stroke." *Embo Journal* **23**, 1729-1738.

Macdonald, A., Crowder, K., Street, A., McCormick, C. and Harris, M. (2004). "The hepatitis C virus NS5A protein binds to members of the Src family of tyrosine kinases and regulates kinase activity." *Journal of General Virology* **85**, 721-729.

Macdonald, A. and Harris, M. (2004). "Hepatitis C virus NS5A: tales of a promiscuous protein." *Journal of General Virology* **85**, 2485-2502.

McConkey, G.A., Pinney, J.W., Westhead, D.R., Plueckhahn, K., Fitzpatrick, T.B., Macheroux, P. and Kappes, B. (2004). "Annotating the Plasmodium genome and the enigma of the shikimate pathway." *Trends in Parasitology* **20**, 60-65.

McCormick, C.J., Challinor, L., Macdonald, A., Rowlands, D.J. and Harris, M. (2004). "Introduction of replication-competent hepatitis C virus transcripts using a tetracycline-regulable baculovirus delivery system." *Journal of General Virology* **85**, 429-439.

Michalopoulos, I., Torrance, G.M., Gilbert, D.R. and Westhead, D.R. (2004). "TOPS: an enhanced database of protein structural topology." *Nucleic Acids Research* **32**, D251-D254.

Mocquet, C.M. and Warriner, S.L. (2004). "Metal-free bifunctional catalysis of the asymmetric Baylis- Hillman reaction." *Synlett*, 356-358.

Morten, I.J., Hewitt, E.W. and Radford, S.E. (2004). Beta-2-microglobulin and dialysis-related amyloidosis. *Protein Misfolding, Aggregation and Conformational Diseases*. V. N. a. F. Uversky, A.L, Kluwer Academic/Plenum Publishers.

Murtagh, M.S., Burgess, S., Sparrow, J., Knight, P.J. and Trinick, J. (2004). "Single particle image processing of muscle thin filaments." *Biophysical Journal* **86**, 212A-212A.

Nelson, A. (2004). "The development of strategies and methods for the synthesis of biologically active compounds." *New Journal of Chemistry* **28**, 771-776.

Paci, E., Friel, C.T., Lindorff-Larsen, K., Radford, S.E., Karplus, M. and Vendruscolo, M. (2004). "Comparison of the transition state ensembles for folding of Im7 and Im9 determined using all-atom molecular dynamics simulations with phi value restraints." *Proteins-Structure Function and Genetics* **54**, 513-525.

Pali, T., Whyteside, G., Dixon, N., Kee, T.P., Ball, S., Harrison, M.A., Findlay, J.B.C., Finbow, M.E. and Marsh, D. (2004). "Interaction of inhibitors of the vacuolar H^+-ATPase with the transmembrane V-o-sector." *Biochemistry* **43**, 12297-12305.

Parkinson, N.M., Conyers, C., Keen, J., MacNicoll, A., Smith, I., Audsley, N. and Weaver, R. (2004). "Towards a comprehensive view of the primary structure of venom proteins from the parasitoid wasp Pimpla hypochondriaca." *Insect Biochemistry and Molecular Biology* **34**, 565-571.

Patching, S.G., Brough, A.R., Herbert, R.B., Rajakarier, J.A., Henderson, P.J.F. and Middleton, D.A. (2004). "Substrate affinities for membrane transport proteins determined by C-13 cross-polarization magic-angle spinning nuclear magnetic resonance spectroscopy." *Journal of the American Chemical Society* **126**, 3072-3080.

Patching, S.G., Herbert, R.B., O'Reilly, J., Brough, A.R. and Henderson, P.J.F. (2004). "Low C-13-background for NMR-based studies of ligand binding using C-13-depleted glucose as carbon source for microbial growth: C-13-labeled glucose and C-13-forskolin binding to the galactose-H^+ symport protein GaIP in *Escherichia coli*." *Journal of the American Chemical Society* **126**, 86-87.

Roberts, R., Lister, I., Schmitz, S., Walker, M., Veigel, C., Trinick, J., Buss, F. and Kendrick-Jones, J. (2004). "Myosin VI: cellular functions and motor properties." *Phil. Trans. Roy. Soc. Lond. B* **359**, 1931-1944.

Rodgers-Gray, T.P., Smith, J.E., Ashcroft, A.E., Isaac, R.E. and Dunn, A.M. (2004). "Mechanisms of parasite-induced sex reversal in *Gammarus duebeni*." *International Journal for Parasitology* **34**, 747-753.

Ryan, D.E., Kim, C.H., Murray, J.B., Adams, C.J., Stockley, P.G. and Abelson, J. (2004). "New tertiary constraints between the RNA components of active yeast spliceosomes: A photo-crosslinking study." *RNA* **10**, 1251-1265.

Sessions, R.B., Thomas, G.L. and Parker, M.J. (2004). "Water as a conformational editor in protein folding." *Journal of Molecular Biology* **343**, 1125-1133.

Sharma-Oates, A., Quirke, P. and Westhead, D.R. (2004). "TmaDB: A tissue microarray database." *Journal of Pathology* **204**, 51A-51A.

Smith, A.J., Mair, L.A., Ponnambalam, S. and Sivaprasadarao, A. (2004). "Trafficking and functional defects associated with genetic mutations in the Kir6.2 subunit of K-ATP channels." *Biophysical Journal* **86**, 580A-580A.

Smith, K.M., Chen, X.Z., Cass, C.E., Baldwin, S.A., Karpinski, E. and Young, J.D. (2004). "Voltage clamp analysis of the presteady-state currents mediated by the human concentrative Na^+-nucleoside cotransporter hCNT3." *Biophysical Journal* **86**, 262A-262A.

Smith, K.M., Ng, A.M.L., Yao, S.Y.M., Labedz, K.A., Knaus, E.E., Wiebe, L.I., Cass, C.E., Baldwin, S.A., Chen, X.Z., Karpinski, E. and Young, J.D. (2004). "Electrophysiological characterization of a recombinant human Na^+-coupled nucleoside transporter (hCNT1) produced in Xenopus oocytes." *Journal of Physiology-London* **558**, 807-823.

Smith, M.C.A. and Thomas, C.D. (2004). "An accessory protein is required for relaxosome formation by small staphylococcal plasmids." *Journal of Bacteriology* **186**, 3363-3373.

Spence, G.R., Capaldi, A.P. and Radford, S.E. (2004). "Trapping the on-pathway folding intermediate of Im7 at equilibrium." *Journal of Molecular Biology* **341**, 215-226.

Street, A., Macdonald, A., Crowder, K. and Harris, M. (2004). "The hepatitis C virus NS5A protein activates a phosphoinositide 3-kinase-dependent survival signaling cascade." *Journal of Biological Chemistry* **279**, 12232-12241.

Tskhovrebova, L., Sader, K., Walker, M., Whiting, A. and Trinick, J. (2004). "Sarcomere structure studied in negatively stained and frozen- hydrated myofibrils." *Biophysical Journal* **86**, 567A-567A.

Tskhovrebova, L. and Trinick, J. (2004). "Properties of titin immunoglobulin and fibronectin-3 domains." *Journal of Biological Chemistry* **279**, 46351-46354.

Turnbull, W.B., Precious, B.L. and Homans, S.W. (2004). "Dissecting the cholera toxin-ganglioside GM1 interaction by isothermal titration calorimetry." *Journal of the American Chemical Society* **126**, 1047-1054.

Turnbull, W.B., Shimizu, K.H., Chatterjee, D., Homans, S.W. and Treumann, A. (2004). "Identification of the 5-methylthiopentosyl substituent in *Mycobacterium tuberculosis* lipoarabinomannan." *Angewandte Chemie-International Edition* **43**, 3918-3922.

Turner, A.J., Hiscox, J.A. and Hooper, N.M. (2004). "ACE2: from vasopeptidase to SARS virus receptor." *Trends in Pharmacological Sciences* **25**, 291-294.

Vickers, M.F., Zhang, J., Visser, F., Tackaberry, T., Robins, M.J., Nielsen, L.P.C., Nowak, I., Baldwin, S.A., Young, J.D. and Cass, C.E. (2004). "Uridine recognition motifs of human equilibrative nucleoside transporters 1 and 2 produced in *Saccharomyces cerevisiae*." *Nucleosides Nucleotides & Nucleic Acids* **23**, 361-373.

Walters, M.S., Hall, K.T. and Whitehouse, A. (2004). "The herpesvirus saimiri open reading frame (ORF) 50 (Rta) protein contains an AT hook required for binding to the ORF 50 response element in delayed-early promoters." *Journal of Virology* **78**, 4936-4942.

Wang, F., Thirumurugan, K., Stafford, W.F., Hammer, J.A., Knight, P.J. and Sellers, J.R. (2004). "Regulated conformation of myosin V." *Journal of Biological Chemistry* **279**, 2333-2336.

Wilkinson, D., Akumanyi, N., Hurtado-Guerrero, R., Dawkes, H., Knowles, P.F., Phillips, S.E.V. and McPherson, M.J. (2004). "Structural and kinetic studies of a series of mutants of galactose oxidase identified by directed evolution." *Protein Engineering Design & Selection* **17**, 141-148.

Williams, G.J., Nelson, A.S. and Berry, A. (2004). "Directed evolution of enzymes for biocatalysis and the life sciences." *Cell. Mol. Life. Sci.* **61**, 3034-3046.

Wilmot, C.M., Saysell, C.G., Blessington, A., Conn, D.A., Kurtis, C.R., McPherson, M.J., Knowles, P.F. and Phillips, S.E.V. (2004). "Medical implications from the crystal structure of a copper- containing amine oxidase complexed with the antidepressant drug tranylcypromine." *Febs Letters* **576**, 301-305.

Xiao, C., Tuthill, T.J., Kelly, C.M.B., Challinor, L.J., Chipman, P.R., Killington, R.A., Rowlands, D.J., Craig, A. and Rossmann, M.G. (2004). "Discrimination among rhinovirus serotypes for a variant ICAM-1 receptor molecule." *Journal of Virology* **78**, 10034-10044.

Xie, H., Patching, S.G., Gallagher, M.P., Litherland, G.J., Brough, A.R., Venter, H., Yao, S.Y.M., Ng, A.M.L., Young, J.D., Herbert, R.B., Henderson, P.J.F. and Baldwin, S.A. (2004). "Purification and properties of the *Escherichia coli* nucleoside transporter NupG, a paradigm for a major facilitator transporter sub-family." *Molecular Membrane Biology* **21**, 323-336.

Yegutkin, G.G., Salminen, T., Koskinen, K., Kurtis, C.R.P., McPherson, M.J., Jalkanen, S. and Salmi, M. (2004). "A peptide inhibitor of vascular adhesion protein-1 (VAP-1) blocks leukocyte-endothelial interactions under shear stress." *European Journal of Immunology* **34**, 2276-2285.