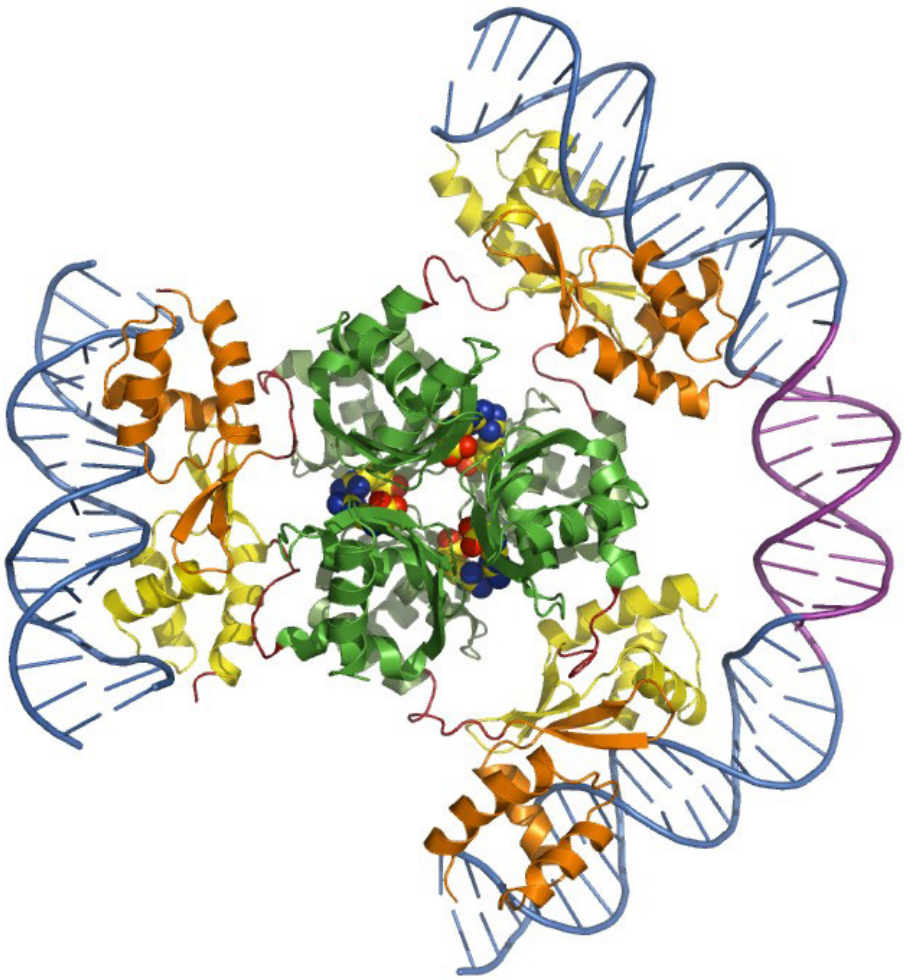
Astbury Centre for Structural Molecular Biology

UNIVERSITY OF LEEDS



ASTBURY CENTRE

Annual Report 2006

Front cover illustration: Model of the *Bacillus subtilis* hexameric transcriptional regulator holo-AhrC in complex with its *argC* operator. ARG boxes are coloured blue and the intervening DNA spacer is purple. For AhrC to bind to its operator site there must be a rearrangement in the N-terminal domains, orientating the wings and recognition helices into the correct position for DNA binding. The third ARG box is believed to be contacted through looping out the intervening DNA. (see page 41).

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

It's a great pleasure to write this Introduction to the Annual Report on the activities within the Astbury Centre during 2006. As I described on these pages last year, the only constant in life is change and this has been yet another hectic period in our development. The University Research Institutes, Molecular Biophysics (IMB) headed by Sheena Radford, and the Institute of Bionanosciences (IoB) headed by myself, have been rolled formally into the Astbury management structure, together with the new developments in Chemical Biology headed by Adam Nelson. With Structural Molecular Biology, these now form the major scientific themes of the Centre, although realistically we all expect that elements of everyone's work will cross backwards and forwards between these headings, as indeed they should in a true inter-disciplinary environment. Significantly, the attractions of the Astbury Centre have been recognised by a large number of newly appointed staff who joined the University over the last year across many different departments and Faculties. We also seem to have more collaborations outside our traditional Faculty base. Reflecting these changes and our increasing membership we have initiated additional meetings between the managing Executive Committee and members to facilitate communications and allowed ever wider participation in our activities.

The Astbury Society formed by our postdoctoral and postgraduate colleagues has also made major contributions to forging the personal links between laboratories, departments and disciplines essential to our future with a series of highly enjoyable social events, including our first Annual Astbury Lecture, given by Prof. Steve Chapman, Pro-Vice Chancellor for Research at Edinburgh University. Steve's exciting and highly amusing seminar was followed by a BBQ and "Sports Day" in which he was an enthusiastic and highly competitive participant (see photos on the Astbury Web Site). A pizza and beer Quiz Evening also sticks in the memory although the senior PI teams were clearly at a disadvantage in what was billed as a "music round".

Astbury members continue to be very successful in raising external grant income. A special mention here must go the award of yet more inter-disciplinary UK BBSRC PhD studentships to Astbury under the auspices of the SCIBS scheme. Many thanks to Adam Nelson and Bruce Turnbull for the hard work they put in to secure this funding on our behalf.

As well as welcoming new members, sadly we must also say goodbye to present and former colleagues. On a personal note I would like acknowledge the help and encouragement I always received from Simon Baumberg who sadly died in April 2007 after a brief illness. Simon had only just retired from the University where he had made major contributions to our understanding of the control of gene expression, his work spawning structural projects in the Stockley, Parsons and Phillips' laboratories, and latterly additional projects in collaboration with Kenny McDowall. He will be sadly missed. Alastair Smith is also leaving us towards the end of the year to pursue the development of a technical spin-out company, Avacta Ltd. Alastair has made major contributions helping to establish advanced biophysics in the University and we all wish him well in the future, especially if Avacta can be persuaded to support some Astbury seminars! Last year I reported that my former Deputy Director, Steve Homans had been appointed as Head of one of the new Institutes within the Faculty of Biological Sciences. It seems there is no holding him back and we congratulate him this year on his appointment as the new Dean in FBS from August 2007.

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, e.g. the developments in mass spectrometry for analysis of non-covalent self-assembling macromolecular complexes. 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.

The Centre produces a regular Newsletter describing its activities. If you would like to receive an electronic copy of this Newsletter it can be downloaded from (http://www.astbury.leeds.ac.uk/). This annual report is also available as a ~17MB PDF document that can be downloaded from our web site.

Finally I would like to thank our Editor, Alan Berry, once again for another outstanding job getting the Report together, ably assisted by Donna Fletcher.

Peter G. Stockley
Director, Astbury Centre for Structural Molecular Biology
Leeds, May 2007

Contents

Contributions indexed by Astbury Centre Principal Investigator

Mass spectrometry research group & facility

Victoria L. Homer, Victoria L. Morton and Alison E. Ashcroft

Overview of facility

The Mass Spectrometry (MS) Facility has a **LCT Premier** nanoelectrospray (ESI)-time-of-flight instrument with a NanoMate automated sampling handling facility and Ion Mobility Spectrometry interface, a **Q-Tof** orthogonal acceleration quadrupole-time-of-flight tandem instrument with nanoESI and on-line capillary HPLC, a **Platform II** ESI-quadrupole instrument with on-line HPLC, and a surface enhanced laser desorption ionisation/matrix assisted laser desorption ionisation (SELDI/MALDI) **ProteinChip** mass spectrometer.

The Mass Spectrometry Facility operates a sample analysis service in addition to having an active research group which is involved in several key projects within the Astbury Centre for Structural Molecular Biology and the Faculty of Biological Sciences, and also with external collaborators.

Research

The research involves the application of mass spectrometry to the structural elucidation of biomolecules and biomolecular processes and can be summarised as follows:

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. In collaboration with Prof. Sheena Radford's group, ESI-MS is being used to characterise amyloidogenic proteins using charge state distribution analysis, proteolysis and H/D exchange to gain insights into folding mechanisms.

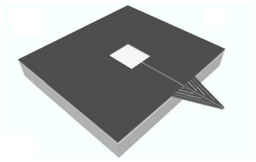
ii). Protein-ligand non-covalent interactions and macromolecular assembly. ESI-MS is being used to investigate non-covalently bound macromolecular structures. Such studies include protein-peptide, protein-protein, and protein-RNA complexes. The latter are important in virus assembly, an area we are investigating with respect to the MS2 and Qβ systems in collaboration with Prof. Peter Stockley and Dr Nicola Stonehouse. Protein-protein macromolecular complexes are critical species in fibrillogenesis and are under investigation as an integral part of our β_2-microglobulin amyloid studies with Prof. Sheena Radford (see above).

iii). Reaction monitoring. We use mass spectrometry to measure the uptake of ATP by the muscle protein myosin, in collaboration with Prof. Howard White, Eastern Virginia Medical School, USA. In a collaborative project with Prof. Gabriel Waksman and Dr Han Remaut of Birkbeck College, London, we are carrying out an in-depth investigation into Saf fibre formation and P pilus assembly by the chaperone-usher pathway.

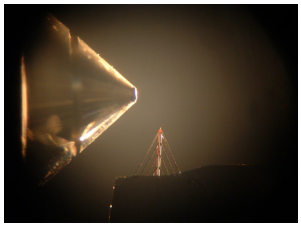
iv). Structural elucidation and proteomics. Tandem MS (MS/MS) sequencing of proteins and peptides is an important bioanalytical technique. Several proteomics-related projects are in progress, including a study into the functional analysis of preproneuropeptide genes from the Drosophila genome with Prof. Elwyn Isaac, and an investigation into the peroxisomal protein import machinery with Dr Alison Baker. We are also developing methodology for membrane protein analysis and structural elucidation in collaboration with Prof. Peter Henderson.

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 oligomers, and for the structural characterisation of small peptides. Additionally we are collaborating with Dr Steve Arscott at the Institute of Electronics, Microelectronics and Nanotechnology (CNRS France) to evaluate their novel miniaturized electrospray sources, developed with the ultimate aim of producing lab-on-a-chip devices (see Figure).

(a)



(b)



(c)

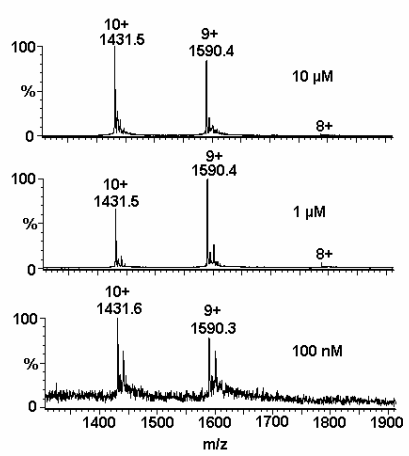


Figure: (a) nanoelectrospray emitter tip design; (b) nanoelectrospray emitter tip positioned orthogonal to the sample cone of the Q Tof 1 mass spectrometer; (c) m/z spectra of lysozyme (14,305 Da) obtained using the nanoelectrospray emitter tip with the Q-Tof 1 mass spectrometer showing the 8+, 9+ and 10+ charge states for a dilution series of the protein[(4)].

Publications

Smith, A.M., Jahn, T.R., Ashcroft, A.E. & Radford, S.E. (2006) Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using mass spectrometry, *J. Mol. Biol.,* **364,** 9-19.

Remaut, H., Rose, R.J., Hannan, T., Hultgren, S.J., Radford, S.E., Ashcroft, A.E. & Waksman, G. (2006) Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta-strand displacement mechanism, *Molecular Cell,* **22,** 831-842.

Myers, S.L., Thomson, N.H., Radford, S.E. & Ashcroft, A.E. (2006) Investigating the structural properties of amyloid fibrils formed from β_2-microglobulin by limited proteolysis and mass spectrometry, *Rapid Commun. Mass Spectrom.,* **20,** 1628-1636.

Arscott, S., Gaudet, M., Brinkman, M., Ashcroft, A.E. & Blossey, R. (2006) Capillary filling of miniaturised sources for electrospray mass spectrometry, *J. Phys. Condensed Matter,* **18**, S677-S690.

Funding

This work was funded by The Wellcome Trust, the BBSRC, Waters Corpn., AstraZeneca, Pfizer and Syngenta.

Directed evolution of new enzyme activities

Lorna Farnsworth, Matthew Edmundson, Amanda Bolt, Adam Nelson and Alan Berry

We are using the power of directed evolution in a number of enzyme systems to alter the specificity and stereochemistry of enzymes to produce novel catalysts. One system involves the enzyme, *N*-acetylneuraminic acid lyase (NAL), which catalyses the reaction of *N*-acetylmannosamine and pyruvate to give *N*-acetylneuraminic acid (Fig 1). NAL exhibits poor facial selectivity, giving a mixture of stereoisomers at the 4 position and its use as a catalyst in synthetic chemistry is therefore limited.

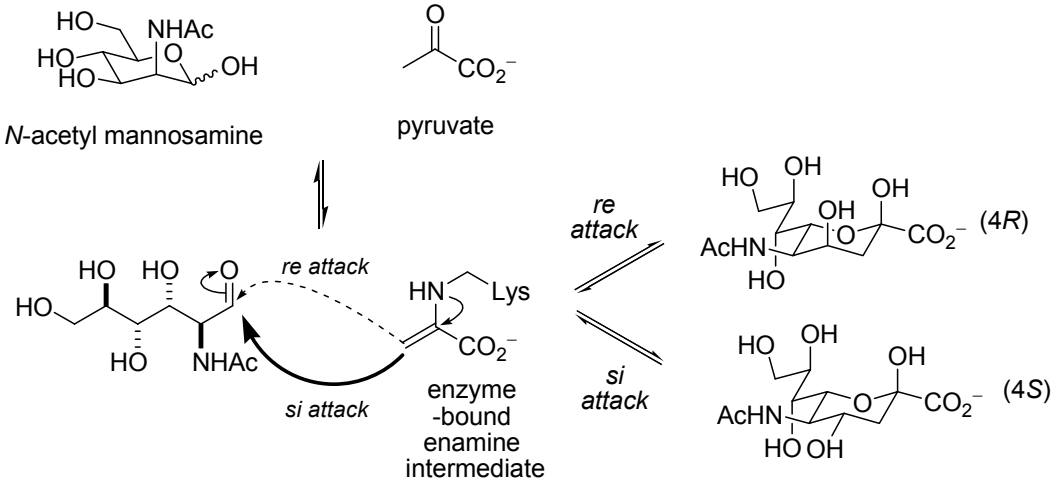


Fig 1. The stereochemical course of the carbon-carbon bond forming reaction catalysed by NAL depends on which face of the carbonyl is attacked by the enamine.

We have used directed evolution to create a pair of stereochemically complementary mutant NALs with broad substrate specificity for the synthesis of sialic acid mimetics, (general structure shown in Fig 2), using a combination of error-prone PCR and site-directed mutagenesis.

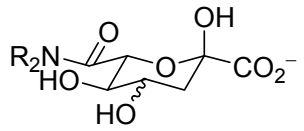
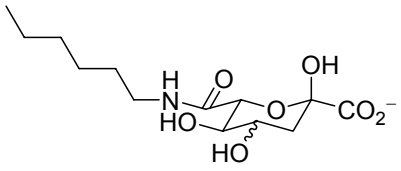


Fig 2. The general structure of the sialic acid mimetics. The screening substrate was the dipropyl amide where R=Pr

The complementary 4*R*- and 4*S*-selective variants, E192N/T167G and E192N/T167V/S208V, both gave ~50-fold selectivity towards the cleavage of the desired 4*S*- and 4*R*-configured products, respectively. The premise of our strategy was that variants which were stereoselective in the cleavage direction would be stereoselective in the synthetic direction and therefore useful biocatalysts. Reactions catalysed by the unselective variant E192N, and the 4*R*- and 4*S*-selective variants were monitored in the synthetic direction by NMR to confirm the stereoselectivity of the variants. The results (Fig 3), confirmed that the variants did exhibit excellent stereoselectivity ($>95:<5$). The variants were then used on a synthetic scale and excellent yields (66% and 70%) and diastereoselectivities ($>98:<2$) were achieved.

The evolved stereoselective enzymes were found to be far less efficient at accepting secondary amides than tertiary amides and the D191X, E192X and S208X saturation libraries were therefore re-screened with the secondary amide shown to find more active enzymes.



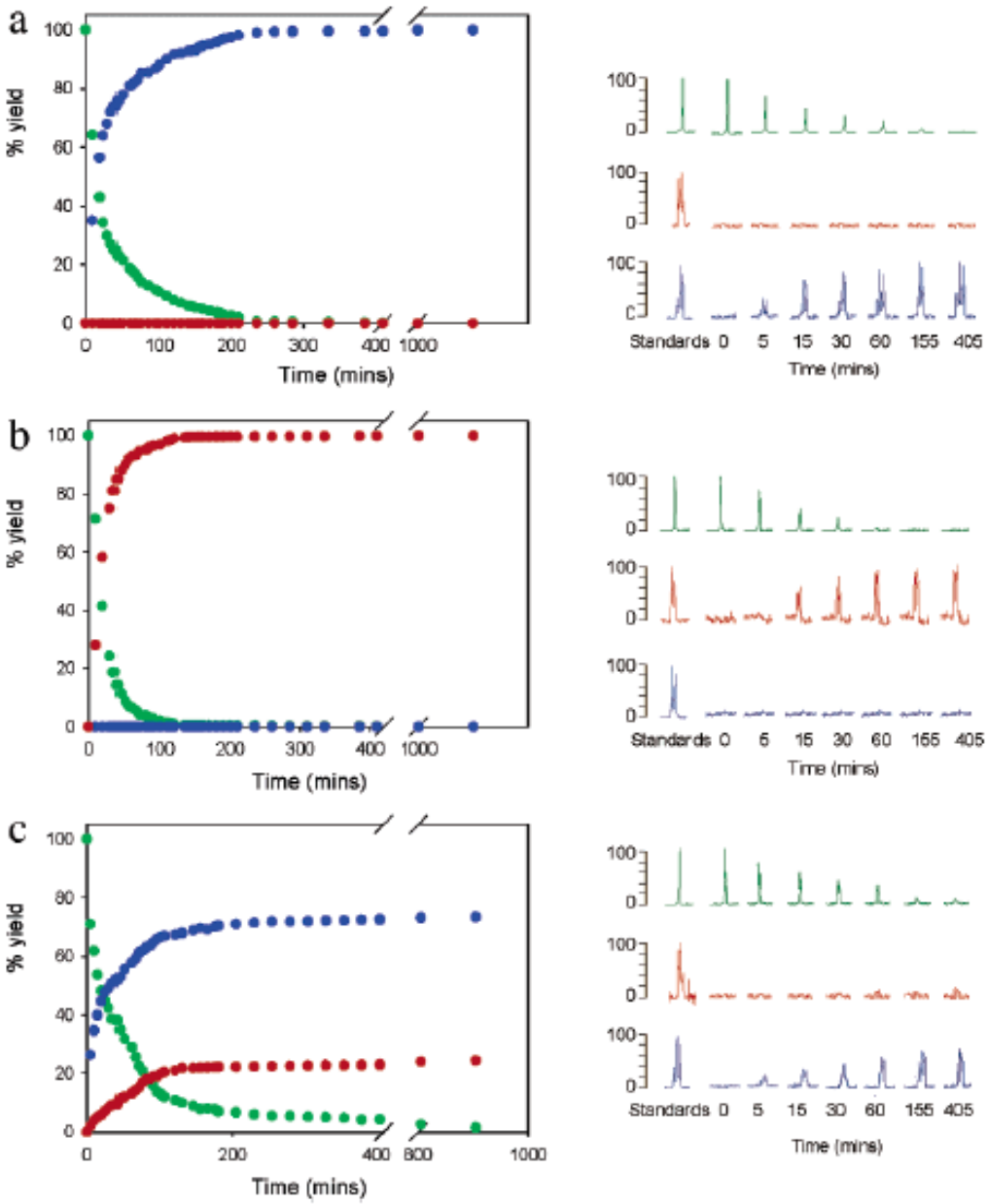


Figure 3. Time course of the stereo-controlled synthesis of dipropylamide analogues of sialic acid using the evolved enzymes. Panel A: Time course with the evolved *S*-specific enzyme; Panel B: Time course with the evolved *R*-specific enzyme; Panel C: Time course with the unselective variant E192N. Green represents aldehyde, red represents the *S*- stereoisomer and blue represents the *R*- stereoisomer

Hits were identified in all three libraries and the most active variant, E192W, is currently being further characterised. The engineering of a non-selective aldolase into a pair of complementary biocatalysts is of enormous interest to synthetic chemists and as the residues identified as critical for stereoselectivity are conserved among members of the NAL superfamily, the approach might be extended to the evolution of other useful biocatalysts for the stereoselective synthesis of biologically active molecules.

Publications

Williams, G.J., Woodhall, T., Farnsworth, L.M., Nelson, A. & Berry, A. (2006) Creation of a pair of stereochemically complementary biocatalysts. *J. Am. Chem. Soc.*, **128**, 16238-16247.

Funding This work was funded by BBSRC and The Wellcome Trust

Mechanical unfolding of proteins using dynamic force spectroscopy

Kirstine Anderson, Colin Grant, Eleanore Hann, James Pullen, David Sadler, Alastair Smith, Sheena Radford and David Brockwell

Determination of the factors that confer mechanical resistance to globular proteins

On the nanoscale, mechanical deformation of biomolecules can result in rapid unfolding and a loss of function. Conversely, the ability to sense an applied force is vital in many signal transduction pathways. Over the past few years, we have used a combination of force-mode atomic force microscopy (AFM) and both all atom, and course grained simulations, to investigate which factors define the mechanical stability of proteins. Whilst the role that secondary structure plays in mechanical resistance is relatively well understood, the effects of other factors such as side-chain packing are not. Work is now being carried out to examine which other factors may play a role in mechanical resistance, as well as characterising the unfolded transition state of a model protein.

Protein L is a 62 amino-acid protein with a simple β-grasp topology that displays significant resistance to mechanical unfolding. However ubiquitin, which has a similar fold to protein L, unfolds at a significantly higher force, suggesting that other factors, in addition to the type and geometry of secondary structural units, are important in determining the mechanical response of proteins. To investigate this observation in greater detail, a systematic mutational analysis is being carried out on protein L using variants that have been designed to alter the sidechain contacts in the native protein. This method will also allow the mechanical transition state of the wild type protein to be visualised and allow benchmarking of molecular dynamics simulations of the mechanism of mechanical unfolding of this protein.

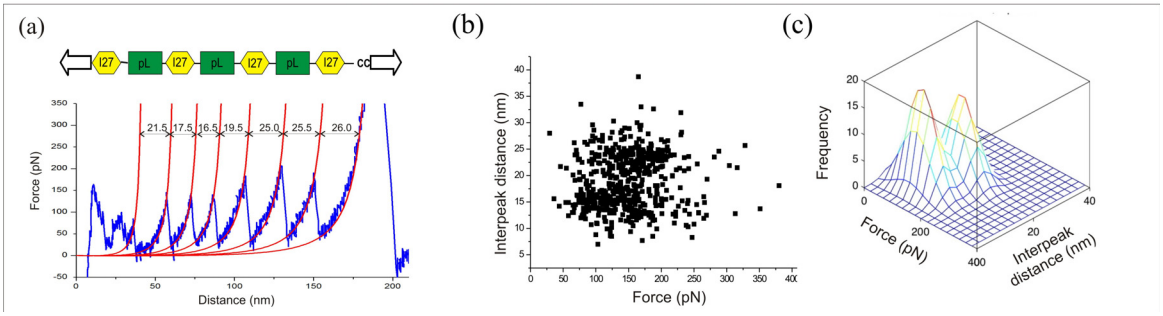


Figure 1: mechanical unfolding of protein L within a heteropolymer. (a) Mechanical unfolding trace obtained on unfolding $(I27)_4(Im9)_3$ illustrating the mixed interpeak distances and different forces obtained upon unfolding the individual domains. (b) Scattergram of force versus interpeak distance shows significant overlap between each type of monomer which can be denconvoluted by the construction and fitting of 3D histograms (c).

The geometry of the applied force strongly affects the apparent mechanical strength of proteins and this anisotropy may be exploited in nature (see below). Indeed, recent work by West *et al*. suggested that protein mechanical energy landscapes (*i.e*. a map of the force at which a protein unfolds when extended by every pairwise combination of residues) are highly anisotropic and may provide a common underlying mechanism linking *in vitro* and *in vivo* unfolding. However, experimentally, altering the points by which a protein is unfolded is technically challenging and the few successful examples in the literature have not employed a generic method. A new linkage method is now being developed in which it will be possible to mechanically unfold any protein by any number of different directions.

How are proteins mechanically denatured *in vivo*?

The forced denaturation of proteins is vital to the translocation of proteins across the mitochondrial membrane and in their degradation by energy dependent proteases. Both

processes require that the protein substrate is tagged then rapidly unfolded by the application of mechanical force. What is the relationship between mechanically unfolding proteins using the AFM and the process that occurs *in vivo*? Simulations suggest that the two processes occur through distinct pathways but that both could be extracted from a common underlying mechanical landscape. Work is now underway to test these observations experimentally.

As well as degrading proteins, the energy-dependent ATPase components of these molecular 'unfoldases' have been observed to rapidly break apart strong protein complexes, probably by the application of a small force to one member of the protein pair. To test this theory the mechanical unfolding of Im9 was measured in the presence and absence of its high affinity protein ligand, E9. Surprisingly, binding of E9 offered little mechanical stabilisation to Im9 despite the high avidity of the complex (K_d ~10 fM), highlighting how the unfolding of a mechanically weak protein partner can be used to break a strong protein : protein interaction within the cell. The molecular determinants of this high affinity are now being studied directly by measuring the force required to rupture the Im9 : E9 interface.

Collaborators

Colin Kleanthous, University of York; Godfrey Beddard, School of Chemistry, University of Leeds; Peter Olmsted, Emanuele Paci, and Dan West, School of Physics and Astronomy, University of Leeds.

Publications

West, D., Brockwell, D., Olmsted, P., Radford, S. & Paci, E. Mechanical resistance of proteins explained using simple molecular models. *Biophys J*. (2006) **90**, 287-297.

West, D., Brockwell, D. & Paci, E. Prediction of the translocation kinetics of a protein from its mechanical properties. *Biophys J.* (2006). **91**, L51-53.

Hann, E. Kirkpatrick, N., Kleanthous, C., Smith, D.A., Radford, S. & Brockwell, D. The effect of protein complexation on the mechanical stability of Im9. *Biophys J.* Submitted.

Funding

We thank Keith Ainley for technical support and the BBSRC, EPSRC, Wellcome Trust and University of Leeds for funding. DJB is an EPSRC funded White Rose Doctoral Training Centre lecturer.

Structural studies of the motor protein dynein

Stan Burgess, Matt Walker, Anthony Roberts and Peter Knight

Introduction

Dynein is a family of minus-end directed microtubule motors that function in an incredible diversity of cellular processes in eukaryotes. These include the transportation of numerous membrane-bound vesicles, mitochondria and mRNA, as well as maintaining the spatial positioning of the nucleus and the Golgi apparatus. Dyneins are also the motors responsible for force generation in the specialized microtubule apparatus within motile cilia and flagella, the most common form being the 9+2 axoneme. Here, sliding of the outer nine doublet microtubules driven by dynein activity is somehow coupled to propagated-bending waves along the entire superstructure.

Dynein is one of three families of molecular motors, the others being kinesin and the actin-based motor myosin, and by far the least well understood. Dynein is ~ ten times larger than the other microtubule motor kinesin and has an evolutionary origin within the AAA+ superfamily of mechanoenzymes, unlike kinesin and myosin.

Little is known about dynein's structure. All our knowledge comes from electron microscope (EM) studies. We showed previously by negative stain EM that dynein has a stalk-head-tail structure and that the head and stalk rotate relative to the cargo-carrying tail during the motor's power stroke (see Fig. 1a). The head is ring-like and contains six AAA+ domains, four of which are capable of binding ATP and at least two of which can hydrolyze it. ATP hydrolysis fuels this motor by driving the conformational changes associated with the power stroke and those governing its binding to, and release from, the microtubule track.

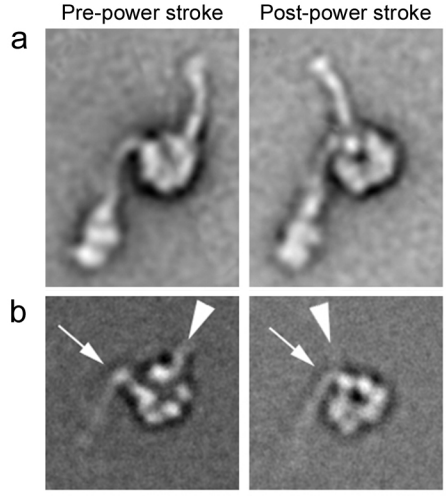


Figure 1
Dynein molecules revealed by electron microscopy and single-particle image processing. Dynein is shown in two different nucleotide conditions which produce the pre- and post-power stroke conformations. (a) Negative stain images showing the stalk-head-tail structure and (b) corresponding frozen-hydrated views of the head showing the emergence points of the stalk (arrowhead) and tail (arrow). Image width is 46 nm.

We have been pursuing the three-dimensional (3D) structure of dynein by cryo-EM. This involves imaging thousands of frozen-hydrated molecules embedded in vitreous water and using single-particle computer image processing to obtain its 3D structure. We have collected images for both pre-, and post-power stroke conformations (Fig. 1b), which show excellent agreement with images from stained and dried molecules. Using recombinant dynein motor domains fused with green fluorescent protein at a variety of positions we have also begun the first sub-domain mapping studies to identify the topology of AAA+ domains within the force-generating head.

Publications

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

Funding

This work was funded by the BBSRC including a Japan Partnering Award.

Acknowledgements

Cryo-EM studies were done in Leeds in collaboration with Prof. Kazuhiro Oiwa and Dr. Hitoshi Sakakibara at the Kansai Advanced Research Centre, Kobe, Japan. Recombinant mapping studies were also done in Leeds in collaboration with Prof. Kazuo Sutoh, University of Tokyo, Japan.

Solution structure of the Vts1 SAM domain bound to RNA

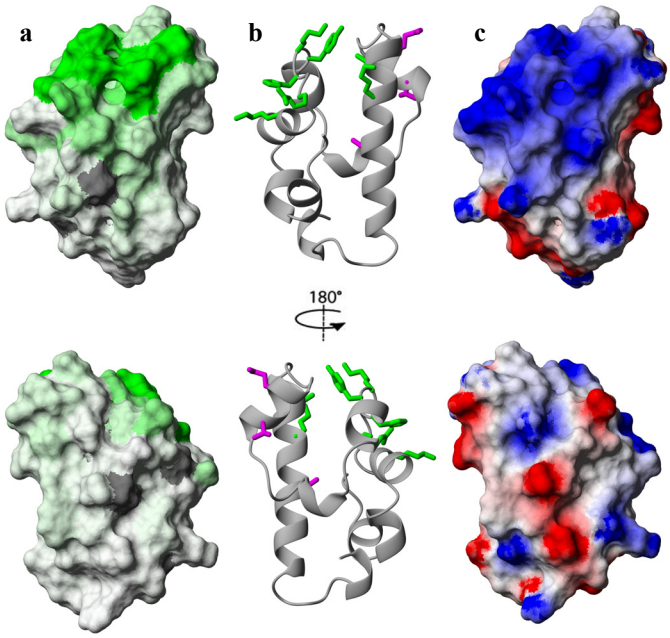
Thomas Edwards

Introduction

The SAM (sterile alpha motif) domain is a common protein–protein interaction module, identified in a wide variety of eukaryotic proteins, including the signal transducing Eph receptors, the transcriptional repressors TEL and polyhomeotic (Ph), and splice variants of the tumor suppressor p53. Recently, we and others showed that the SAM domain of Smaug (Smg), a protein required for proper abdominal segmentation in early *Drosophila melanogaster* embryos, encodes a novel protein–RNA, rather than a traditional protein–protein, interaction module. Although structures are known for several SAM domains that mediate protein–protein interactions, the only structure known for a SAM domain that mediates protein–RNA interactions is the Smg RNA-binding domain, consisting of a SAM domain that interacts specifically with RNA. This finding was bolstered by biochemical analysis of a Smaug homologue in *Saccharomyces cerevisiae*, Vts1, which has a standalone SAM domain and therefore is a more typical member of the new class of RNA-binding SAM domains.

Structure of Vts1

To elucidate the structural basis for RNA binding, the solution structure of the Vts1 SAM domain, in the presence of a specific target RNA (5'-GGCUCUGGCAGCC-3'), has been solved by multidimensional heteronuclear NMR spectroscopy. The Vts1 SAM domain retains the "core" five-helix-bundle architecture of traditional SAM domains, but has additional short helices at N and C termini, comprising a small substructure that caps the core helices. The RNA binding surface of Vts1, determined by chemical shift perturbation, maps near the ends of three of the core helices, in agreement with mutational data and the electrostatic properties of the molecule. These results provide a structural basis for the versatility of the SAM domain.



(**a**) Chemical shift changes are color-coded onto the surface of Vts1. Green corresponds to DdR3 ppm, white corresponds to dZ0, and intermediate values are linearly interpolated. (**b**) Mutations in either Vts1 or Smg that reduce RNA binding are shown in green (Lys461, Arg465, His466, Lys467, Leu498, Ala497), while for comparison, a subset (Thr487, Glu489 and Ala505) of mutations that do not perturb RNA binding are shown in magenta to aid in delineating the RNA binding surface. (**c**) Electrostatic potential at the surface of the Vts1 SAM domain determined using MOLMOL. The orientation of the Vts1 SAM domain is the same in the top panels of (**a**–**c**); all structures are rotated 180° in the lower panel.

Collaborators

Aneel Aggarwal (Mount Sinai Medical School), Art Palmer III (Columbia University), Robin Wharton (HHMI, Duke University Medical School).

Funding

This work was funded by the NIH. Funding for future projects is available from the Royal Society, European Union, Yorkshire Cancer Research, and the University of Leeds.

Publications

Edwards, T.A., Butterwick, J.A., Zeng, L., Gupta, Y.K., Wang, X., Wharton, R.P., Palmer, A.G. & Aggarwal, A.K. (2006). Solution structure of the Vts1 SAM domain in the presence of RNA. *J. Mol. Biol.* **356,**1065-72.

Lipocalin receptors

Clara Redondo, Renske Hesselink, Mari Lopez Ruiz and John Findlay

Introduction

The research interests of our group are concerned with the biological role and mechanism of action of lipocalin receptors as a platform for understanding their structure/function relationships and the interaction networks through which they fulfil their biological role. The lipocalins are a large family of small proteins that occur throughout all eukaryotic life forms, and bind small ligands with a range of specificities (retinoids, pheromones, odours). Lipocalins and their ligands play critical roles in crucially important biological processes such as vision, reproduction, olfaction, stress responses, development, inflammation, infection control and so on. Almost 20 years ago, we produced evidence that at least some of these functions were mediated through membrane-bound receptors. That for retinol binding protein was characterised and shown to mediate the uptake of vitamin A. Since then, further data have confirmed the presence of these receptors for a number of lipocalins in a variety of biological systems. The first of these receptors to be cloned - that for tear lipocalin - was shown to be a member of a quite new family of proteins (Figure 1A). Most recently, the one for the retinol-binding-protein (RBP) has also been described and shown to possess all the features described in our earlier reports (Figure 1B). It too belongs to yet another new family, the only previously know property being its induced expression by retinoic acid.

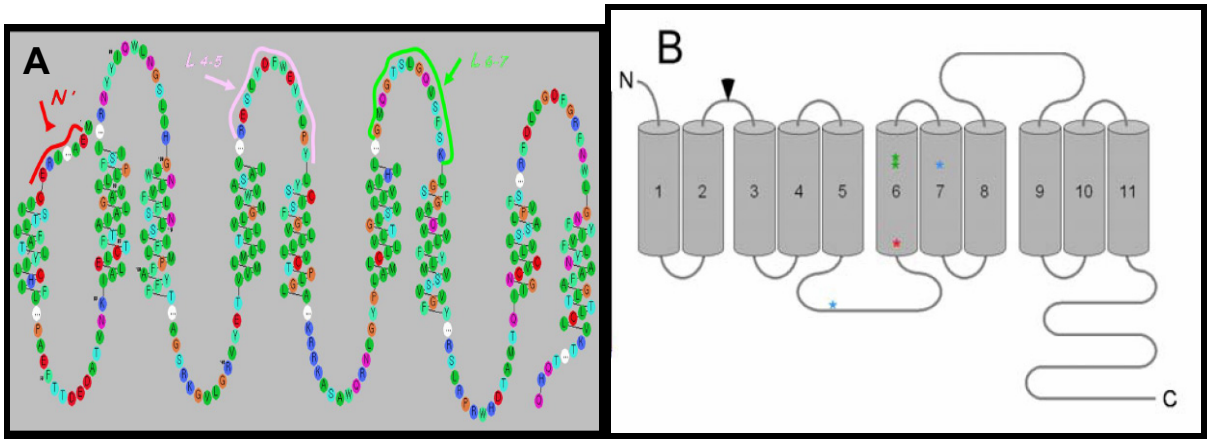


Figure 1. Putative Lipocalin Receptor Topology. Panel A for Lcn 1, and Panel B. for RBP

The RBP receptor-system

The RBP : RBP receptor system has been studied only in this laboratory. Recent work has demonstrated that as well as being the receptor for RBP on the outside of the cell, it also appears to be the receptor for cellular RBP at the cytosolic surface.

Using surface plasmon resonance, and observing the binding phenomenon of the two purified lipocalins with preparations of solubilised HEK membranes, we obtained binding of both soluble partners to a protein entity in the membrane preparations, but only when RBP was conjugated with retinol and CRBP was in its free form. These results supported our hypothesis on the mechanism of action of the system *in vivo* (Figure 2). Further, domain swapping experiments both confirmed the site on RBP responsible for its interaction with the receptor and demonstrated that occupancy of the binding site on a heterologous lipocalin could signal to the RBP domain thereby generating a high affinity conformation for receptor interaction.

Current work utilising 2-hybrid systems for integral membrane proteins has identified receptors, ion channels and soluble proteins which can interact with RBP and CRBP (used as

baits). In the light of the recent dramatic paper that RBP may be heavily involved in causing insulin resistance (and hence its development into type 2 diabetes), this work promises to reveal new receptors and pathways important in homeostatic regulation and as drug targets.

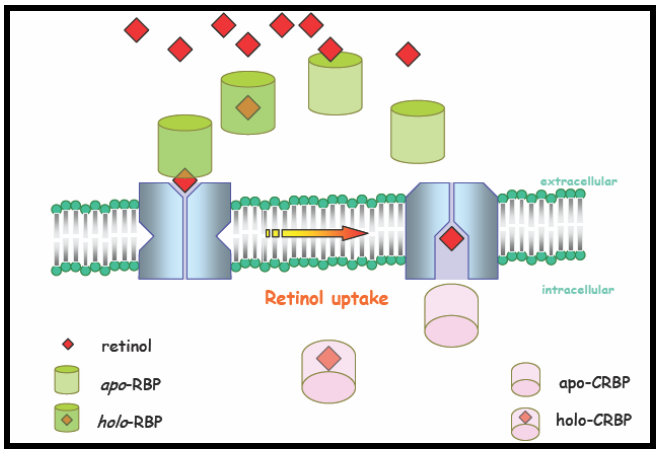


Figure 2. The mechanism of RBP/CRBP-receptor binding

Lipocalin-1 interacting membrane protein (LIMR)

The cloning of the first validated receptor for a lipocalin was achieved via the demonstration of an interaction between tear lipocalin (Lcn1) and a phage bearing what turned out to be the N-terminal putative extracellular domain of the intact receptor. Subsequent work confirmed this region as the receptor site for lipocalin binding. We are expressing this receptor, and that for RBP, in transient and stable Drosophila S2 cultures to determine the oligomeric structure, topography and key functional sites in both proteins, using a combination of biochemical, cell biological and biophysical techniques The study will also begin to reveal the interacting networks in which these receptors participate.

Publications

Redondo, C., Burke, B.J. & Findlay, J.B.C. (2006) The retinol binding protein system: a potential paradigm for steroid binding globulins? *Horm. Metab. Res* **38**, 269-78.

Burke, B.J., Redondo, C., Redl, B. & Findlay, J.B.C. (2005) "Lipocalin Receptors: Into the Spotlight" in Lipocalins (Akerström, B., Borregaard, N., Flower Edward, D.R. and Salier, J.P., eds.) Eureka Biosciences.

Funding

This work was funded by BBSRC and the EU Marie Curie programme.

Studies on the functions of hepatitis C virus proteins

Steve Griffin, Anna Nordle, Andrew Milward, Jamel Mankouri, Sarah Gretton, Corine St.Gelais, Philip Tedbury, Mair Hughes, Lynsey Corless 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, however, recently a clone of the virus that is able to replicate in cell culture has been identified and we are using this model extensively to understand how viral protein products function within the viral lifecycle.

Current projects include the study of the p7 protein – we previously demonstrated that this protein forms an oligomeric cation channel. Recent work has analysed the oligomeric nature of p7 in more detail and defined it as a heptamer. We are currently generating mutants of p7 in the context of the cell culture infectious clone of HCV in order to analyse the role of p7 in the virus lifecycle. Additionally we have established a robust liposome based *in vitro* assay to measure p7 function that may be useful in the search for novel p7 inhibitors.

We are analysing a novel auto-proteolytic event in the virus lifecycle – the cleavage between the NS2 and NS3 proteins. We have shown that this protease requires zinc and have identified a key cysteine residue that is required for cleavage, current work is addressing the role of this cleavage event in viral replication.

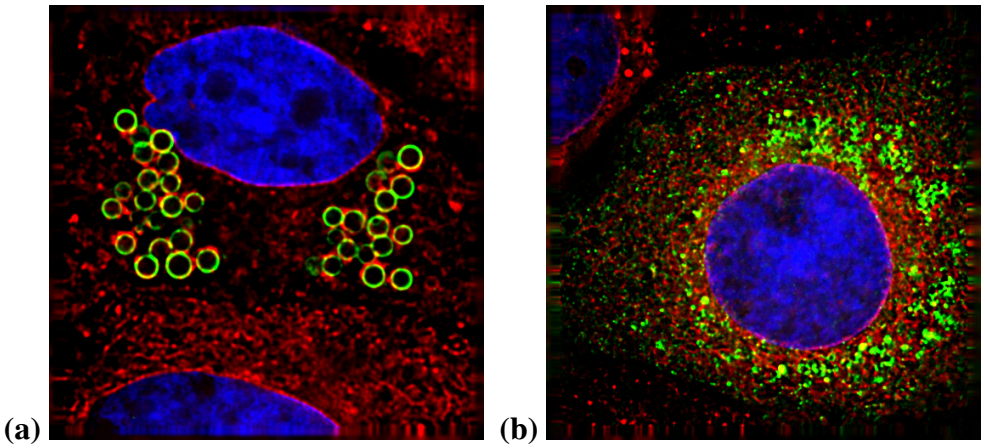


Figure 1: Distribution of the HCV Core protein **(a)** or NS5A protein **(b)**, in human hepatoma cells infected with the JFH-1 strain of HCV. The HCV proteins were detected using specific antibodies followed by FITC-conjugated secondary antibodies (green), the ER was detected using Texas Red–conjugated ConA (red). Note the association of Core with the surface of lipid droplets. Image taken by Dr Steve Griffin in the Wellcome

Thirdly we are studying the role of the NS5A protein, firstly in terms of perturbation of cellular signalling pathways - we have previously shown that NS5A perturbs two key mitogenic pathways within the cell; it inhibits the Ras-ERK MAP kinase pathway and stimulates PI3K signalling pathways, the latter promotes cell survival and activates the proto-oncogene β-catenin with implications for the link between HCV and hepatocellular carcinoma. We are currently funded by the MRC and Yorkshire Cancer Research to analyse these signalling events in more detail. We are also investigating the role of NS5A in virus replication. We are addressing the role of NS5A phosphorylation in viral replication initially by using mass spectrometry to identify phosphorylation sites. We are also generating a series of NS5A mutants – particularly in PxxP motifs that interact with cellular SH3 domain – to

characterise the role of these interactions in viral replication. Lastly we have inserted into NS5A either GFP or a protein domain that is efficiently biotinylated in eucaryotic cells (from the Propionibacterium shermanii transcarboxylase enzyme). This will allow us to visualise or purify the viral RNA replication complexes with a view to undertaking a proteomic analysis of this complex.

Publications:

Clarke, D, Griffin, S.D., Beales, L., Burgess, S., StGelais, C., Harris M. & Rowlands D. (2006) Evidence for the formation of a heptameric ion channel complex by the HCV p7 protein *in vitro*. *J. Biol. Chem.* **281**, 37057-68.

Tedbury, P. & Harris, M. (2007) Characterisation of the role of zinc in the hepatitis C virus NS2/3 auto-cleavage and NS3 protease activities. *J. Mol. Biol.*, **366**, 1652-1660

McCormick, C.J., Brown, D., Griffin, S., Challinor, L., Rowlands, D.J. & Harris, M. (2006) A link between translation of the hepatitis C virus polyprotein and polymerase function; possible consequences for hyperphosphorylation of NS5A. *J. Gen. Virol,* **87,** 93-102.

McCormick, C.J., Maucourant, S., Griffin, S., Rowlands, D.J. & Harris, M. (2006) Tagging of NS5A expressed from a functional HCV replicon. *J. Gen. Virol,* **87**, 635-640.

Collaborators:

Kalle Saksela, University of Helsinki, Finland
John McLauchlan, MRC Virology Unit, Glasgow
Phil Tedbury hold a Glaxo-Smith-Kline-sponsored CASE studentship (NS2/3 cleavage), Corine StGelais holds a Pfizer CASE studentship (p7) and Mair Hughes holds an Arrow Therapeutics CASE studentship (NS5A).

Funding:

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

Studies on the HIV-1 Nef protein

Matthew Bentham and Mark Harris

HIV-1 Nef is a 205 amino acid N-terminally myristoylated protein that plays a critical role in viral pathogenesis. Myristoylation is an eukaryotic specific co-translational modification that is catalysed by a ribosomal associated enzyme - N-myristoyltransferase (NMT). Two projects are ongoing. Firstly, we are attempting to understand the mechanisms by which Nef interacts with cellular membranes – using a combination of *in vitro* liposome binding assays and sucrose gradient fractionation of lysates from Nef-expressing cells we have determined that both the myristate and basic amino acids near the N-terminus of the protein are required. In further experiments, we have anchored Nef irreversibly to the membrane with a C-terminal farnesyl tag and shown that this abrogates the ability of Nef to down-modulate cell-surface CD4.

Secondly, we are raising RNA aptamers to native, myristoylated Nef. The latter can be co-expressed in *E.coli* with N-myristoyl transferase to generate large amounts of purified myristoylated Nef. This has been used to select pools of randomised RNA aptamers and we are currently in the process of characterising these aptamers. They will be tested for the ability to inhibit Nef functions – both *in vitro* (using ELISA based protein-protein interaction assays to measure effects on the interactions of Nef with cellular SH3 domains or CD4 cytoplasmic tail), and *in vivo*. The latter include FACS analysis of down-modulation of cell surface CD4 by Nef, and effects of Nef on virus replication - these experiments are being carried out in the Category III containment facility.

Publication:

Bentham, M., Mazaleyrat S. & Harris, M. (2006) A cluster of arginine residues near the N-terminus of the HIV-1 Nef protein is required both for membrane association and CD4 down-modulation. *J. Gen. Virol.,* **87**, 563-571.

Collaborators:

Kalle Saksela, University of Tampere, Finland
Matthias Geyer, Max Planck Institute, Dortmund
J Victor Garcia, University of Texas at Dallas

Funding:
This work is funded by the MRC.

The contributions of backbone protein dynamics and ligand reorientation to the thermodynamics of ligand binding by arabinose binding protein

Christopher MacRaild, Agnieszka Bronowska and Steve Homans

Protein dynamics make important yet poorly understood contributions to molecular recognition and ligand binding events. In order to estimate these, we measured changes in fast (picosecond-to-nanosecond scale) protein dynamics which accompany the interaction of the arabinose binding protein (ABP) with its ligand, D-galactose, using NMR relaxation and molecular dynamics (MD) simulation. These two approaches present an entirely consistent view of the dynamic changes which occur in the protein backbone upon ligand binding.

A common view of ligand-protein interactions sees them arising from the shape complementarity between the protein and ligand together with the hydrophobic effect (exclusion of water from hydrophobic surfaces). It is generally assumed that upon binding the ligand molecule adopts a single conformation, and thus decreases its entropy (this entropy loss reflects the loss of conformational flexibility the ligand possesses in solution). Likewise, the protein binding site might be expected to undergo a loss of entropy as it adopts only that subset of available conformations which are relevant to binding of the ligand. This loss of internal conformational entropy of the ligand is challenging to assess, particularly for a ligand such as D-galactose which displays very complex conformational behaviour in solution.

The contribution of protein dynamics to the ABP-D-galactose interaction was assessed by means of Lipari-Szabo analysis of nuclear magnetic relaxation. The results of this formalism can be interpreted in terms of the conformational entropy associated with the measured

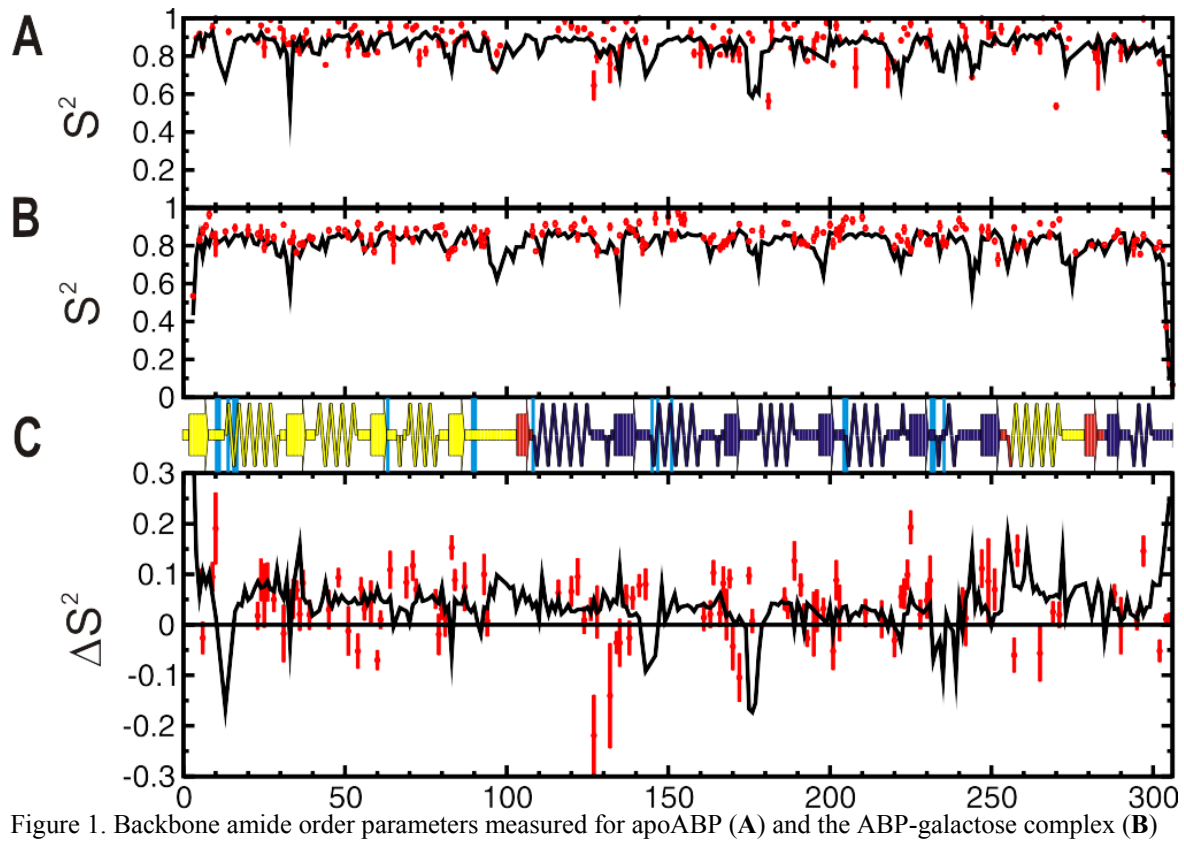


Figure 1. Backbone amide order parameters measured for apoABP (**A**) and the ABP-galactose complex (**B**) from NMR relaxation data (red) and calculated from MD simulation (black). The change in order parameter ($\Delta S^2 = S^2_{apoABP} - S^2_{ABPgal}$) is in **C**, with secondary structure elements identified above. The N domain is coloured yellow, the C domain dark blue, and the domain hinge regions red. Residues directly involved in the ligand interaction are highlighted in cyan.

motions. The order parameters (S2), which are a measure of the extent of angular motion of individual bond vectors, can be obtained from relaxation data and MD simulation trajectories.

In both NMR experiments and MD simulations, we saw almost exclusively the large order parameters, which is typical for relatively rigid, globular proteins. This happened for both apo-ABP and ABP-galactose complex. However, we found a global increase in the extent of protein dynamics upon D-galactose binding, i.e. order parameters for apo-ABP were generally larger than for the ABP-galactose complex (Fig.1).

It is likely that the enhanced protein dynamics in ligand-bound state will reduce the entropic cost of D-galactose binding to ABP.

In order to estimate ligand's contribution to the conformational entropy loss upon binding, upper limits of configurational entropies were calculated using the Schlitter method. This method of entropy estimation is based on a quantum-mechanical quasiharmonic approximation: it uses the covariance matrix of the atom-positional fluctuations, which can be readily calculated from the MD trajectory.

Again, we found that entropy loss upon binding of D-galactose to ABP is much lower than expected. It results from relative flexibility of ligand inside the binding pocket – MD simulations suggest that ligand is allowed to rotate about 90 degrees while bound (Figure 2). Contrary to the picture derived from x-ray studies, the ligand is not rigid inside the binding pocket.

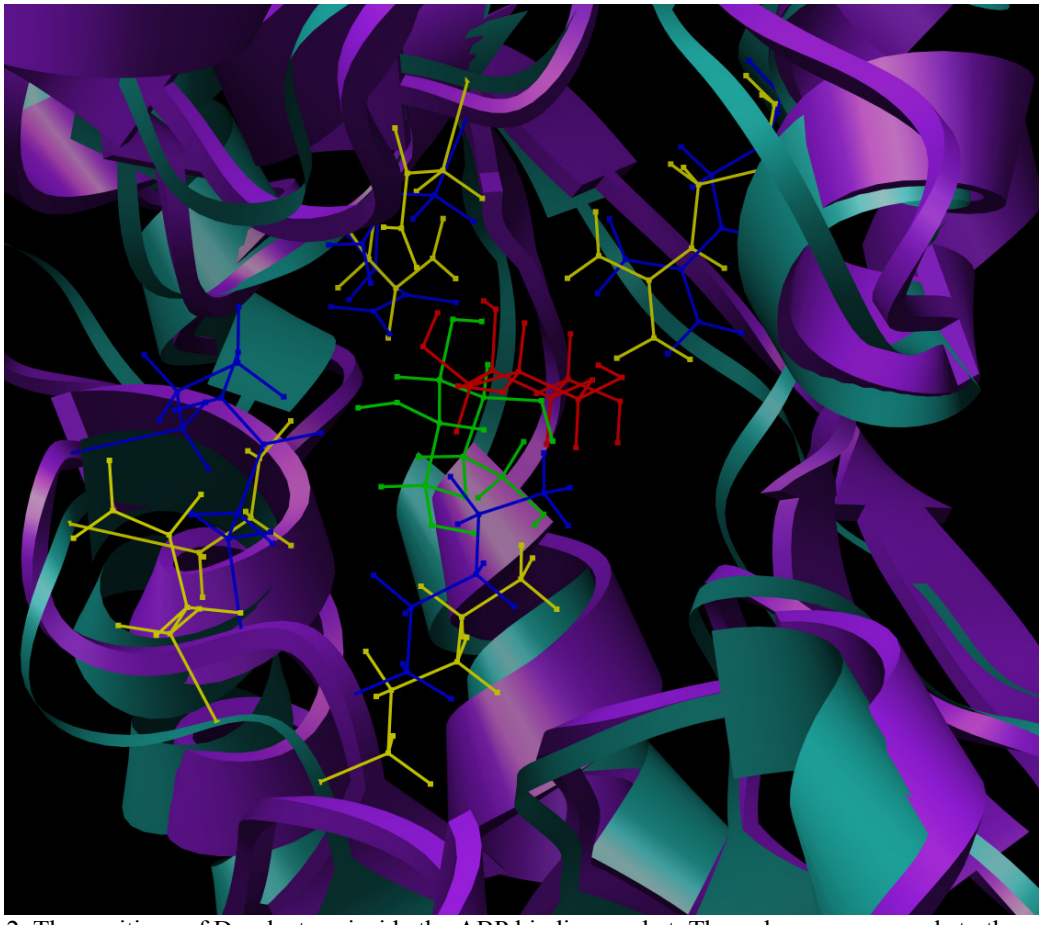


Figure 2. The positions of D-galactose inside the ABP binding pocket. The red one corresponds to the x-ray structure, while green one occurs upon multi-nanosecond MD simulation.

Publications

MacRaild, C.A., Daranas, A.H., Bronowska, A. & Homans, S.W. (2007) Changes in protein dynamics reduce the entropic cost of carbohydrate binding in the Arabinose Binding Protein, *J Mol Biol* in press

Funding
This work was funded by The Wellcome Trust.

Contribution of ligand desolvation to binding thermodynamics of ligand-protein interactions

Natalia Shimokhina and Steve Homans

Introduction

Despite enormous advances in the structure determination of protein complexes, our ability to predict binding affinity from structure remains severely limited. Affinities are governed by both structure and dynamics, including solvent rearrangement. While a number of studies have examined the contribution of water molecules in the protein binding pocket, ligand solvation has received little attention. Here, we examine the latter in the major urinary protein (MUP), an abundant pheromone-binding protein where subtle recognition of a series of related compounds is essential to its biological function. Our approach involves the experimental determination of solvation thermodynamics of relevant ligands using air/water partition coefficient measurements supported by free energy perturbation calculations. These are interpreted in the context of thermodynamic binding data, leading to a decomposition of interaction thermodynamics at a level of detail unreported for any system.

Ligand desolvation

We determined experimental standard free energies of solvation for two pyrazine-derived ligands of MUP, namely 2-methoxy-3-isopropylpyrazine (IPMP) and methyl-pyrazine (MP). In each case we found that the temperature dependence of the standard free energy of solvation is approximately linear, and, therefore, standard enthalpies and entropies could be derived with good accuracy. Further, we used these contributions of ligand desolvation along with earlier measured intrinsic terms (changes in ligand and protein structures, degrees of freedom, etc.) to complete decomposition of binding thermodynamic in this system. With this data it became possible to calculate the entropy and enthalpy contributions from protein desolvation upon binding (0.4 and -12.3 kJ/mol correspondingly) which confirmed the earlier assumption about a dearth of well-ordered water molecules in MUP binding pocket. It became clear that the characteristic entropy driven thermodynamic signature of 'hydrophobic binding' is not realized for MUP, since the favourable entropic contribution arising from ligand desolvation itself is insufficient to overcome the unfavourable contribution from 'freezing' protein and ligand degrees of freedom on binding. The enthalpic contribution to binding is more difficult to deconvolute with accuracy, but the obtained data indicate that the unfavourable enthalpic contribution arising from desolvation of the ligand is offset both by favourable solute-solute dispersive interactions, with a contribution from desolvation of the protein binding pocket.

Collaborators

Charles A. Laughton at the University of Nottingham

Publications

Shimokhina, N., Bronowska, A. & Homans, S.W. (2006) Contribution of ligand desolvation to binding thermodynamics in a ligand-protein interaction. *Angew. Chem. Int. Ed.* **118**, 6522-6524

Funding

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

Contributions to the change in heat capacity occurring on ligand-binding to the major urinary protein

Neil Syme and Steve Homans

Introduction

A better understanding of the factors that determine the affinity for a ligand-protein interaction remains an important goal in biophysical and medicinal chemistry. Three-dimensional structures can reveal a significant amount of information about a complex, but this represents only part of the picture. Affinities are governed not only by the energetics of an interaction, concerning the spatial disposition of interacting groups, but also by the dynamics of these groups. Proteins undergo motions on a wide range of time-scales in order to perform their functions. Thus, to more accurately predict the affinity of a protein for a ligand, it is essential to understand both the enthalpic and entropic contributions to binding. However, these components are enormously complex, and involve contributions from new ligand-protein interactions, solvent rearrangement and changes in protein and ligand degrees of freedom. We employ isothermal titration calorimetry (ITC) and NMR relaxation measurements to study the mouse major urinary protein (MUP) as a model system for investigating the thermodynamics of ligand-protein interactions.

Table 1. Changes in heat capacity recorded for the binding of IBMP (2-methoxy-3-isobutyl pyrazine) and 1-hexanol to the major urinary protein (MUP).

Ligand	ΔC_p (kJ mol^{-1} K^{-1})
IBMP	-0.51 ± 0.03
Hexanol	-0.72 ± 0.18

The binding site of MUP is hydrophobic, a factor essential for its biological function which is to transport hydrophobic molecules in urine. Based on the classical hydrophobic effect we would expect the interactions between MUP and the various ligands it binds to display an entropy-driven thermodynamic signature. Such a signature is generally thought to result from the release of solvating water molecules to bulk solution. Most of the ligands studied, however, demonstrate an enthalpy driven signature. Recent work from our group revealed that the binding site of MUP is poorly solvated, suggesting that dispersive interactions make a significant enthalpic contribution to ligand-MUP interactions. While this explains the absence of an entropy-driven thermodynamic signature, there remains another indicator of the hydrophobic effect. A negative change in heat capacity (ΔC_p) on ligand-binding is associated with the removal of solvating water molecules. The ΔC_p has been determined for the interaction of several ligands with MUP and is negative in each case (Table 1). But, if the MUP binding site contains very few water molecules and does not display the hallmarks of the classical hydrophobic effect, what is the major contribution to ΔC_p?

Firstly, while the protein is believed to be poorly solvated, the ligand will be fully solvated prior to binding. Therefore, we must ask what contribution ligand desolvation makes to the negative ΔC_p? The fortunate discovery that MUP binds the primary aliphatic alcohols pentanol to nonanol has provided a means of answering this question. The solvation thermodynamics of the primary aliphatic alcohols are well characterised in the literature with solvation enthalpies available at several temperatures. Crystal structures have revealed that there are few water molecules present in the complex between MUP and the primary alcohols so we can therefore view the association of the alcohols with MUP as a dehydration event. By taking the enthalpies of solvation for the alcohols hexanol to nonanol from the literature, we can obtain ΔC_p values for desolvation. Using isothermal titration calorimetry (ITC) we can determine experimentally the ΔC_p of binding for each of these alcohols by recording the enthalpy of binding (ΔH_b) at several temperatures. The dehydration ΔC_p values can then be compared with the binding ΔC_p values in order to determine the contribution from ligand

desolvation.

Another possibility is that a change in protein dynamics could affect the ΔC_p. As well as calculating ΔC_p from the relationship between ΔH_b and temperature, it can be calculated from the relationship between binding entropy (ΔS_b) and temperature. In order to study the contribution of protein dynamics to ligand-binding, we employ NMR relaxation-time measurements. By acquiring longitudinal and transverse relaxation times along with the heteronuclear NOE (nuclear Overhauser effect) we can obtain order parameters (S^2) that describe the restriction of internal motion for backbone amide and side-chain methyl sites in ^{15}N, ^{13}C and ^{2}H-labelled samples. If we acquire S^2 values for two states (e.g. apo and holo) it is possible to use these parameters to calculate the difference in backbone and side-chain conformational entropy between the two states. By calculating ΔS_b at several temperatures, we can determine whether changes in protein dynamics have any effect on ΔC_p.

We have completed ITC measurements on the interaction between MUP and the alcohols hexanol to nonanol and ^{15}N-relaxation measurements on the interaction between MUP and hexanol. Our initial results suggest that the principal contribution to the ΔC_p of binding comes from desolvation of the ligand. Analysis of the backbone dynamics implies there is little if any contribution arising from protein backbone motions. Experiments to determine whether side-chain methyl group motions make a more significant contribution are in preparation.

Funding
We acknowledge the support of the Wellcome Trust.

Docking and structure based drug design

Monika Rella, Alasdair T. Laurie, Peter R. Oledzki and Richard M. Jackson

Introduction

Several projects within the group continue to develop novel strategies for structure-based drug design, in addition to applying them to new targets such as Angiotensin Converting Enzyme 2. ACE2 is a new member of the M2 metalloprotease family and a unique human homologue of the Angiotensin Converting Enzyme (ACE), which may be a drug target for the treatment of hypertension and heart disease.

Binding Site Prediction

Q-SiteFinder locates ligand binding sites by clustering favourable regions for van der Waals (CH_3) probes on the protein surface. It uses the GRID forcefield parameters to estimate the interaction energies of probes placed at all points on a three dimensional grid that encompasses the entire protein. Probes with favourable interaction energies are retained and are clustered according to their spatial proximity. The clusters are ranked according to their total interaction energy.

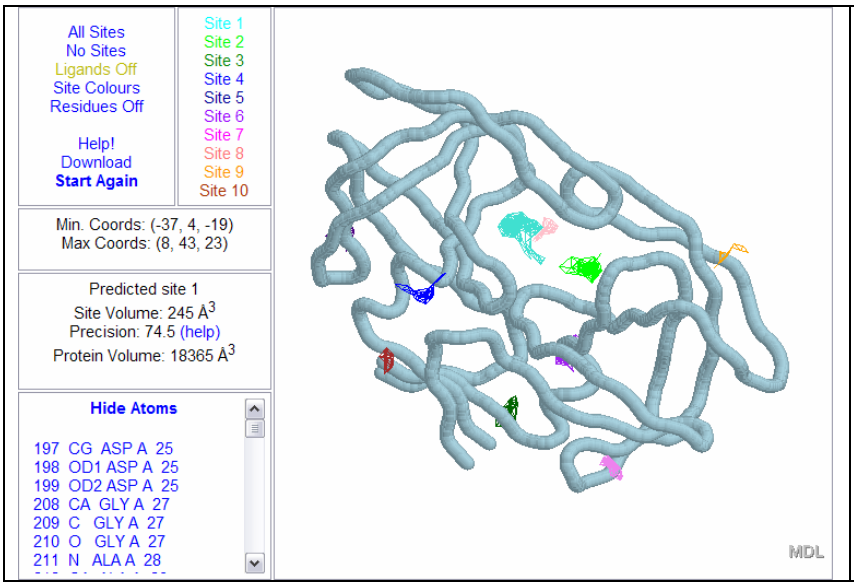


Fig. 1 Ligand binding site prediction for HIV protease (1aaq). The protein is shown in grey tubing with ten colour-coded predicted binding sites, each represented by a cluster of methyl probes. The location of the ligand binding site is identified by the 1st (turquoise), 2nd (light green) and 8th (pink) predicted binding sites.

The algorithm was shown to have a 90% success rate in the top three predicted sites when tested on 134 protein-ligand complexes. The success rate showed a small decrease (to 86%) when tested on proteins in the unbound state, possibly because of the effect of induced fit. The high precision and success rate of Q-SiteFinder will be of benefit in SBDD studies and functional site analysis. Q-SiteFinder and has been made available online (www.bioinformatics.leeds.ac.uk/qsitefinder).

Flexible ligand docking

A novel algorithm for flexible docking, FlexLigdock, is being developed in our lab. 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 65%. When the entire ranked solution set is considered, FlexX docks 70% of the dataset <2Å RMSD, whereas FlexLigdock successfully docks 84%. This level of accuracy is achievable in just under a minute of CPU time on average for each ligand in the data set. A different high-throughput parameterisation for docking has been developed. This

has an average speed of 15 seconds per ligand, with an overall success rate of 78% when considering the entire solution set.

Identification of novel ACE2 inhibitors by structure-based pharmacophore modelling and virtual screening

A computational virtual screening approach was applied to identify novel ACE2 inhibitors based on the ACE2 crystal structure in complex with a potent inhibitor. Pharmacophore modelling has been used to identify steric and electronic features responsible for specific drug-receptor interactions in the ACE and ACE2 family of proteins. This study generated a set of protein structure-based ACE2 pharmacophore models to apply as filters for the virtual screening of a large commercial database comprising ~3.8 M compounds. Model selectivity was assessed by hit reduction of an internal ACE inhibitor database and the Derwent World Drug Index.

A subset of 25 compounds was proposed for bioactivity evaluation derived from high geometric fit values and visual inspection as well as diverse structure. Seventeen compounds were purchased and tested in a bioassay at 200 μM. The six most promising candidates were subjected to IC50 studies, all revealing novel ACE2 inhibitors in the range of 62-179 μM. This corresponds to a hit rate of 41%.

Publications

Laurie, A.T. & Jackson, R.M. (2006) Methods for the Prediction of Protein-Ligand Binding Sites for Structure-Based Drug Design and Virtual Screening. *Current Protein and Peptide Science* **7**, 395-406.

Rella, M., Rushworth, C.A., Guy, J.L., Turner, A.J., Langer, T. & Jackson, R.M. (2006) Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. *Journal of Chemical Information and Modelling* **46**, 708-716.

Bartlett S., Beddard G.S., Jackson R.M., Kayser V., Kilner C., Leach A., Nelson A., Oledzki P.R., Parker P., Reid G.D. & Warriner S.L. (2005) Comparison of the ATP binding sites of protein kinases using conformationally diverse bisindolylmaleimides. *J Am Chem Soc.* 33, 11699-708.

Funding

These projects are funded by the BBSRC, University of Leeds and The Wellcome Trust.

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 uses a method based on geometric hashing to compare the structures and properties of all known ligand binding sites and assess the extent of their similarity. Geometric hashing gives a similarity score (number of superimposable atoms of the same type e.g. carbon, nitrogen, oxygen), 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 (Figure 1). A multiple alignment of structurally equivalenced atoms is also provided.

Comparison of each binding site to all other binding sites provides a matrix of similarity scores. Figure 2 shows a matrix for a subset of proteins all known to bind nucleotide-containing ligands. 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, which have different overall folds but share a common structural P-loop.

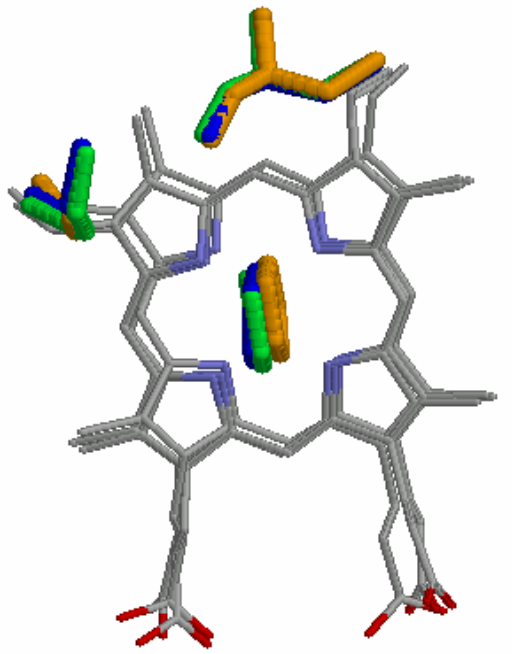


Figure 1. Superposition of the haem binding site of cytochrome c4 (1m6z), flavocytochrome c (1dxr) and cytochrome f (1e2z). These proteins have different folds. 1m6z has a cytochrome c fold, 1dxr has a multihaem cytochrome fold and 1e2z has a common fold of diphtheria toxin/transcription factors/cytochrome f fold.

The project has advanced considerably and in 2006 we finalised an industry contract with PharmaDesign Inc. in Tokyo to licence our software. Work is now underway to cluster binding sites according to their structural similarity with the purpose of identifying binding sites to represent each cluster. Our aim is to allow new and putative binding sites to be compared to the representative set of binding sites using geometric hashing. If sufficient structural similarity can be found between a putative or uncharacterised binding site and a known binding site a prediction of related function or common ligand recognition can be made.

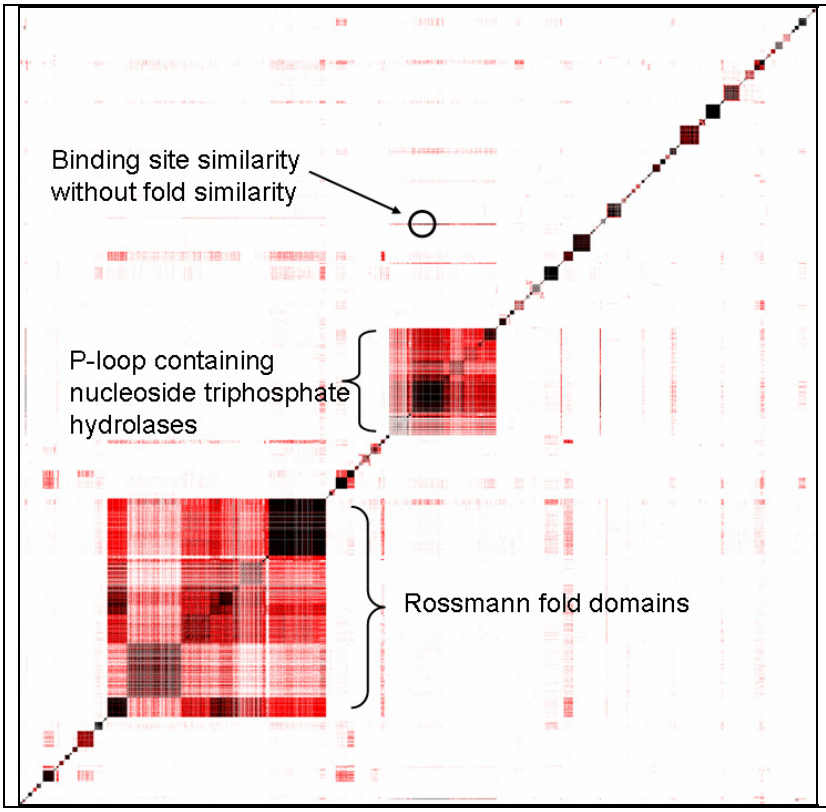


Figure 2. Similarity matrix from the database of nucleotide binding sites. Strong similarity (score>39) is indicated by a black dot and weaker but still significant similarity (score 25-39) is indicated with a red dot. Binding sites have been ordered along the axes by their evolutionary relationships *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.

Publications

Gold, N.D. & Jackson, R.M. (2006) Fold independent structural comparisons of protein-ligand binding sites for exploring functional relationships. *Journal of Molecular Biology* **355** 1112-1124

Gold, N.D. & Jackson, R.M. (2006) A searchable database for comparing protein-ligand binding sites for the analysis of structure-function relationships. *Journal of Chemical Information and Modelling* **46** 736-742

Gold, N.D. & Jackson, R.M. (2006) SitesBase: A database for structure-based protein-ligand binding site comparisons. *Nucleic Acids Research* **34** (Database issue) D231-4.

Funding

We wish to acknowledge the support of the BBSRC.

Predicting protein-protein complexes

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 can know both the structure of a protein and the other molecules it may bind to, the structure of the complex is still difficult to determine. Bioinformatic predictions of the structure adopted by two interacting proteins can be instructive for the molecular-biologist, guiding experimental analysis of the complex and being useful for the design of interaction inhibitors.

Like docking for the purpose of generating protein-ligand structures, protein-protein docking is usually a two step process where one of the proteins is placed on the surface of the second in many different orientations, based on an easily quantifiable scoring function. These complexes are then ranked, clustered and refined using ever more computationally expensive procedures until a manageable number of suitable complexes are obtained.

Recent work in the rescoring of docked protein-protein complexes, and the analysis of protein interfaces (see Figure), has shown that the energy associated with the desolvation of the protein surface plays an important role in the prediction of the correct protein complex. Although useful, the problem arises that desolvation potentials are specific to the protein structures from which they are derived. This is most significant when the contact potentials are derived for their role in protein folding and then applied to protein-protein interactions. Surfaces of different chemical composition interact in the two types of interaction; therefore different measures of desolvation could be expected.

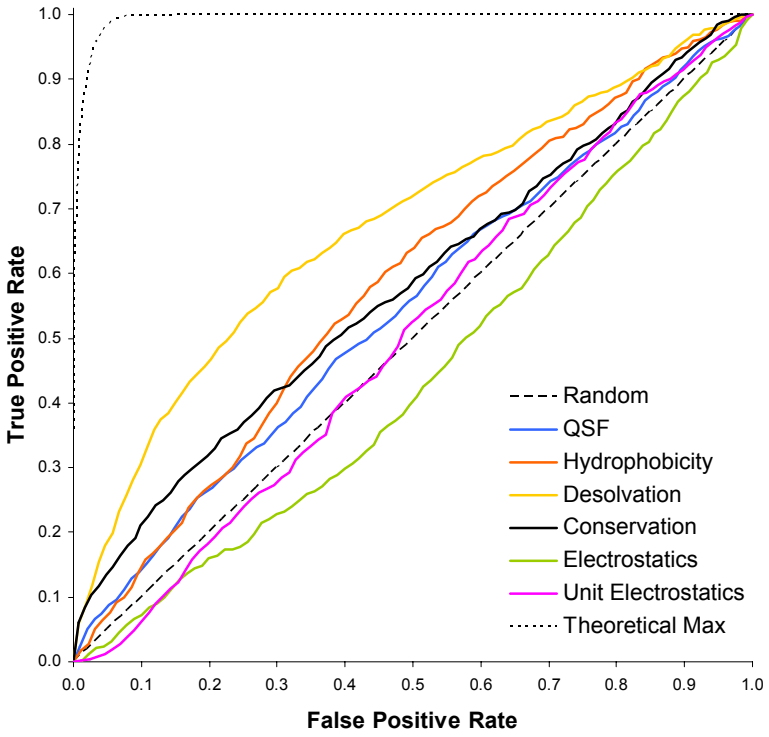


Figure. A Receiver Operating Characteristic (ROC) curve showing the prediction of clefts that lie in a the protein interface relative to those on the rest of the protein surface, for a collection of protein complexes. Desolvation of the clefts is the property that best correlates with the interface, as shown by the largest curvature from the diagonal. Image taken from Burgoyne and Jackson, 2006.

Computationally, desolvation is assessed as either: 1) the weighted loss of surface area when molecules interact; or 2) the weighted number of contacts which when formed break contacts with water molecules. We have generated the first weighted contact potential specifically for the assessment of protein-protein interactions which we have shown to aid the ranking of the native-like complexes in a set of decoy structures. We intend for this to form part of a protein-protein docking strategy.

Funding
This work is funded by the MRC.

Publications
Burgoyne, N.J., Jackson, R.M. (2006) Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces, *Bioinformatics*, **22**, 1335-1342

Electrodes for redox-active membrane proteins

Andreas Erbe, Sophie Weiss, Nikolaos Daskalakis, 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' membranes

We have prepared electrode surfaces which enable the characterisation of redox-active membrane enzymes in a native-like environment. For this, we have used the methodology of tethered bilayer lipid membranes (tBLM), in which the lipid bilayer is attached to the electrode surface via special chemical anchors that are bound to the surface on one side and insert into a bilayer leaflet at the other (Figure 1). For this purpose cholesterol derivatives have been used, which, via a hydrophilic linker, are connected to a thiol group that forms self-assembled monolayers (SAMs) on gold electrodes. These cholesterol-lipids have been mixed with small thiols to provide space for transmembrane proteins. The surfaces of these mixed SAMs have been characterised in detail and shown to exhibit a complex pattern of phase separation.

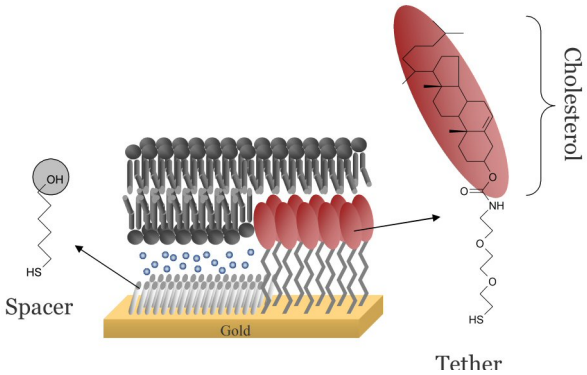


Figure 1: Chemical structures of 6-mercaptohexanol (spacer) and the cholesterol tether molecule used to form tBLMs.

Spectroscopic studies

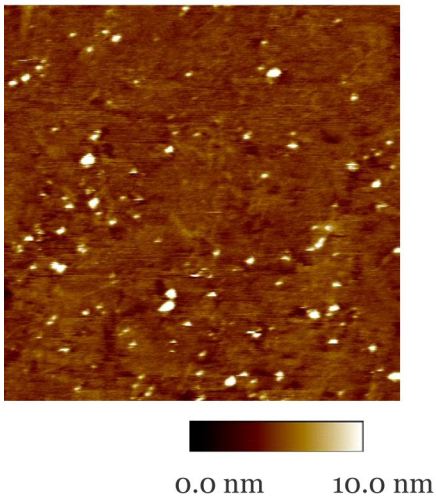
The formation of tBLMs has also been studied in detail with attenuated total reflectance infrared spectroscopy combined with impedance spectroscopy. This novel combination of spectroscopies was made possible by the design of a new experimental setup and the results indicate that the structure of the tBLM varies with varying the mixed SAM composition. On SAMs with a high content of cholesterol tether, tBLMs with reduced fluidity were formed. For SAMs with low content, the results are consistent with the adsorption of flattened vesicles, while spherical vesicles have been found in a small range of surface compositions.

Electrochemistry

Quinones (also known as Coenzyme-Q) are important electron mediators in biology and function by transporting electrons between membrane enzymes. The electron transfer properties of quinones are closely coupled its protonation and biology has utilised this to build up a proton motive force (*pmf*) across the membrane, which is the key feature in the generation of energy in all living organisms. We have used the tBLM system to study the quinone redox properties and have found that the protonation can only occur at the membrane-solution interface as protons are not able to penetrate the membranes.

Importantly, however, it was shown that quinones can also accept protons in the membrane directly from commonly-used proton uncouples like CCCP, which shows an previously unknown action of this class of compounds that are often used to disrupt *pmf* in experimental systems.

The quinone in the tBLM system still acts a substrate for quinone enzymes that have been co-immobilised on the surface. In this way the activity of an ubiquinol oxidase from *Escherichia coli*, cytochrome bo_3 (cbo_3), was studied using voltammetry techniques. Enzyme coverages observed with microscopic (AFM) and voltammetric techniques were shown to coincide (2-8.5 fmol·cm^{-2}) indicating that most - if not all - cbo_3 on the surface is catalytically active and thus retains its integrity during immobilisation.



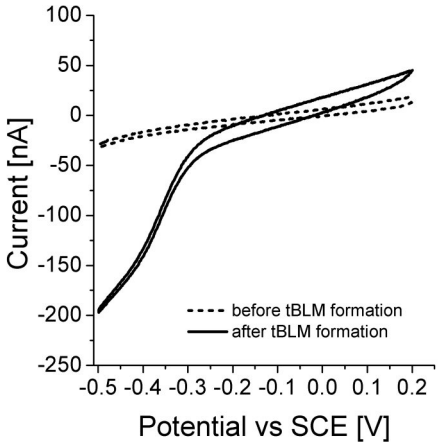


Figure 2: (Left) TM-AFM results of a tBLM with cbo_3.The membrane proteins are observed as heightened (5-10 nm) areas protruding above the planar lipid bilayer. (Right) CV results indicating catalytic oxygen reduction after formation of a tBLM

Future directions

We are currently preparing tBLMs using complete bacterial membranes. In contrast to the systems prepared previously with purified membrane proteins, these systems are easier to prepare and more robust in nature. We are currently using these systems to study cbo_3 in more detail. In a second project, intact vesicles with fluorescent pH indicators are adsorbed on the surface in order to monitor simultaneously the electron and proton transfer properties of the redox enzymes of interest.

Collaborators and Funding

Dr. Robert B. Gennis (University of Illinois, Urbana, USA)

This work was funded by the BBSRC (David Phillips fellowship).

Publications

Jeuken, L..J.C., Connell, S.D., Henderson, P.J.F., Gennis, R.B., Evans, S.D. & Bushby, R.J. (2006). Redox enzymes in tethered membranes *J. Am. Chem. Soc.* **128**, 1711-1716.

Jeuken, L.J.C., Bushby, R.J. & Evans, S.D. (2007) Proton transport into a tethered bilayer lipid membrane, *Electrochem. Comm.*, **9**, 610-614.

Jeuken, L.J.C., Daskalakis, N.N., Han, X., Sheikh, K. Erbe, E., Bushby, R.J. & Evans, S.D. (2007) Phase separation in mixed self-assembled monolayers and its effect on biomimetic membranes, *Sensors and Actuators B: Chemical*, **In press**

Erbe, A., Bushby, R.J., Evans, S.D., Jeuken, L.J.C. (2007) Tethered bilayer lipid membranes studied by simultaneous attenuated total reflectance infrared spectroscopy and electrochemical impedance spectroscopy, **Submitted.**

Structure and properties of regulated myosin 5

Kavitha Thirumurugan and Peter Knight

Introduction

Myosin 5 is involved in a multitude of functions, including hearing, muscle contraction, neural transmission, membrane trafficking and vesicular transport. In mice, a brain-specific isoform moves synaptic vesicles and the melanocyte-specific isoform carries melanosomes in the skin. Myosin 5 is a cargo-carrying motor protein that moves along actin filament without letting go. It is regulated: when it has no cargo it is in an inhibited state that interacts only rarely with actin. Whereas there is a vast amount of research on functions performed by the active state of myosin 5, the structure and properties of the inactive molecule are not well known. In 2004, we showed by electron microscopy that the inactive molecule has a folded conformation. Our initial negative stain images of the inhibited state showed a compact conformation, but identification of substructure was ambiguous. Our recent detailed study of the structure and properties of the inhibited state have improved our understanding of the mechanism of myosin regulation in cargo transport.

Folded myosin 5 structure

Following on from our initial work, we have studied in detail the structure of the folded myosin 5, and the role of the C-terminal cargo-binding globular tail domain (GTD) in regulation. We find the folded, inactive conformation is achieved by the interaction of the GTD with the motor domain. This tail domain interaction inhibits the motor domain function. Single-particle image processing of folded myosin 5 molecules negative-stained in ATP at nanomolar calcium reveals detailed substructure (Fig. 1a-f).

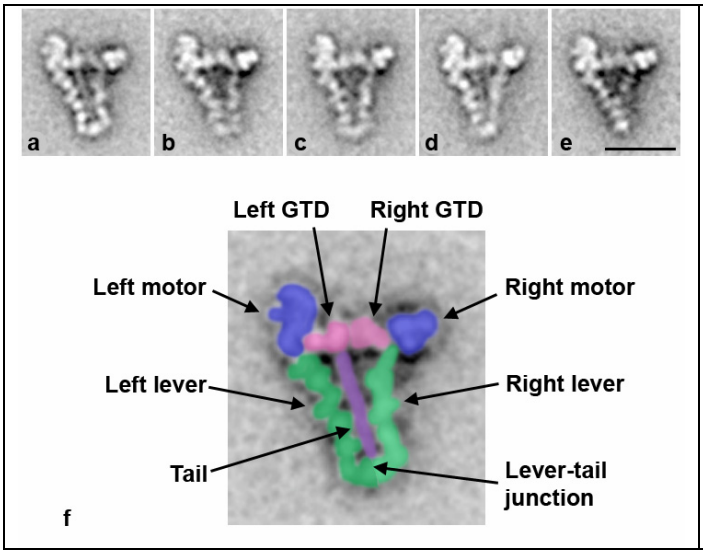


Fig. 1 Structure of switched-off myosin 5. **a-e,** averaged images of negative stained, folded whole myosin 5 molecules. Scale bar 20 nm. **f,** enlargement of **a,** coloured and labelled to show domains within folded myosin 5.

The two heads lie either side of a proximal 17nm segment of the tail, with the head on the left of the tail more clearly delineated and closer to the tail. Within each head, we can see the calmodulin subunits of the lever domain and overlap of the flexible levers at their junction with the tail. The GTDs appear as two globules located between the two motor domains. We can identify well the left GTD, which is a small lobe (~7 nm) touching the left motor domain closer to the lever (Fig. 1f). This GTD has a consistent location on the motor domain indicating its site specificity. Comparison with the atomic model shows that this putative GTD-binding site is restricted to a specific segment Pro117-Pro137 of the motor domain (pink in Fig. 2a). The right GTD (~7 nm) is less prominent, but extends into the junction between motor domain and lever, which is the location of Pro117-Pro137 in right head models. Within the motor domain, Pro117-Pro137 is near the entrance to the ATP-binding pocket, loop 1 and the β-bulge of the transducer, but is far from the actin-binding interface. Thus each GTD binds at an equivalent site and acts allosterically, not by physically blocking

binding to actin. Alignment of the sequences of the GTD-binding site from many sources of myosin 5 shows a conserved GDMDP sequence which is probably the site of specificity for myosin 5 regulation. Four outward-facing acidic residues in this segment of the motor domain (Fig. 2b) may make ionic interactions with basic residues of the GTD.

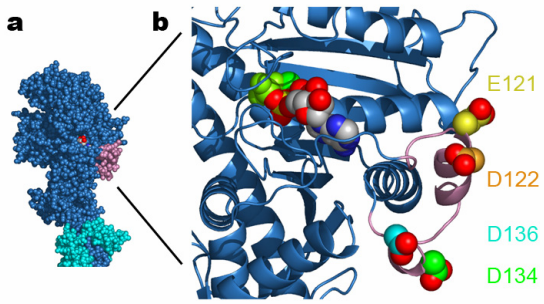


Fig. 2 GTD-binding region of myosin 5 motor domain.
a, motor domain region of the myosin 5.ADP.BeF_x complex oriented to match the appearance of the left motor of Fig. 1a. The putative GTD-binding region is pink. Actin would bind at the top left.
b, enlargement of the GTD-binding region of **a**. The polypeptide chain is shown in cartoon form, except the outward-facing four acidic residues (E121, D122, D134, D136) shown in spacefill. The bound MgADP.BeF_x is shown as a spacefilling model, with ADP shown with carbon silver-grey, oxygen red, nitrogen dark blue and phosphorus yellow, BeF_x and Mg^{2+} green.

The role of the GTD in regulation was studied using a GST-GTD fusion protein. This dimeric construct $(GST\text{-}GTD)_2$ inhibits the actin-activated MgATPase of myosin 5 lacking the GTD domain with an inhibitory dissociation constant K_i of 0.53 μM. This construct alone is sufficient to cause the inhibition. Electron microscopy of the inhibited complex shows a compact, triangular shape similar in conformation to that of the full-length folded myosin 5.

Our additional EM and TIRF microscopy results indicate that the folded molecules have a low affinity for actin in the presence of ATP so may diffuse freely in cells. This low affinity allows the inactive molecule to let go of the actin thereby saving the energy that otherwise be needlessly wasted to keep it moving on the track. Free diffusion allows the folded molecules to be recycled, ready to pick up a new cargo and move along the track when required.

Collaborators
James R Sellers, John A Hammer III and Takeshi Sakamoto (NIH, MD, USA).

Publications
Thirumurugan, K., Sakamoto, T., Hammer, III, J.A., Sellers, J.R. & Knight, P.J. (2006) The cargo-binding domain regulates structure and activity of myosin 5. *Nature.* **442,** 212-215.

Funding
This work was funded by the Wellcome Trust.

Two metals are better than one: *E. coli* amine oxidase

Pascale Pirrat, Mark Smith, Peter Knowles, Simon Phillips and Mike McPherson

Copper amine oxidases (CuAOs) are a ubiquitous class of enzymes important in the cellular and extracellular metabolism of amine substrates. *Escherichia coli* CuAO (ECAO) is a 160 kDa homodimer protein containing one copper ion and a post-translationally modified tyrosine cofactor; 2,4,5-trihydroxyphenylalanine quinone (TPQ) per subunit. TPQ biogenesis is an autocatalytic event requiring copper and oxygen. Studies have identified that amine oxidation occurs via a 'ping-pong' mechanism with two halves involving reductive and oxidative steps. The oxidation of reduced TPQ requires the reduction of O_2. However, the role of the TPQ and the copper in the oxidative half cycle remains elusive. Insights into the mechanism of O_2 activation and the role of both active site and auxiliary metals are likely to be important for mechanistic understanding of many metalloenzymes.

In ECAO, two calcium ions (Ca) have been identified per monomer. The buried calcium and its liganding residues lie at the ends of two anti-parallel β-strands the other ends of which carry the three histidines that coordinate the active site copper ion (Figure 1). Mutagenesis studies at the two calcium binding sites have demonstrated that the buried site is likely to play a role in protein folding, whereas the peripheral site, which lies some 30Å from the active site, is linked to activity and more surprisingly to TPQ processing. Crystal structure analysis has been carried out on two "peripheral site" mutants to determine the conformational changes giving rise to these observations. We are also investigating the role of copper through a series of active site metal replacement studies. Further structural, spectroscopic and biochemical analyses are under way.

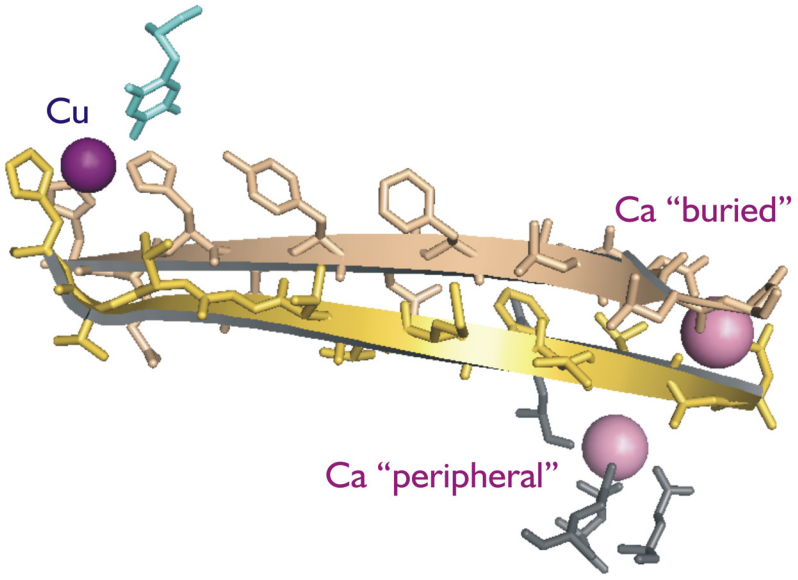


Figure 1: Structural relationship between the buried Ca and the active site Cu (blue). TPQ, in pink, is shown in its "on-copper" conformation.

Acknowledgements
This work is supported by funding from BBSRC and the Wellcome Trust PhD programme.

Second co-ordination sphere residues in galactose oxidase: tryptophan 290

Melanie Rogers, Nana Akumanyi, Sarah Deacon, Peter Knowles,
Simon Phillips and Michael McPherson

Introduction

Galactose oxidase (GO) is a copper-containing 68kDa enzyme, secreted by species of *Fusaria*, and post-translationally modified to yield the mature peptide of 639 amino acids. GO catalyses the stereo- and regio-specific oxidation of a broad-range of primary alcohols, from small aliphatics to large polysaccharides generating their corresponding aldehydes with associated reduction of oxygen to hydrogen peroxide. The catalytic mechanism involves the formation of a post-translational covalent cross-link between the active site residues Tyr272 and Cys228. Tyr272 is the site for a protein radical and the cross-linked feature involves a stacking interaction with Trp290.

Second co-ordination sphere

It has become increasingly obvious that the environment of protein based radical can have a major effect on modulating and tuning its properties and reactivity. We are systematically investigating the influence of the active site microenvironment on the structure and reactivity of the tyrosyl radical in galactose oxidase. To define the roles of W290 we compared the properties of three mutational variants, W290F, W290H and W290G. Crystal structures for these proteins were determined and the reactivity and electronic structures investigated through kinetic and UV-vis, CD and EPR spectroscopic approaches.

We found that W290 is involved in recognition and binding of some substrates through hydrogen bonding interaction. It also plays an important role in lowering the redox potential of the tyrosyl radical and influences the stability of this radical. Interestingly it may also modulate the reactivity of the active site base Tyr495. It is clear that perturbations of weak interactions modulated by W290 have substantial effects on reactivity without major change to the electronic structure and bonding of the Cu(II) and its liganding residues.

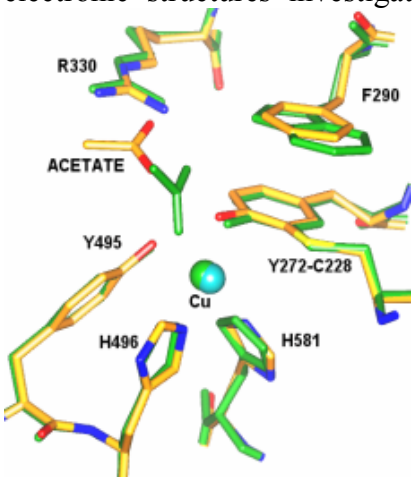


Fig 1. Active site of W290F (atom colours) superimposed with wild-type

Publications

Rogers, M.S., Tyler, E.M., Akyumani,N., Kurtis, C.R.P., Spooner, R.K., Deacon, S.E., Tamber, S., Firbank, S.J., Mahmoud, K., Knowles, P.F., Phillips, S.E.V., McPherson, M.J. & Dooley, D.M. (2007) The stacking tryptophan of galactose oxidase: a second coordination sphere residue that has profound effects on tyrosyl radical behaviour and enzyme catalysis. *Biochemistry*. In Press.

Collaborators

Dave Dooley, Department of Chemistry and Biochemistry, Montana State University, Bozeman, USA.

Funding

This work was sponsored by the BBSRC.

Recombinant production of peptidic bionanomaterials

Kier James, Jessica Riley, Stephen Parsons, Stuart Kyle and Mike McPherson

Replacement materials for those derived from non-renewable petroleum feedstocks are required. Biological production of new self-assembling materials which can be functionalised to increase their versatility and suitability for manufacturing applications offers one solution. The most promising example of this approach has been the use of short peptides capable of self-assembly. These peptides undergo self association driven by the formation of non-covalent interactions such as electrostatic, hydrophobic, van der Waals, and metal ligand interactions. Such materials offer numerous advantages such as biodegradability and the potential to tailor make 'smart' materials with specific surface, mechanical and responsive properties.

Currently, self-assembling peptides are predominantly produced using solid or liquid phase chemistry. However, these methods are cannot be used cost effectively to produce sufficient quantities of peptides for many manufacturing applications. One aim of our research is the high yield recombinant production of self-assembling peptides. These peptides have been rationally designed by the Centre for Self Organising Molecular Systems (SOMS) to form β-sheets that hierarchically self-assemble into helical tapes, twisted ribbons (double tapes), fibrils (twisted stacks of ribbons) and fibres (entwined fibrils) with increasing concentration driven by environmental changes, such as pH. The recombinant peptides are characterised by a range of biophysical approaches including TEM, AFM, (Fig1) CD and FT-IR.

We have used *Escherichia coli* expression hosts where yields of 600mg/L of recombinant self-assembling peptide have been obtained. In addition, we are exploring alternative hosts including *Pichia pastoris, Aspergillus niger* and *Nicotiana tabacum* in order to increase yields. A range of constructs containing various numbers of peptide repeats, purification tags, fusion proteins and cleavage sequences are being tested with the aim to maximise yields in each of the host organisms. In addition, toxicological studies will be undertaken in order to assess the viability of recombinantly-produced, self-assembling peptides for *in vivo* applications such as tissue engineering.

We are also using our expression systems to explore the production of medically important peptides including those conferring bacteriocidal and anti-proliferative functions.

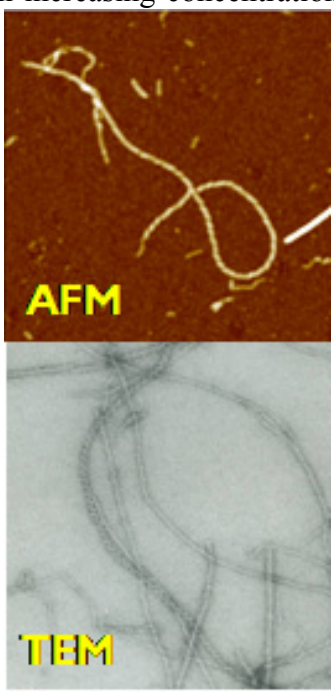


Fig 1: Images of recombinant self-assembled peptide fibrils

Collaborators

Amalia Aggeli, Centre of Self Organising Molecular Systems (SOMS); Eileen Ingham, IMCB, University of Leeds; Rudi Koopmans, Dow Chemicals.

Funding

We would like to acknowledge studentship funding from BBSRC, The Wellcome Trust and The Dow Chemical Company and the University of Leeds Institute of Bionanoscience.

Applications of synthetic organic chemistry to biological problems

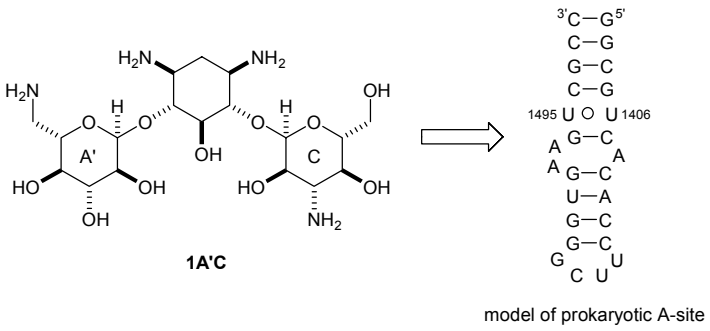
Chris J. Adams, Mark Anstiss, Alison E. Ashcroft, Andrew J. Baron, Blandine Clique, Claire Crawford, Teresa Damiano, Robert Hodgson, Alan Ironmonger, Adam Nelson, Alexis Perry, Peter G. Stockley, James Titchmarsh and Ben Whittaker

Introduction

Synthetic organic chemistry is an immensely powerful tool for Chemical Biology, which we exploit in a wide range of biological problems: from the directed evolution of enzymes for use in synthetic chemistry (using biology to control synthetic chemistry), to chemical genetic studies (using chemistry to control biology). A summary of some work accepted for publication in 2006 is provided. You might like to browse our group webpages at www.asn.leeds.ac.uk to find out more about what we do!

Applications of aminoglycoside libraries in the discovery of ligands for RNA hairpins

We have previously described the preparation of a library of around forty aminoglycosides with diverse stereo- and regiochemistry. In order to double the stereochemical diversity of the library, the compounds were screened against both enantiomers of three RNA hairpins. The affinity of three active compounds (including **1A'C**) and an isomeric compound was determined using surface plasmon resonance. Remarkably, the three tight-binding aminoglycoside derivatives were found to discriminate rather poorly between the alternative RNA sequences. Within the aminoglycoside derivative library studied, high affinity for an RNA target was not accompanied by good discrimination between alternative RNA sequences.



Applications of conformationally constrained bisindolylmalemides in chemical genetics studies of signalling pathways

The conformational constraint of bisindolylmaleimides can allow the selective inhibition of protein kinases. The length of a tether linking the indolyl nitrogen atoms determines the position of equilibrium between the limiting *syn* and *anti* conformations. Some bisindolylmaleimides with longer tethers (e.g. fourteen atoms) have a *syn* conformation as their ground state, and are selective inhibitors of the protein kinase GSK-3. In collaboration with Professor Melanie Welham (University of Bath), we have exploited such compounds as tools in chemical genetic studies of signalling pathways in murine embryonic stem cells.

We also published papers on new synthetic methods, and the development of efficient strategies and approaches for natural product synthesis. Results from our collaboration with Alan Berry on the directed evolution of enzymes for use in synthetic organic chemistry are described elsewhere in this report.

Acknowledgements

ASN holds an EPSRC Advanced Research Fellowship (2004-2009). We thank EPSRC, BBSRC, the Wellcome Trust, GSK, AstraZeneca, Hoffman La-Roche, Avecia and Organon for funding. We would like to thank the following collaborators for their contribution to the

work described here: Professor Melanie Welham (University of Bath), Ian Patel (AstraZeneca) and Andrew Kennedy (GSK).

Publications

B. Clique, A. Ironmonger, B. Whittaker, J. Colley, J. Titchmarsh, P. Stockley & A. Nelson, (2005) Synthesis of a library of stereo- and regiochemically diverse aminoglycoside derivatives, *Org. Biomol. Chem.,* **3**, 2776-2785.

S. Bartlett, G. S. Beddard, R. M. Jackson, V. Kayser, C. Kilner, A. Leach, A. Nelson, P. R. Oledzki, P. Parker, G. D. Reid & S. L. Warriner, (2005) Comparison of the ATP binding sites of Protein Kinases using conformationally diverse bisindolylmaleimides", *J. Am. Chem. Soc.* **127**, 11699-11708.

C. Crawford, A. Nelson and I. Patel, (2006) Development of an approach to the synthesis of the ABC Ring System of Hemibrevetoxin B, *Org. Lett.* **8**, 4231-4234.

M. Anstiss and A. Nelson, (2006) Catalytic and stoichiometric approaches to the desymmetrisation of centrosymmetric piperazines by enantioselective acylation: A total synthesis of Dragmacidin ", *Org. Biomol. Chem.* **4**, 4135-4143.

G. J. Williams, T. Woodhall, L. Farnsworth, A. Nelson and A. Berry, (2006) Creation of a Pair of Stereochemically Complementary Biocatalysts", *J. Am. Chem. Soc.* **128**, 16238-16247.

A. Ironmonger, B. Whittaker, A. J. Baron, B. Clique, C. J. Adams, A. E. Ashcroft, P. G. Stockley & A. Nelson, (2007) Scanning conformational space with a library of stereo- and regiochemically diverse aminoglycoside derivatives: The discovery of new ligands for RNA hairpin sequences", *Org. Biomol. Chem.* **5**, in press.

M. P. Storm, H. K. Bone, C. G. Beck, P.-Y. Bourillot, V. Schreiber, T. Damiano, A. Nelson, P. Savatier & M. J. Welham, (2007) Regulation of Nanog expression by phosphoinositide 3-kinase-dependent signalling in murine embryonic stem cells, *J. Biol. Chem.* in press.

R. Hodgson, A. Kennedy, A. Nelson and A. Perry, (2007) Synthesis of 3-sulfonyloxypyridines: Oxidative ring expansion of α-furylsulfonamides and *N→O* sulfonyl transfer, *Synlett* in press.

Computational self-association of amyloid fibril forming peptides

Oliver Clarke, Malcolm McLean and Martin Parker

Introduction

Amyloid, a highly ordered protein aggregate state with a characteristic cross-β structure, is associated with several diseases, including Alzheimer's disease and the prion diseases. Recently, there have been several studies suggesting that toxicity is not caused by amyloid fibrils themselves, but by soluble oligomers formed *en route* to the mature fibre. Amyloid fibres and oligomers are difficult to study experimentally and there is little high resolution structural data available. Computational methods can offer valuable insights into the structures of these species and the forces holding them together. This information is vital for the rational design of therapeutic agents which prevent amyloid formation. Owing to the long timescales involved, classical molecular dynamics simulations are not possible. Alternative, less CPU-intensive methods are needed; therefore, using a reduced but effective protein model and force field, in tandem with a novel Monte Carlo sampling algorithm, we have successfully simulated the self-association of amyloid forming peptides.

A fast and effective force field

Reduced Ramachandran angle and force field for tertiary structure prediction (RAFT) is an accurate and fast simplified protein model. Proteins are represented using six pairs of backbone torsion angles, vastly reducing conformational space. Each residue is represented by a sphere centred upon its backbone atoms and side chains are represented by between zero and three spheres fixed along the C_α-C_β vector, with the total volume approximating that which the real side chain occupies in water. The force field comprises hydrophobic forces, excluded volume effects, hydrogen bonding and a simple electrostatic potential. RAFT has been shown to be effective on small peptide systems and, unlike most simplified models, is equally good at predicting β secondary structures as α secondary structures.

A highly efficient sampling method

Peptides are very dynamic. Thus, it is important to calculate their time-averaged structures, i.e. one must calculate the Boltzmann distribution. This requires a sampling method. Standard methods such as Metropolis Monte Carlo are inefficient, as at physiological temperatures they become trapped in local minima. We have developed a method called Density guided importance sampling (DGIS) that guides simulations towards under-sampled regions of the energy spectrum and recognises when equilibrium has been reached, thus preventing arbitrary and excessively long execution times. It has been shown to perform better than other hybrid Monte Carlo methods such as Replica Exchange.

Self-association of amyloid forming peptides

12 non-homologous fragments of amyloidogenic proteins that have been shown experimentally to form amyloid fibrils were sampled using the RAFT model with DGIS. Three copies of each peptide were placed in a periodic box of dimensions equivalent to three fully extended copies of the peptide being simulated. Starting from a random conformation, the simulations were run until equilibrium was reached. Simulations required on average six days to complete. In every case except one, the most probable structure is a β-sheet. Each peptide was also sampled as a monomer. The most probable structures are not β-strands. 4 representative examples are shown in Figure 1. Each is associated with an amyloid disease: Alzheimer's disease, dialysis-related amyloidosis, Diabetes type II and the prion diseases. That the most likely monomer structures are not β-like, whilst the three peptide systems form β-sheets, suggests aggregation is a co-operative phenomenon. These results also provide further proof of RAFT's general predictive capacity.

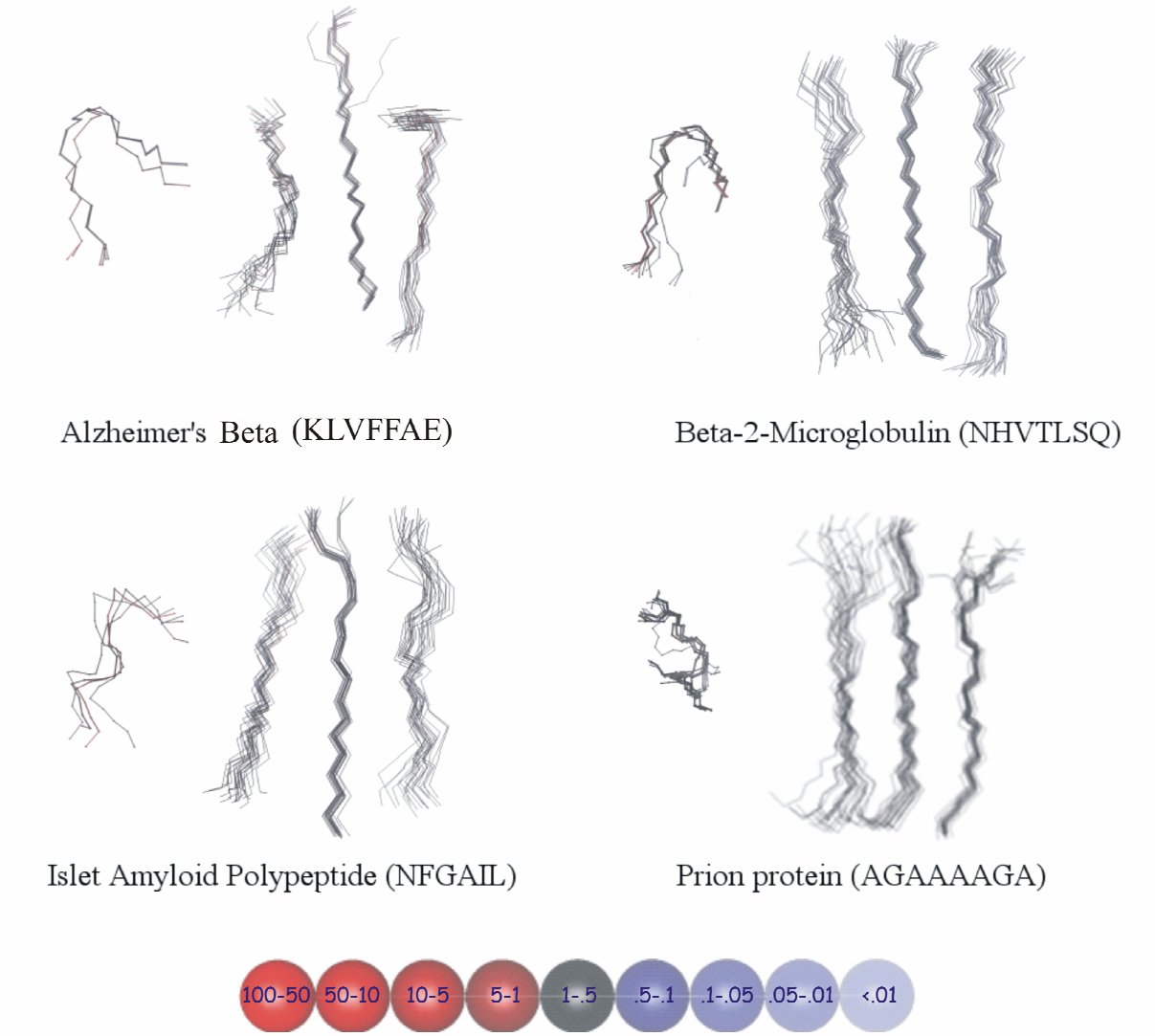


Figure 1. Sampled structures are superimposed on top of one another. Each individual structure is assigned a probability, and is coloured according to the key. For each image: (LHS) ensemble structure of autonomous amyloid protein fragment. (RHS) ensemble structure of three amyloid protein fragments placed in a periodic box. Labels indicate peptide name and sequence.

Sequence effects on propensity to form amyloid

Several mutants of the amyloidogenic peptide STVIIE were synthesised and structurally characterised by Serrano and colleagues. Some of the mutants aggregated, but did not form amyloid fibrils. We simulated STVIIE and several mutants using RAFT and DGIS, as above. An interesting example is shown in Figure 2. The ensemble structure of STVIIE forms an ordered β-sheet. As STVIIE forms fibrils experimentally, this is expected. The ensemble structure of STLEFE shows aggregation, but is significantly less ordered than STVIIE. This correlates with the experimental data shown in the electron micrograph, where XXLEFX forms small aggregates, but not the ordered fibril seen for STVIIE.

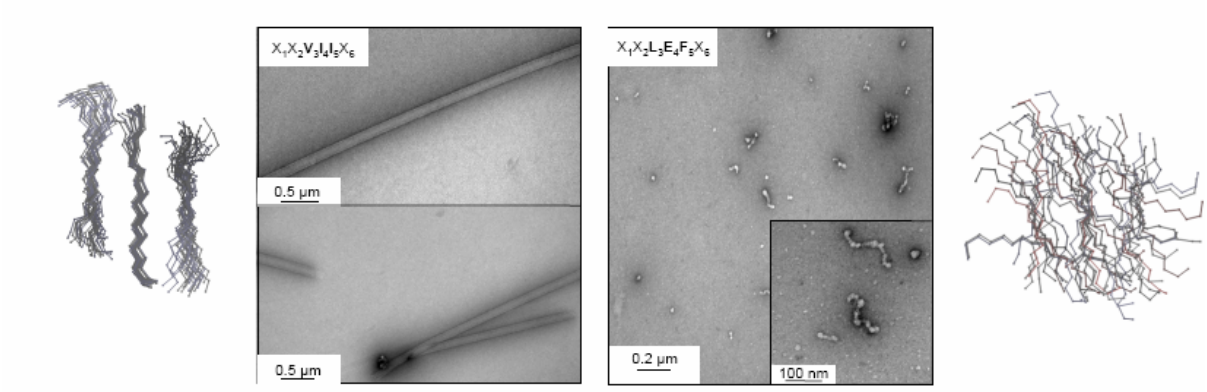


Figure 2. Ensemble structures of STVIIE (LHS) and STLEFE mutant (RHS), produced from sampling with RAFT and DGIS, coloured according to individual probability as in Figure 1. Electron micrographs of XXVIIX and XXLEFX (centre), where X represents a mixture of all residues, reproduced from Serrano and colleagues.

Future directions

The results presented here are very encouraging. However, in order to fully differentiate between β-aggregation and fibrillogenesis a higher level of detail is required in our model. To this end, a more realistic model is being developed, which incorporates rotameric side chains, to allow cross-sheet interactions to be adequately modelled.

Publications

Thomas G.L., Sessions R.B. & Parker M.J. (2005) Density guided importance sampling: application to a reduced model of protein folding. *Bioinformatics.* **21,** 2839-2843.

Funding

This work was funded by the BBSRC and the MRC.

Employing sparse NMR restraints in a protein folding algorithm: towards rapid structure calculations

Moza Al-Owais, Arnout Kalverda, Malcolm McLean and Martin Parker

Introduction

The quality of NMR-derived protein structures is heavily dependent on the number of inter-proton distance restraints obtained by NOE experiments. It is the assignment of a sufficient set of such NOEs that is the most time consuming part of NMR structure calculations. For large proteins, to obtain high-quality spectra, it is often necessary to deuterate samples to a high level (sometimes > 90%) in order to assign backbone resonances using traditional methods. The concomitant reduction or loss of proton NOEs is a serious problem. Thus, novel approaches that decrease the dependence on NOE restraints, and which operate successfully at high levels of deuteration, are necessary to increase the efficiency of NMR in the structural characterisation of proteins. We are developing novel experimental methods for the rapid acquisition of NMR distance restraints and exploring the use of these in our own *ab initio* protein folding algorithm.

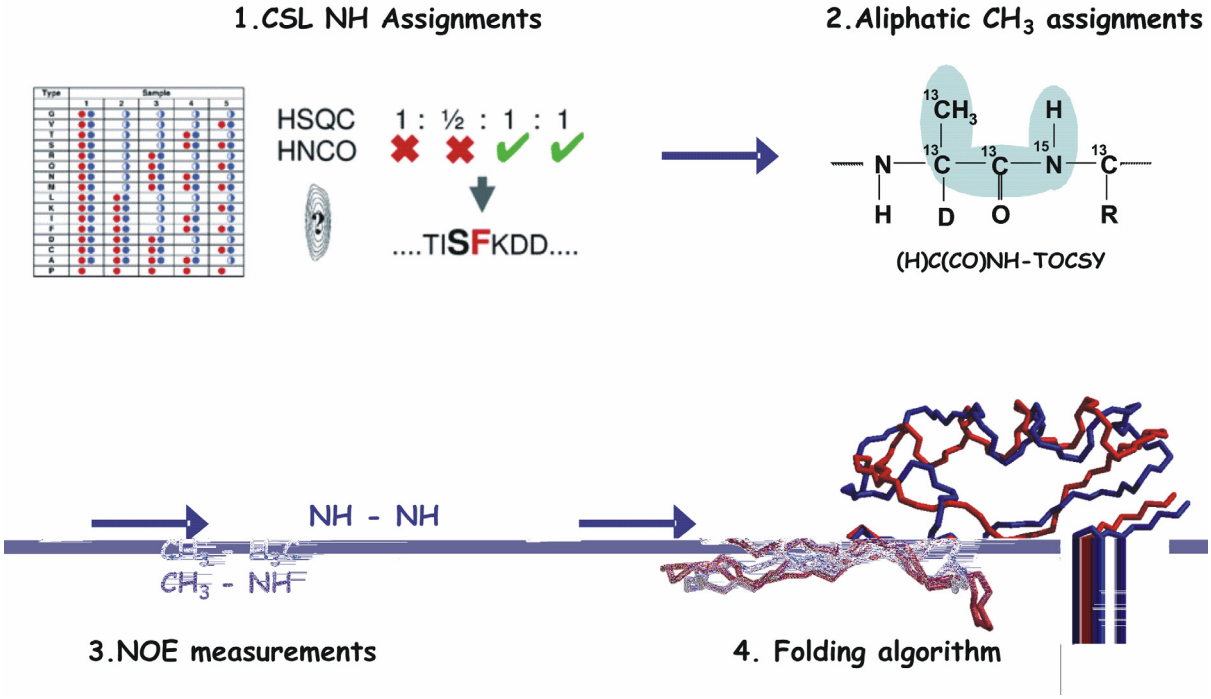


Figure 1. The 4 steps of our NMR protein structure calculation method. Step 4 shows superimposition of PDB structure of protein G (red) with structure calculated using our method (blue).

The method

A schematic of our method is shown in Figure 1. ***Step 1:*** backbone NH assignments are obtained using Combinatorial Selective Labelling (CSL) of amino acids. The CSL method uses a limited number of samples (5 in total), each containing a different combination of ^{13}C,^{15}N- and ^{14}N-labelled amino acids. The labelling pattern for a particular amino acid type across the five samples is unique, and assignments are made simply by looking at the patterns of cross-peak intensities in HSQC and HNCO spectra across the samples. This is a very robust, sensitive and technically undemanding method, compared with traditional sequential assignment methods. ***Step 2:*** Sample 1 in the CSL method is fully ^{13}C,^{15}N labelled. It is also fully deuterated, except at the backbone amide groups and at specific side chain CH_3 groups of the aliphatic amino acids Ala, Val, Leu and Ile. By running a special TOCSY experiment, it is possible to correlate the spins of these CH_3 groups with those of the backbone NH groups. Thus, we can assign the CH_3 group resonances to specific positions in the protein

sequence. ***Step 3:*** NOE experiments are run to obtain NH-HN, CH_3-H_3C and CH_3-HN distances. ***Step 4:*** the set of resulting NOE distances are used as guiding restraints in an otherwise *ab initio* folding algorithm.

Results for a model system

We have applied our method to a model study object as proof of principle: protein G from *Streptococcus*. Using only ~30 unique NOE distances we obtain a backbone structure with a C_α RMS of 2.6 Å with respect to the PDB structure (final panel in Figure 1). This is an encouraging result, considering that the PDB structure required 854 NOE restraints, 60 H-bond restraints, and 54 ϕ and 51 ψ angle restraints.

Publications

Parker, M. J., Aulton-Jones, M., Hounslow, A. & Craven, C. J. (2004) A Combinatorial Selective Labelling Method for the Assignment of Backbone NMR Resonances. *J. Am. Chem. Soc.* **126**, 5020-5021.

Craven, C. J., Al-Owais, M. & Parker, M. J. (2007) A Systematic Analysis of Backbone Amide Assignments Achieved via Combinatorial Selective Labelling of Amino Acids. *J. Biomol. NMR* (in press).

Funding

This work was funded by the BBSRC.

Structure and function of AhrC, the arginine transcription factor from *Bacillus subtilis*

James. A. Garnett, Peter. G. Stockley & Simon. E. V. Phillips

Introduction

In *Bacillus subtilis* the concentration of L-arginine is controlled by the hexameric transcriptional regulator AhrC, which interacts with 18 bp pseudo-palindromic ARG boxes in the promoters of arginine biosynthetic and catabolic operons. The biosynthetic promoters contain two or three binding sites and a unique property of AhrC is its apparent ability to also stimulate the catabolic pathway through a single operator site.

The crystal structure of the apo-form of AhrC, in the absence of the arginine corepressor, was solved in Leeds (**Fig. 1a**). Each subunit has two domains, with the C-terminal domains forming the core of the hexamer. The N-terminal domains contain a winged helix-turn-helix DNA-binding motif, and are arranged around the periphery.

Crystallographic studies

To study the effects of ligand binding in the C-terminal hexamer (CAhrC) and ARG box interactions with the N-terminal domains of AhrC (NAhrC), the two independent domains were expressed, purified and crystallized, with CAhrC in the presence of L-arginine. In addition we have crystallized a complex between a dimer of NAhrC and an 18 bp blunt ended oligonucleotide, corresponding to an ARG box. The three crystal structures have been solved at 1.00 Å (NAhrC), 1.95 Å (CAhrC) and 2.85 Å (NAhrC operator complex).

The structures reveal how AhrC functions as a transcriptional repressor. AhrC has 32 symmetry and L-arginine binding to AhrC triggers an approximate rotation of 15˚ of one trimer with respect to the other, when compared to the core of apo-AhrC. When NAhrC domains recognise and bind to their ARG box, the flexible wing of the DNA binding motif is stabilised across the minor groove and the recognition helices (α3) lie in the major groove, bending the ARG box by approximately 30° towards the NAhrC dimer.

We have modelled the effects of L-arginine and operator binding to intact AhrC using the structures of the domains and the operator complex (**Fig. 1a,b,c**). These models suggest that the rotation induced by ligand binding results in the DNA binding domains coming together, although this alone will not bring the N-terminal domains into the correct arrangement. A reorientation of these domains is needed to reproduce the NAhrC dimer observed in the operator complex and create the fully activated transcription factor. We have also constructed models of the repression complexes at the *argC* biosynthetic operator, where the promoters are bent around AhrC, disrupting any attempts by RNA polymerase to bind and initiate transcription.

Funding

We are grateful to the BBRSC for financial support.

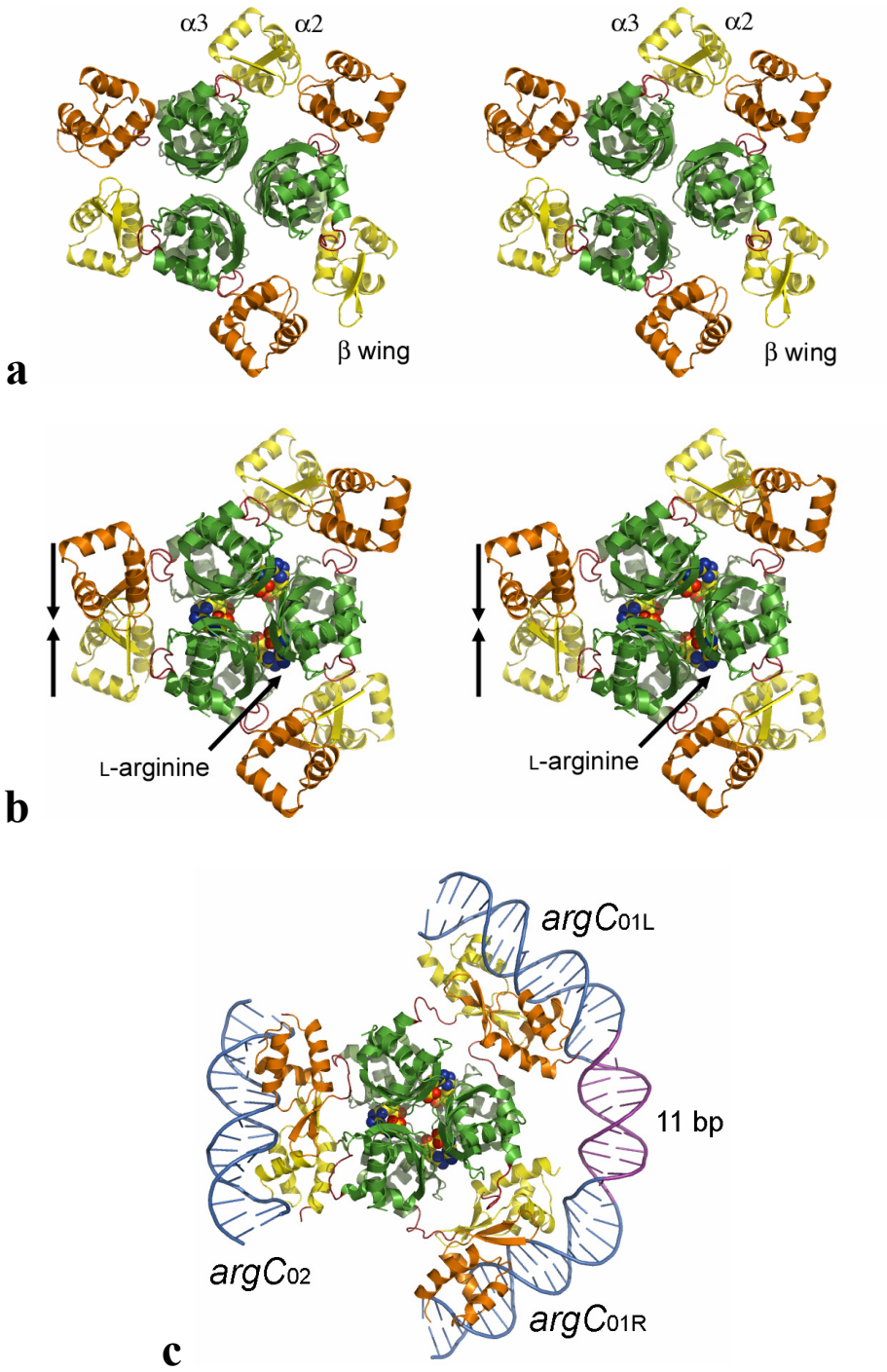


Figure 1: Models of L-arginine activation of AhrC and the consequent binding to its *argC* biosynthetic operator site. The C-terminal domains of AhrC are coloured light green (top trimer) and dark green (lower trimer). L-arginine ligands are shown as spheres. The N-terminal domains of AhrC are coloured yellow (associated with the upper trimer) and orange (associated with the lower trimer), with the flexible linker connecting the N- and C-terminal domains coloured red. (**a**) Crystal structure of apo-AhrC[1]. The N-terminal domains lie around the periphery of the protein core with imperfect symmetry. The DNA binding structures are labelled. (**b**) Model of holo-AhrC. Binding of L-arginine causes a rotation between the trimers, bringing the N-terminal domains together. (**c**) Model of holo-AhrC in complex with its *argC* operator. ARG boxes are coloured blue and the intervening DNA spacer is purple. For AhrC to bind to its operator site there must be a rearrangement in the N-terminal domains, orientating the wings and recognition helices into the correct position for DNA binding. The third ARG box is believed to be contacted through looping out the intervening DNA.

The structure of the complex between Bacteriophage T7 Endonuclease I and a synthetic Holliday Junction

Jonathan M. Hadden, Stephen B Carr 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 and the wild-type protein with metals bound, both in the absence of DNA.

Recently, we have grown crystals of a stable complex of endonuclease I and a synthetic Holliday junction and have collected X-ray diffraction data to 3.1Å resolution using station ID23-2 at the European Synchrotron Radiation Facility (ESRF). The structure of this complex has been solved using molecular replacement with the structure of free endonuclease I as the search model.

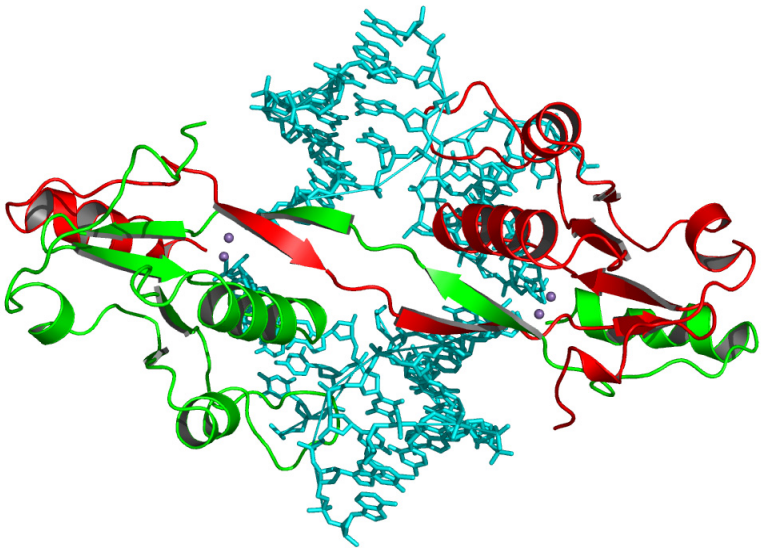


Figure 1: Crystal structure of the complex between T7 endonuclease I and a synthetic Holliday junction. Endonuclease I subunits are shown in red and green, DNA in cyan and calcium ions in purple.

Model fitting and refinement of the complex is underway (R_{cryst} 31.6 %, R_{free} 35.1%), with all four of the helical DNA arms present, and the paths of the exchanging strands clear. A number of interesting features regarding the mechanism of DNA binding and cleavage have already become apparent.

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.

In vitro studies of sensor kinase and response regulator proteins of the Prr two-component signal transduction pathway in *Rhodobacter sphaeroides*

Christopher Potter, Eun-Lee Jeong, Peter Henderson and Mary Phillips-Jones

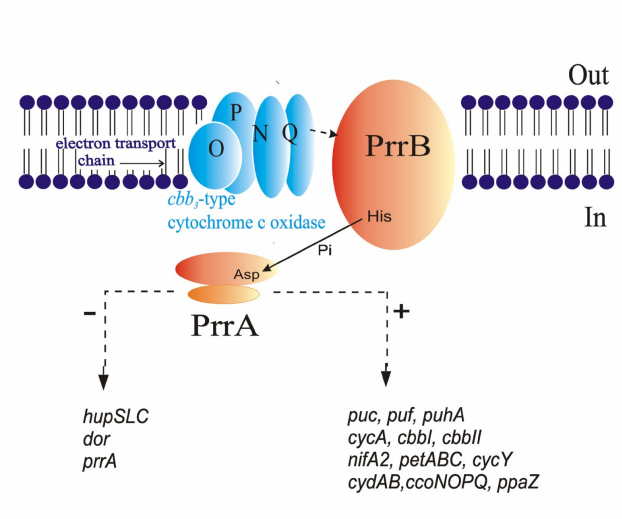


Fig. 1. The Prr two-component signal transduction pathway in Rhodobacter sphaeroides

Prr is a global regulatory system belonging to the 'two-component' family of signal transduction proteins (Fig. 1). It controls a large and diverse range of genes in *Rhodobacter sphaeroides* including those for synthesis of the photosynthetic apparatus, electron transport, nitrogen and carbon fixation, anaerobic respiration, [NiFe] hydrogenase and aerotaxis, in response to changing conditions of environmental redox potential; Prr is therefore a pivotal regulator in the complex switch between aerobic and anaerobic lifestyles and the optimum use of reducing power in this versatile bacterium. The aim of our work is to characterise fully the mechanisms of signal sensing, signal transduction and DNA target binding, in order to provide insights into how globally-important processes such as carbon and nitrogen fixation, as well as photosynthesis, are coodinately regulated by just one system in one bacterium. PrrB is the membrane-bound sensor kinase that senses changes in redox potential. Upon anaerobiosis, PrrB becomes autophosphorylated and transfers the phosphoryl signal to response regulator PrrA. Once phosphorylated, PrrA~P then positively or negatively regulates gene expression.

Recent advances: PrrB.

Previously we showed that the purified, detergent-solubilised intact membrane sensor PrrB is functional in autophosphorylation, phosphotransfer and phosphatase activities. Here, we confirm that it also senses and responds directly to its environmental signal, redox potential, *in vitro*; strong autophosphorylation of PrrB occurred in response to dithiothreitol-induced reducing conditions (and levels increased in response to a wide 0.1 - 100 mM DTT range), whilst under oxidising conditions, PrrB exhibited low, just detectable levels of autophosphorylation (Fig. 2). The clear response of PrrB to changes in reducing conditions confirmed its suitability for *in vitro* studies to identify modulators of its phosphorylation signalling state, and so far has

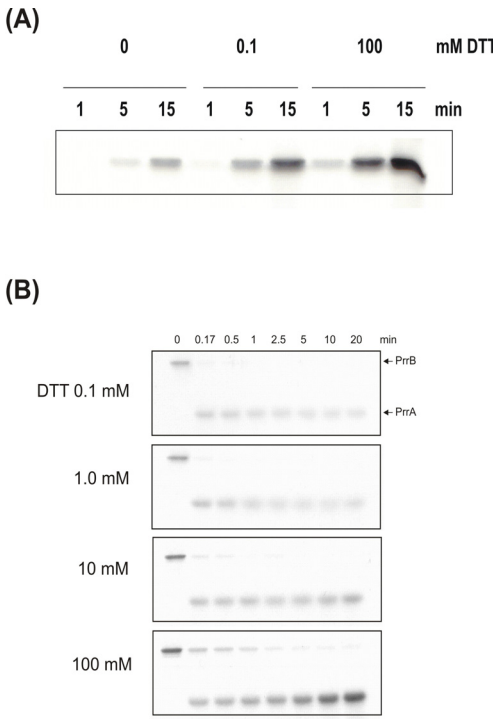


Fig. 2. Modulation of PrrB autophosphorylation and phosphotransfer to PrrA by dithiothreitol in vitro. (A) Effect of DTT on PrrB autophosphorylation in assays initiated using[γ-32P]ATP. (B) Effect of DTT on PrrB autophosphorylation and subsequent phosphotransfer to PrrA. PrrB was allowed to autophosphorylate in the presence of DTT for 20 min; PrrA was then added (final PrrB:PrrA ratio of 1:2.5), and levels of PrrB~P and PrrA~P measured.

been used to investigate whether PrrB might sense more than one redox-related signal, such as signals of cell energy status. NADH, ATP and AMP were found to exert no detectable effect on maintenance of the PrrB~P signalling state. By contrast, ADP produced a very strong increase in PrrB~P dephosphorylation rate, presumably through the back-conversion of PrrB~P to PrrB.

Recent advances: PrrA.
Using gel filtration, analytical centrifugation and NMR diffusion measurements, our recent studies of PrrA have shown that treatment of the protein with a phosphate analogue, BeF_3^-, results in dimerisation of the protein, producing a protein that binds DNA. Upon addition of BeF_3^-, the inhibitory activity of the N-terminal regulatory domain on the C-terminal DNA-binding domain is relieved, after which PrrA becomes capable of binding DNA as a dimer. The interaction surface of the DNA-binding domain with the regulatory domain of PrrA was identified by NMR as being a well-conserved region centred on helix α6, which is on the face opposite from the DNA recognition helix. This suggests that there is no direct blockage of DNA binding in the inactive state, but rather that PrrA dimerisation promotes a correct arrangement of two adjacent DNA-binding domains that recognises specific DNA binding sequences.

Publications
Potter, C.A., Jeong, E-L., Williamson, M.P., Henderson, P.J.F. & Phillips-Jones, M.K. (2006) Redox-responsive *in vitro* modulation of the signalling state of the isolated PrrB sensor kinase of *Rhodobacter sphaeroides* NCIB 8253. *FEBS Lett*. **580**, 3206-3210.

Laguri, C., Stenzel, R.A., Donohue, T.J., Phillips-Jones, M.K. & Williamson, M.P. (2006) Activation of the global gene regulator PrrA (RegA) from *Rhodobacter sphaeroides*. *Biochem*. **45**, 7872-7881.

Collaborators
Mike P. Williamson, Department of Molecular Biology & Biotechnology, University of Sheffield.
Timothy J. Donohue, Department of Bacteriology, University of Wisconsin, Madison, Wisconsin USA.

Funding
This work was funded by the BBSRC.

Crystal structure of the response regulator VicR DNA-binding domain

Chi Trinh, Yang Liu, Simon E.V. Phillips and Mary Phillips-Jones

VicR is the response regulator component of the VicKR (YycFG) two-component signal transduction system (TCS) of infection agent *Enterococcus faecalis* (Fig. 1). TCSs are the main mechanism by which bacteria sense and respond to environmental change, and they are found in almost all species so far examined. The VicKR system is a member of the YycFG/Vic subfamily that is essential for cell viability. Only a few members of this important subfamily of essential regulators have been identified and characterised so far, and all exhibit homology with known members (e.g. 79 - 90 % identity with *E. faecalis* VicR). Their essentiality makes them attractive targets for novel classes of antibacterial agents. In all cases, it is the response regulator (RR) proteins that have been shown to be essential, whilst the sensor kinase components are not essential for all members of the YycFG/Vic TCS subfamily.

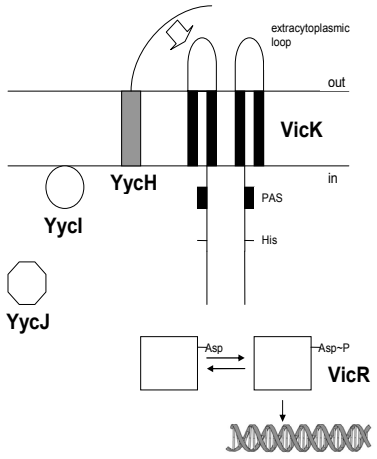


Fig. 1. The Vic pathway of *Enterococcus faecalis*

The target genes regulated by YycF/VicR include those critical for cell wall biosynthesis or cell membrane composition, cell division (*ftsAZ*), virulence and genetic competence and biofilms. The DNA sequence recognised by this subfamily of regulators is highly conserved amongst species. In spite of an expanding list of target genes identified, and elucidation of the crystal structure of the highly conserved *N*-terminal receiver domain of one homologue, RR02rec of *S. pneumoniae* YycF, no other structural information is available, including none at all on the effector domains of this 'essential' group of DNA-binding proteins. Here we present the crystal structure of the putative DNA-binding effector domain of VicR, $VicR_c$ (PDB: 2HWV).

$VicR_c$ belongs to the winged helix-turn-helix family of DNA-binding proteins. It is very similar to the DNA-binding domains of *E. coli* PhoB and OmpR, despite low sequence similarity, but differs in two important loops (Fig. 2): (1) the α-loop, that links the two helices of the helix-turn-helix motif, is different from OmpR, but similar to that in PhoB, where it has been implicated in contacting the σ-subunit of RNA polymerase (Fig. 2); and (2) the loop following the helix-turn-helix motif is different from PhoB (Figs 2 and 3). YycF/VicR, PhoB and *Bacillus subtilis* PhoP regulators all recognise almost identical DNA sequences and, although there is currently no experimental evidence linking this loop with the DNA, the structure is

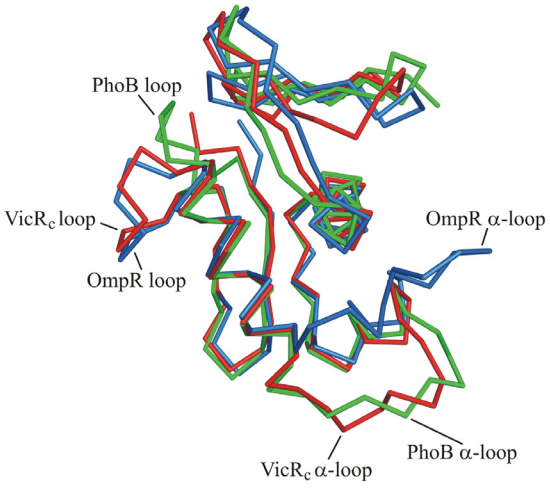


Fig. 2. Structural comparison of the Cα backbones of VicRc (red), PhoB (green) and OmpR (blue). The α-loops and the loops linking α3 to the C-terminal β-hairpin are labelled.

consistent with possible involvement in selective DNA recognition or binding (Fig. 3).

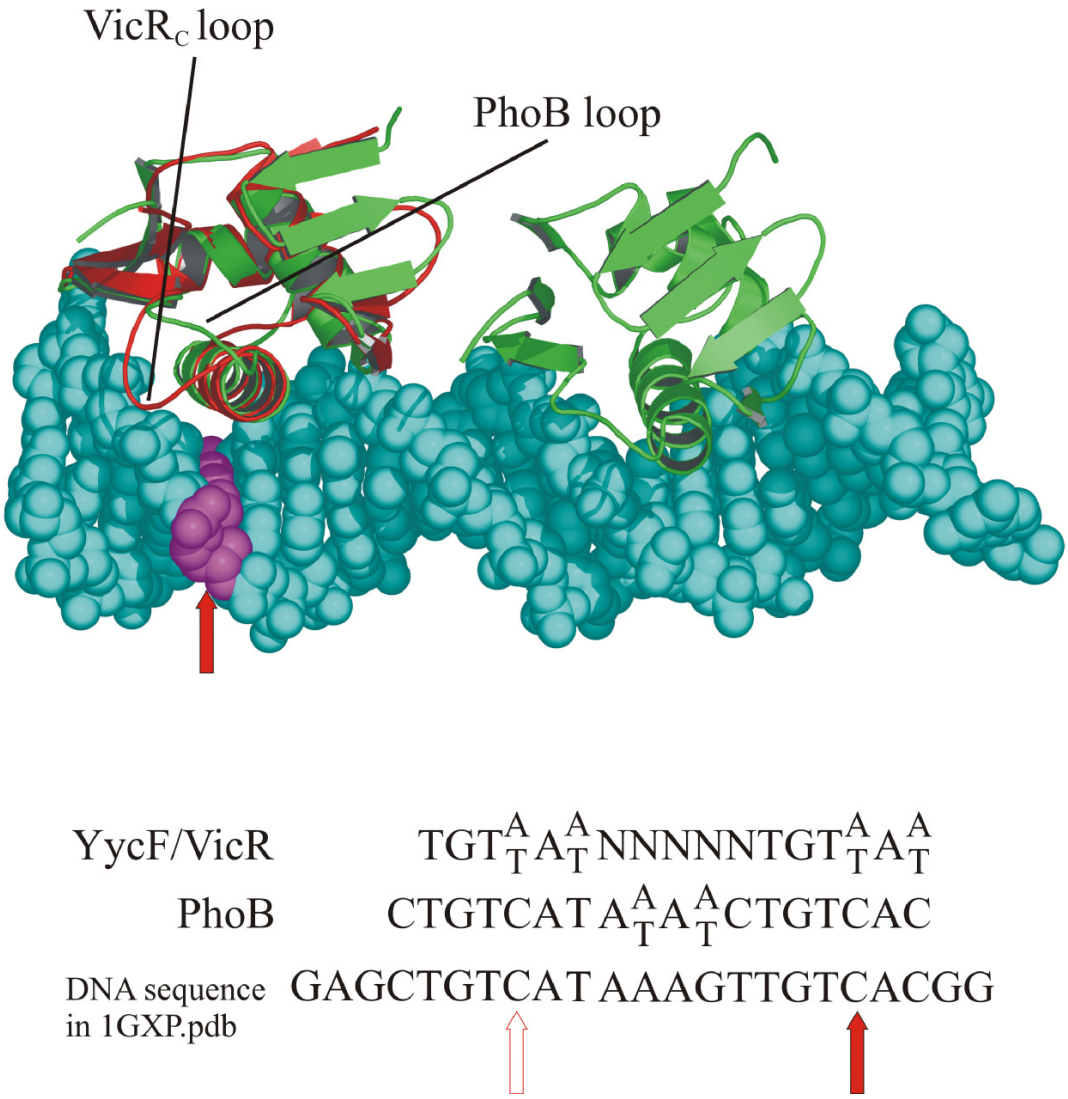


Fig. 3. Ribbon representation showing two molecules of PhoB (green) in complex with DNA shown in space filling representation (pdb code: 1gxp). VicRc is shown in red, superimposed onto one PhoB monomer. The consensus sequences for YycF/VicR homologues and PhoB targets are also shown, with the C to A/T substitution mentioned in the text indicated by red arrows in each repeat. The filled red arrow indicates the corresponding base pair in the model, that has been coloured magenta and lies close to the recognition helix. The loop region of PhoB does not reach the DNA, while the VicRc loop could make contact with it.

Publications

Trinh, C.H., Liu, Y., Phillips, S.E.V. & Phillips-Jones, M.K. (2007) Structure of the response regulator VicR DNA-binding domain. *Acta Cryst.* **D63**, 266-269.

Funding

This work was funded by the BBSRC.

Early states during protein folding

Alice Bartlett, Claire Friel, Stuart Knowling, Daniel Lund, Victoria Morton, Graham Spence, Alastair Smith and Sheena Radford

Introduction

To understand how a protein folds, we need to define the properties of the folding landscape and how the polypeptide chain moves through this landscape. For example, does a protein fold quickly on a smooth landscape directly from the unfolded to the folded state, or is the landscape more rugged, involving partially folded intermediates that serve to complicate the folding kinetics by adding kinetic traps? In order to answer these questions, we need to be able to detect all the species populated during folding and to characterise their structural, dynamic and spectroscopic properties in as much detail, and at as high a resolution, as possible. Whilst this can be readily achieved for the native state using NMR or X-ray methods, the transient nature, conformational dynamics and conformational heterogeneity of earlier states makes them much more difficult to study. We have previously shown that Im7 folds to its native state *via* an on-pathway three-helical intermediate, which is stabilised by both native and non-native interactions (Figure 1). We also know that the six residues that comprise helix 3 in the native state play no part in the intermediate, only docking onto the developing structure after the rate-limiting transition state has been traversed. We have been using a range of techniques, from experimental methods to computer simulation, to determine information about the unfolded and intermediate states of Im7, with a view to defining its folding mechanism in atomic detail.

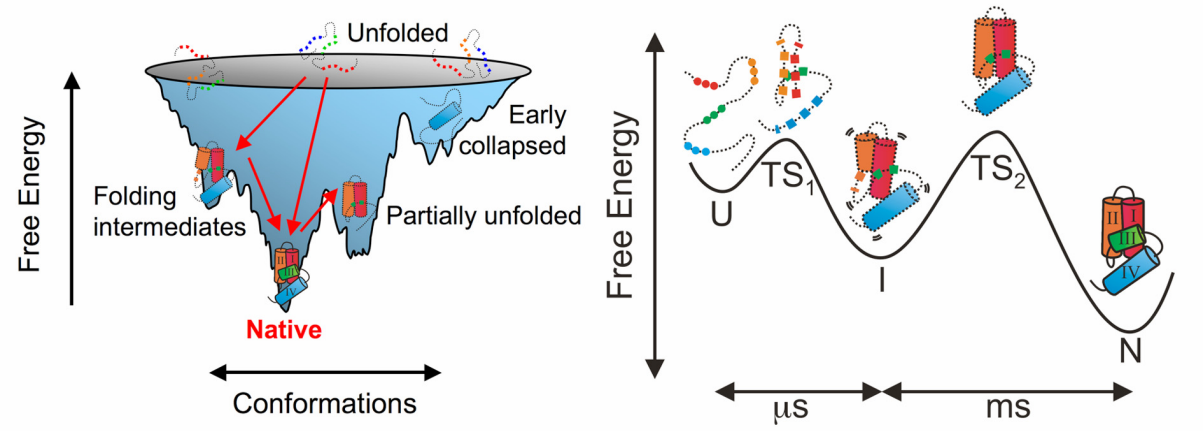


Figure 1: (Left) Schematic representation of a multi-dimensional folding landscape, with a large number of unfolded conformations at the top of the landscape, and a small number of native conformations at the bottom. (Right) Schematic representation of the folding landscape of Im7. The unfolded state (U) is highly dynamic, but contains four independent clusters of hydrophobic residues. By the early transition state (TS1), these clusters have collapsed, forming the native topology. As folding proceeds to the intermediate (I), helical secondary structure is consolidated and stabilised by native and non-native interactions, although helix 3 (green) is not involved. As the protein traverses the rate-limiting transition state (TS2) there is a structural rearrangement, revealing the binding site for helix 3, which acts to lock the protein in its native conformation (N).

The intermediate state

It has been possible to trap the normally transiently populated intermediate species at equilibrium by truncation of two key side chains in helix 3, preventing Im7 reaching its native state. By studying this variant with a range of biophysical techniques it has been possible to determine new information about the structural properties of the folding intermediate. The use of small-angle X-ray scattering and a pulsed-field gradient NMR technique has revealed that the intermediate species is as compact as the native state, but with a slightly different shape. More complex multi-dimensional NMR experiments were used to

determine structural and dynamic information, confirming the three-helical nature of the intermediate at the residue-specific level and revealing that whilst the intermediate remains a compact species, it is much more conformationally dynamic than the native state of the wild type protein. Using molecular dynamics simulations restrained by experimental parameters (ϕ-values, hydrogen exchange protection factors and chemical shifts) it has been possible to calculate an ensemble of structures with which we can examine side-chain contacts, helical orientations and the breadth of the ensemble, helping us interpret experimental data and giving us a fuller understanding of the intermediate ensemble.

The unfolded state

Folding in our experiments is initiated from a protein unfolded in high concentrations of urea. Again, we have used NMR to determine the conformational properties of this state (in 6 M urea at 10°C). Pulse field gradient NMR experiments showed the unfolded state has a hydrodynamic radius of ~30Å, much larger than the 19Å displayed by the native state. Analysis of NOE's shows that the urea unfolded state is highly solvated by both water and urea and does not contain any residual helical structure. However, measurements of the backbone relaxation dynamics suggest the residues that go on to form helices in the native state form four hydrophobic clusters in the urea unfolded state, although these clusters do not interact with one another. These data suggest that, even in the highly unfolded state populated at the start of a folding reaction, Im7 is primed to fold rapidly towards the native helical state.

Future work

We now have structural and dynamic information on every state we can detect on the Im7 folding pathway, with experiments to provide structural information about the first transition state almost complete. Combining these data with restrained molecular dynamics simulations, our aim is to push our knowledge of the folding of Im7 towards atomic detail. Experiments are also underway to explore the early events on the folding pathway in more detail by determining how the helices collapse to form the first transition state ensemble.

Collaborators

Geoffrey R Moore, University of East Anglia
Michele Vendruscolo, University of Cambridge

Publications

Gsponer, J., Hopearuoho, H., Whittaker, S.B., Spence, G.R., Moore, G.R., Paci, E., Radford, S.E. & Vendruscolo, M. (2006) Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Nat. Acad. Sciences USA.* **103,** 99-104.

Le Duff, C.S., Whittaker, S.B., Radford, S.E. & Moore, G.R. (2006) Characterisation of the conformational properties of urea-unfolded Im7: Implications for the early stages of protein folding. *J. Mol. Biol.* **364**, 824-835.

Whittaker, S.B., Spence, G.R., Grossmann, J.G., Radford, S.E. & Moore, G.R. (2007) NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J. Mol. Biol.* **366**, 1001-1015.

Funding and acknowledgments

This work is funded by the BBSRC and The Wellcome Trust. We thank Keith Ainley and Anne Holland for technical support.

Single molecule protein folding and kinetics

Jennifer Clark, Sara Pugh, Tomoko Tezuka-Kawakami, Chris Gell,
Alastair Smith and Sheena Radford

Introduction

Single molecule fluorescence resonance energy transfer (smFRET) permits the study of biological macromolecules, and can reveal new information about rarely populated species that cannot be detected using traditional biochemical or biophysical assays. Using smFRET, proteins are labelled with suitable fluorophores at selected points within the protein of interest and the energy transfer between fluorophores can be used to provide information as to their spatial separation. This energy transfer is generally only possible over distances between 20-100Å, depending on the fluorophores employed.

Diffusion based FRET studies

The folding mechanisms of Im7 and Im9 have been extensively studied in our laboratory using ensemble techniques. Interestingly, despite possessing 60% sequence identity, Im9 folds *via* a two-state mechanism, while Im7 folds *via* a three-state mechanism in which an on-pathway intermediate is populated transiently during folding. Using smFRET methods, the folding of Im7 and Im9 have been probed as a function of urea concentration and, for Im7, with particular attention to the changes in compactness of both the native, intermediate and denatured ensembles as measured by the relative FRET efficiency (the proximity ratio). In addition, the widths of the resolved species in the proximity ratio histogram can also contain information about the dynamics and heterogeneity of the species within each state; information which is not readily available using ensemble studies.

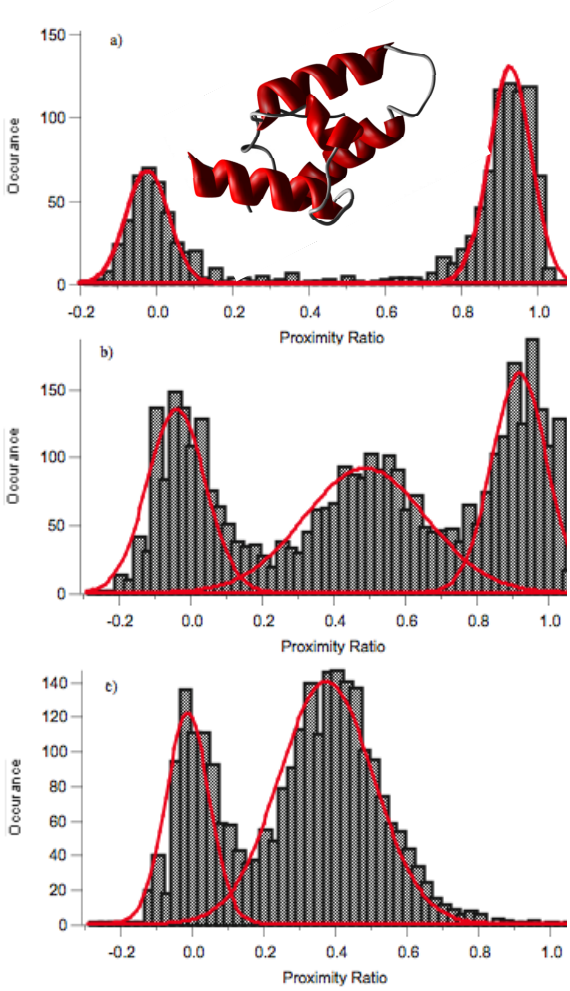


Fig 1: Frequency histograms of smFRET data for Im7 Q17C–K70C at pH 7, 10°C, 0.4M sodium sulphate, 50mM sodium phosphate, displaying a) native, b) 50% folded and c) unfolded. The peak at proximity ratio ~0 results from molecules with only donor

The results revealed that at equilibrium Im7 populates two species, folded and unfolded, as measured by smFRET. Intriguing changes in the proximity ratio of both the native and denatured states were also observed as a function of the concentration of urea, suggesting that the conformational properties of these species depend on the concentration of denaturant.

In addition to the diffusion equilibrium experiments carried out using the immunity proteins, diffusion studies have been carried out using a labelled variant of the B domain of protein A (BdpA), a 58 residue three-helix bundle protein which has been extensively studied using ensemble techniques and been the subject of a range of folding simulations. BdpA has been shown to be one of the fastest folding proteins known (k_f = 249,000s^{-1} at 45°C and 0M

GuHCl). Using smFRET over a range of temperatures, and measuring the response of the protein to denaturant (GuHCl), we have modelled the proximity ratio distributions using Monte Carlo simulations and used the resulting fits to extract the rates of folding and unfolding. The methods developed have added a new dimension to the information available from single molecule diffusion experiments that should be widely applicable to many systems.

Immobilised protein FRET

Using a tethering and encapsulation system involving a lipid bilayer and liposome, fluorescently labelled immunity proteins can be studied individually using a total internal reflection fluorescence microscope (TIRFm) setup. This permits us the opportunity to observe the folding and unfolding of an immunity protein in real time. This technique has the potential to elucidate the transient population of rare states that are lost to bulk averaged methods, thus providing the exciting potential to map the protein folding landscape in a previously unexplored manner. We have set up such experiments in Leeds and studies of the real time folding of the bacterial immunity proteins are in progress.

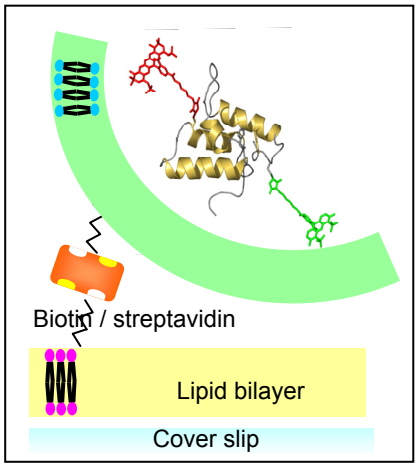


Fig 2: Schematic representation of the immoblisation strategy employed in the TIRFm experiment.

Publications

Tezuka-Kawakami, T., Gell, C., Brockwell, D.J., Radford, S.E., & Smith, D.A. (2006) Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys J.* **91**, L42-L44.

Funding and Acknowledgements

We would like to thank Graham Spence for helpful discussions. We acknowledge the Wellcome Trust, the BBSRC and the University of Leeds for funding.

Investigating fibril structure and assembly using mass spectrometry

Andrew Smith, Thomas Jahn, Sarah Myers, John Hodkinson,
Sheena Radford and Alison Ashcroft

Introduction

The assembly of normally soluble proteins into pathological aggregates is of great interest due to their role in important illnesses such as Alzheimer's and Parkinson's diseases. The human protein beta-2-microglobulin (β_2m) assembles into amyloid fibrils *in vivo* causing the debilitating condition known as dialysis related amyloidosis (DRA). We have been investigating the mechanism of aggregation of β_2m into amyloid using electrospray ionisation mass spectrometry (ESI-MS), focusing on the conformational properties of aggregation-prone monomers, as well as the structure of the fully assembled fibril.

Investigating fibril structure

Fibrils which have the cross-β structure characteristic of amyloid fibrils found in the joints of DRA sufferers can be formed from wild-type β_2m under acidic conditions *in vitro*. The morphology of the fibrils formed *in vitro* depends on the conditions employed. Incubation at low ionic strength at pH 2.5 yields long (>μm) straight fibrils, while at pH 3.6 short (<500nm) fibrils form. At higher ionic strength (0.2M-0.4M) fibrils which exhibit a curved, nodular morphology are formed. To determine the conformation of β_2m within fibrils with these different morphologies, limited pepsin proteolysis of the fibrils was performed, followed by analysis of the fragments by ESI-MS and tandem mass spectrometry (MS/MS). For comparison, proteolytic degradation of the monomer and seven synthetic peptides representing the entire sequence of β_2m was undertaken. The differing morphologies of the fibrils resulted in distinct digestion patterns as shown in Figure 1.

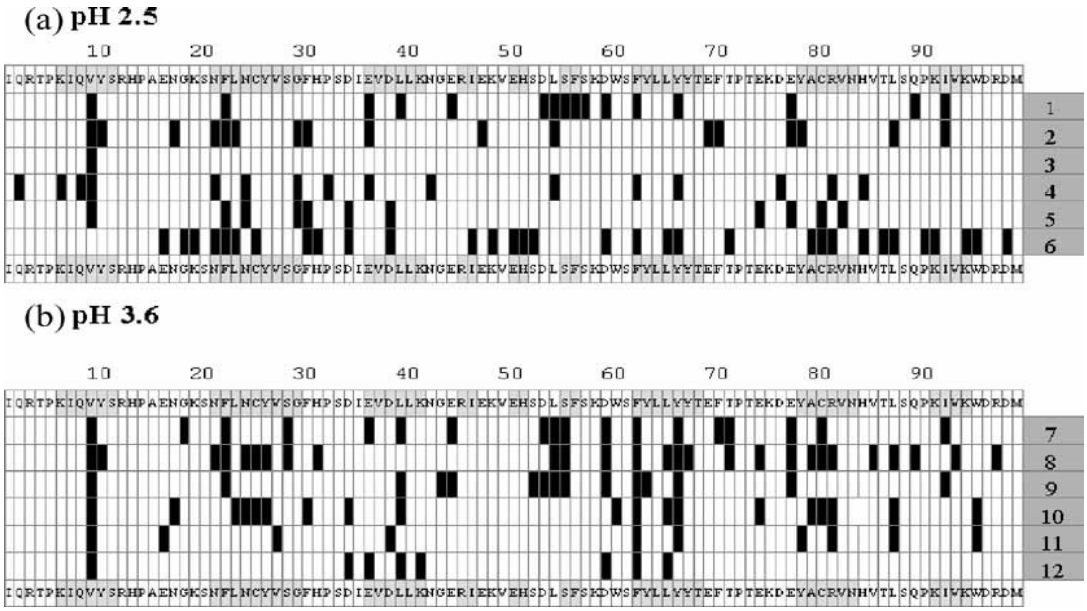


Fig 1. Summary of the cleavage sites observed by ESI-MS(/MS) after pepsin proteolysis of β_2m fibrils of different morphological type. (a) Samples at pH 2.5: (1) and (2) monomeric β_2m, (3) and (4) long, straight fibrils formed at low ionic strength, and (5) and (6) nodular fibrils formed at high ionic strength. (b) Samples at pH 3.6: (7) and (8) monomeric β_2m, (9) and (10) short, straight fibrils formed at low ionic strength, and (11) and (12) nodular fibrils formed at high ionic strength. Each pair of reactions (e.g (1) and (2)) were incubated with pepsin (1:100 w/w) for 15min or for 24hr, respectively, at 25°C. The 99-residue amino acid sequence is also shown. The regions in grey shading indicate the location of β-strands in native β_2m. The vertical bars in (a) and (b) depict the C-terminal residue of an identified peptide fragment and hence indicate the proteolytic cleavage sites.

The results shown in Figure 1 indicate that the long, straight fibrils formed at pH 2.5 (low ionic strength) show enhanced protection from limited proteolysis (only one cut site at valine 9) relative to the fibrils formed at pH 3.6. The protected fibril core is reduced from ~90 residues in the long, straight fibrils to only ~30 residues in the nodular forms. Only the A strand (first 11 residues) of β_2m is unprotected in all fibril types.

Investigating fibril assembly

As the assembly mechanism of β_2m fibrils remains unclear, the identification and characterisation of early oligomeric species was undertaken. Using ESI-MS the formation of β_2m oligomers was monitored over time at pH 2.5 and 3.6 and validated using analytical ultracentrifugation (AUC) and Thio-T fluorescence (Thio-T). The oligomeric species detected at each time point for both pH values is shown in Figure 2.

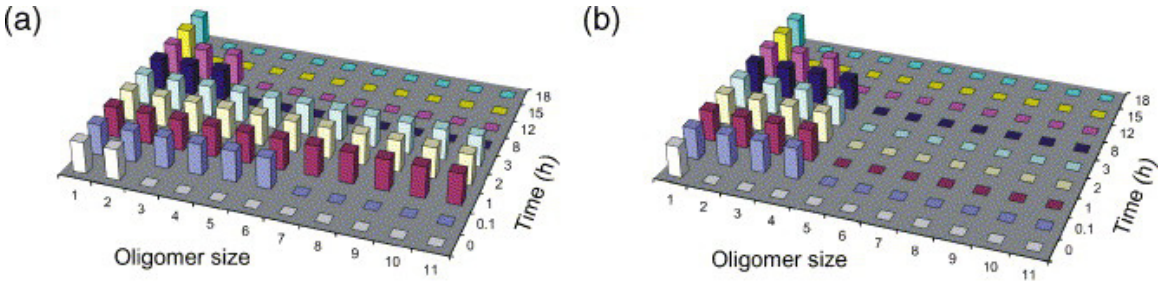


Fig 2. Oligomer distributions observed during β_2m assembly measured by nanoESI-MS over a range of *m/z* 3200–5500. (a,b) Summary of species detected during fibril assembly at 37°C at (a) pH 3.6 and (b) pH 2.5 using nanoESI-MS. Note that as the concentration of each oligomer cannot be quantified precisely by nanoESI-MS, the population of individual species is denoted as either present or absent in this chart.

The MS data, confirmed by AUC, shows that in the lag phase of pH 2.5 fibril assembly (~7 h as monitored by Thio-T fluorescence) monomeric to tetrameric species are formed (Fig. 2b). By contrast, at pH 3.6 wherein nodular fibrils are formed without a lag phase, an extensive series of oligomeric forms are detected (Fig. 2a), consistent with a mechanism involving monomer addition. This work highlights the power of ESI-MS to identify protein aggregation intermediates in complex heterogeneous systems in real time.

Investigating aggregation-competent conformations

In the next step of this work we are using hydrogen-deuterium exchange (HX) monitored by MS to investigate rarely populated, partially unfolded conformations of monomeric β_2m within the native ensemble. Comparison of the data, with parallel experiments using NMR, is beginning to unravel the complex dynamics of β_2m and to define the role of these conformational changes in fibril assembly.

Publications

Myers, S.L., Thomson, N.H., Radford, S.E. & Ashcroft, A.E. (2006) Investigating the structural properties of amyloid-like fibrils formed *in vitro* from β_2-microglobulin using limited proteolysis and electrospray ionisation mass spectrometry. *Rapid Commun. Mass Spectrom*. **20**, 1628-1636.

Smith, A.M., Jahn, T.R., Ashcroft, A.E. & Radford, S.E. (2006) Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry. *J. Mol. Biol.* **364**, 9-19.

Funding

We thank the BBSRC, The Wellcome Trust, Waters Corporation and the University of Leeds for financial support.

Determining the mechanism of bacterial fibre assembly using non-covalent electrospray ionisation mass spectrometry

Rebecca J. Rose, Sheena E. Radford and Alison E. Ashcroft

Bacterial fibre assembly

The highly pathogenic *Salmonella enterica* subspecies I produce extra-cellular Saf fibres, composed of β-sheet subunits. The Saf operon codes for four proteins: SafA, the fibre subunit; SafB, the chaperone; SafC, the usher; and SafD, a fibre capping adhesin. In the periplasm of the bacterium, a chaperone-subunit complex exists, whereby completion of a seven-strand immunoglobulin-like fold in a subunit protein is achieved by donation of a β-strand from the chaperone. Fibre assembly is known to occur by a β-strand exchange mechanism, in which an N-terminal extension 'Nte' strand from one subunit is donated to an adjacent subunit molecule, causing subunit polymerisation. This donor strand exchange mechanism causes dissociation of the chaperone-subunit complex and formation of the subunit-subunit complex (and thus fibre) at the site of the usher in the bacterial outer membrane. The details of this mechanism were largely unknown, but it was hypothesised to proceed *via* a ternary intermediate in a concerted manner, whereby the Nte strand 'zippers' into the chaperone-subunit complex, gradually displacing the chaperone β-strand.

Probing the mechanism of DSE using mass spectrometry

Mass spectrometry has been used to monitor the reaction between the chaperone-subunit complex (SafA-SafB) and the Nte peptide of an incoming subunit. Using nano-electrospray ionisation, the reaction has been followed over time (Figure 1), allowing direct observation of the decrease in concentration of the SafA-SafB non-covalent complex, and the subsequent appearance and increase in concentration of the SafA-Nte peptide non-covalent product. The presence of a SafA-SafB-Nte peptide ternary complex has been detected (Figure 1), confirming directly, and for the first time, that the reaction proceeds by a concerted mechanism involving a stable ternary intermediate.

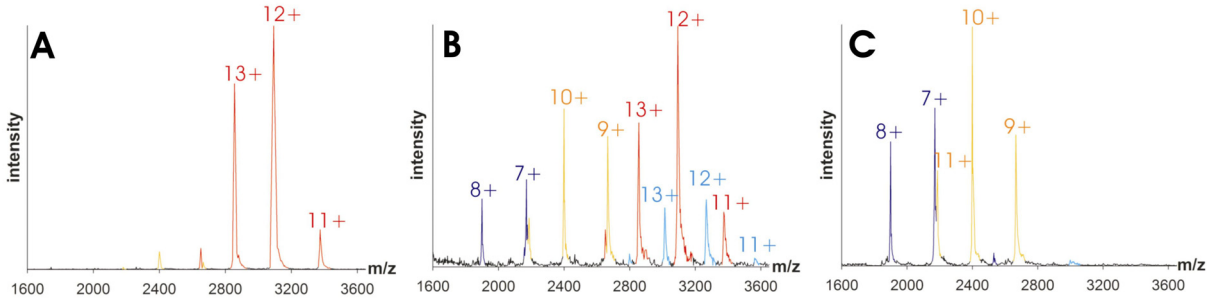


Figure 1. Mass spectra of samples taken during a donor strand exchange reaction between SafA-SafB (red peaks) and wild-type Nte peptide (not shown). (A) Sample prior to peptide addition, (B) 4 minutes after reaction initiation and (C) 24 hours after reaction initiation. These data provide the first evidence of the existence of a ternary intermediate (cyan peaks) during donor strand exchange (B). The products of the reaction, free SafB and SafA-Nte, are shown by the yellow and dark blue peaks, respectively.

Peptides of the Nte strand with single amino acid substitutions, plus variants of SafB with single-point mutations, have also been used to probe the roles of individual residues in the subunit exchange mechanism. The results showed that the reaction rates change significantly depending on nature of the residues at the interface between subunit and chaperone or within the Nte of the incoming subunit (Figure 2). This allowed residues involved in the initial docking site of the Nte to SafA-SafB to be identified, in addition to a capping interaction which stabilises the donor strand exchange product. The data point to a mechanism whereby hydrophobic residues on the N-terminal extension sequentially replace residues of the

chaperone, causing dissociation of the chaperone-subunit complex in a zip-in zip-out mechanism.

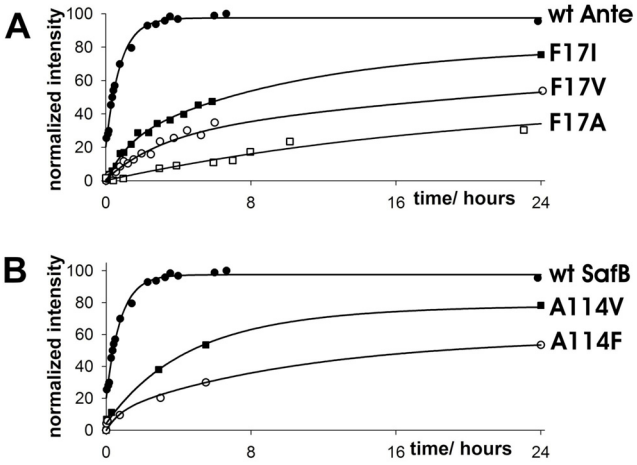


Figure 2. Kinetic traces of donor strand exchange reactions between SafB-SafA and the Nte peptide. (A) The rate of DSE decreases as residue F17 in the Nte peptide is substituted with smaller sidechains; (B) The rate of DSE also decreases as residue A114 in SafB is substituted with larger sidechains, stabilising the SafA-SafB complex. Since F17 binds to SafA at the same location as A114 in SafB, the data suggest that F17 is the initial site of Nte binding.

Current work is focussing on more complex pilus systems, including the P-pilus from *Escherichia coli*, in which six different subunits are assembled in a precise order and copy number. Using non-covalent mass spectrometry techniques, the specificity of the donor strand exchange reactions of each subunit is being monitored, providing detailed mechanistic insights into the manner by which the order of assembly is dictated and controlled.

Collaborators

This work has been in collaboration with Prof. Gabriel Waksman and Dr. Han Remaut from the School of Crystallography, UCL/ Birkbeck College, London. On-going work is in collaboration with this group, including additionally Dr. Denis Verger and Dr. Tina Daviter.

Publications

Remaut, H., Rose, R.J., Hannan, T.J., Hultgren, S.J., Radford, S.E., Ashcroft, A.E. & Waksman, G. (2006) Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted β-strand displacement mechanism. *Mol Cell* **22**, 831-842.

Funding

This work was funded by the BBSRC, Micromass/ Waters Corp. and the University of Leeds. We thank Keith Ainley for technical support.

The β_2m folding free energy landscape

Thomas Jahn, Geoffrey Platt, David Smith, Sarah Myers, Katy Routledge,
Wei Feng Xue, Timo Eichner, Martin Parker, Steve Homans and Sheena Radford

Project aims

Protein aggregation has received considerable attention in recent years because of its association with a wide range of neurological and systemic human disorders. Despite the severe impact of these diseases on the welfare of our society, we still know very little about how proteins aggregate and the structure of assembly intermediates. Furthermore, despite the ability of most proteins to assemble into amyloid-like fibrils *in vitro* under extreme conditions, how proteins form amyloid fibrils *in vivo* remains unresolved. To address this question we have been investigating the aggregation pathway of β_2-microglobulin (β_2m), a protein known to aggregate into amyloid fibrils as a consequence of long-term haemodialysis, using a large variety of biophysical techniques.

A folding intermediate as aggregation precursor

A series of folding experiments revealed that a partially folded species, the folding intermediate I_T, is significantly populated in the native state ensemble of β_2m under physiologically-relevant conditions (Figure 1). By creating the mutant protein P32G and correlating the concentration of different species populated in the native state ensemble with the rate of fibril elongation, the folding intermediate I_T was shown to be the direct precursor of β_2m fibril elongation. Structural analysis of this species demonstrates that it possesses a highly native-like structure with significant conformational changes to the β-strands A and D. The results indicate that specific perturbations to these edge-strands, which normally protect

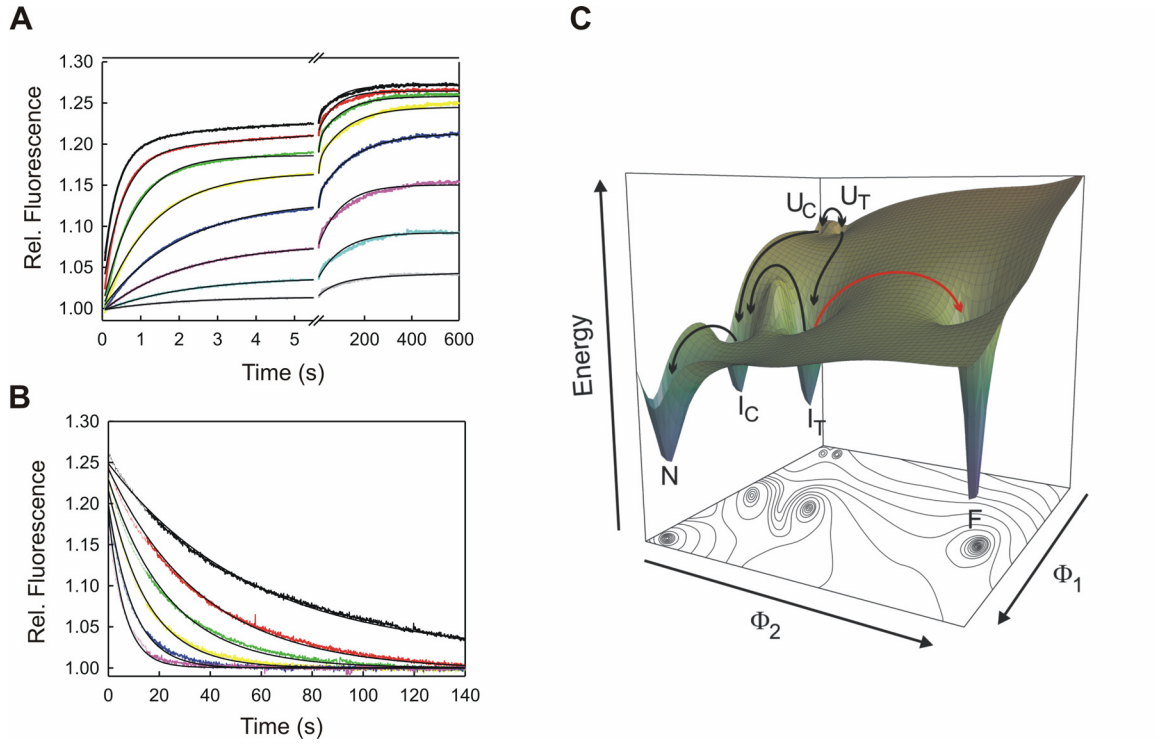


Figure 1: Linking the wild-type β_2m folding and aggregation free energy landscapes. (**A**) Folding and (**B**) unfolding kinetics of β_2m at pH 7.0, 37°C and at different guanidinium chloride concentrations, fitted to a five-state folding mechanism (solid lines). (**C**) Representative free energy surface, indicating the gain of structure folding (black arrows) from the unfolded protein (U) containing *cis* (C) or *trans* (T) Pro32 to the native state (N), as well as the aggregation pathway (red arrows) into amyloid fibrils (F) from the folding intermediate I_T.

β-sandwich proteins from self-association, promote β_2m self-assembly at neutral pH. We are currently using a series of protein variants and NMR spectroscopy to shed further light onto the structural properties of this aggregation precursor species.

Analysing aggregation on a per residue basis

Although the mechanism of β_2m amyloid fibril formation is now one of the best-described *in vitro*, very little knowledge exists about the role of individual amino acids in this process. To address this question we are currently analysing a large set of protein variants carrying point mutations, initially targeting residues in the native β-strand E (Figure 2). Combining the power of systematic mutant analysis with biophysical analysis of the effect of mutation on the assembly kinetics we aim to derive a better understanding of this complex assembly mechanism. Furthermore, the aim of our approach is to gather valuable insights into the structural characteristics of all species involved in the amyloid assembly, including the structure of the final amyloid fibril itself.

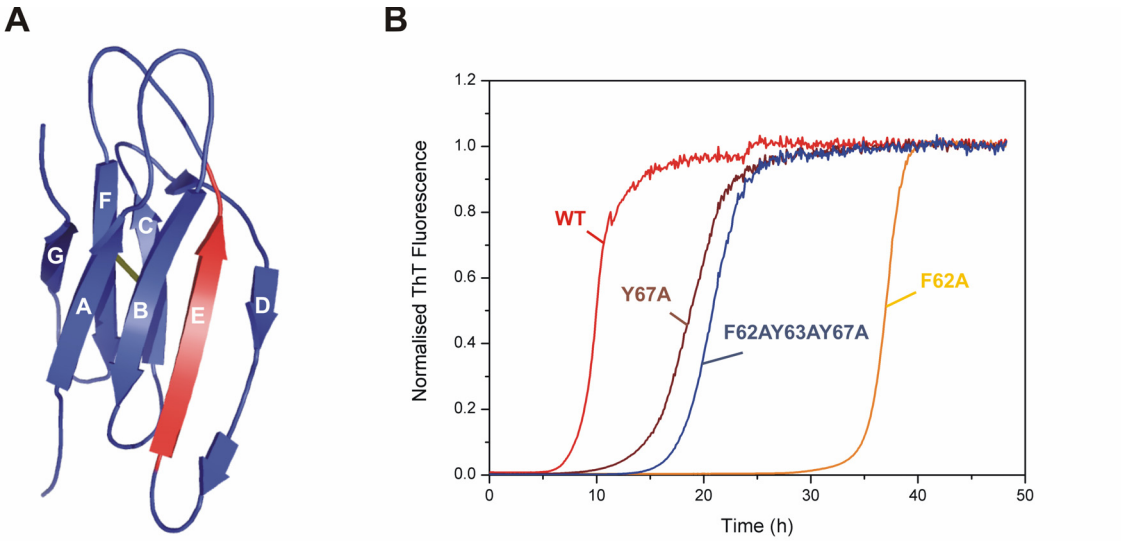


Figure 2: Mutational analysis of β_2m fibril formation. (**A**) Ribbon diagram of native β_2m, highlighting the position in the native structure of the β-strand E (red). (**B**) Mutations in the E-strand alter the rate of amyloid fibril formation at pH 2.5.

Publications

Myers, S.L., Jones, S., Jahn, T.R., Morten, I.J., Tennent, G.A., Hewitt, E.W. & Radford, S.E: (2006) A systematic study of the effect of physiological factors on β_2-microglobulin amyloid formation at neutral pH. *Biochemistry*. **45**, 2311-2321.

Jahn, T.R., Parker, M.J., Homans, S.W. & Radford, S.E: (2006) Amyloid formation under physiological conditions proceeds *via* a native-like folding intermediate. *Nature Struct Mol Biol.* **13**,195-201.

Funding

We are grateful to the University of Leeds, BBSRC and The Wellcome Trust for financial support.

Targeting the functions of viral proteins with RNA aptamers

Clare Nicol, Mark Ellingham, David Bunka, David J. Rowlands,
G.Eric Blair and Nicola J. Stonehouse

Introduction

RNA aptamers are ideal laboratory tools to study the molecular details of viral replication and assembly. They are derived from the process of *in vitro* selection which results in ligands of high affinity and specificity to almost any target. We have selected RNA aptamers to the human papilloma virus 16 (HPV16) E7 protein and 3Dpol from foot-and-mouth disease virus (FMDV).

Foot-and-mouth disease virus (FMDV) is a highly important animal pathogen. 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. The 3Dpol enzyme is a key component of the complex responsible for the replication of the FMDV genome. Some of the aptamers selected against this enzyme (see previous Astbury Report for more information) were found to effectively inhibit function with an IC_{50} of less than 20 nM and K_i values of 18-75 nM. Furthermore, the aptamers were specific to FMDV – having no effect on polio 3Dpol.

HPV16 is a high-risk papilloma virus associated with the development of cervical cancer. Cervical cancer is the second most common cancer in women and there are more than a quarter of a million deaths due to this disease each year, 80% of which are in developing countries. More than 95% of cervical cancers have been shown to contain HPV DNA; 50% of these are HPV16 specific. Carcinogenesis occurs due to the integration of HPV DNA into the host genome, with resulting loss of the negative-feedback control of the oncogenes E6 and E7. E7 has many cellular functions, the most well characterized is the binding and destabilization of the retinoblastoma protein (pRB) which regulates the progression of the cell cycle from the G1 to the S phase by inhibiting major transcription factors.

Aptamers raised to GST-E7 alter cell cycle progression in cervical cancer cells

After 10 rounds of selection, we isolated aptamers with high affinity for a GST-fusion of E7. These were screened in a cell line derived from a human cervical carcinoma expressing E7 (SiHa). Cell cycle analysis showed that 3 aptamers (denoted A2, C3 and I1) reduced progression to the G2-M phase by 50%, while 1 aptamer (denoted J1) increased the proportion of cells in the S-phase by 20%. Further analysis showed that aptamers A2 and C3 significantly increased apoptosis in these cells while the presence of J1 appeared to inhibit apoptosis. Cell growth rates appeared to be slowed by aptamers A2 and C3 whilst stimulated by aptamer J1. It is proposed that instead of blocking the interactions of E7 with cell cycle control proteins, aptamer J1 may have a stabilising or enhancing role. Transfection of aptamers into a control cell line (HaCat) had no effect.

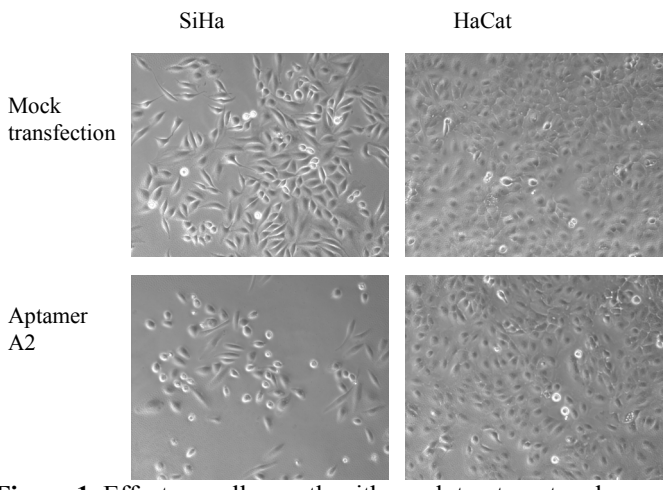


Figure 1 Effect on cell growth with mock treatment and transfection with aptamer A2.

Collaborators
We are grateful to Graham Bottley, University of Leeds for help with analysis of FACS data.

Funding
Funding from Yorkshire Cancer Research (to NJS and GEB) is gratefully acknowledged.

Publications
Ellingham M., Bunka D.H.J., Rowlands D.J. and Stonehouse N.J. (2006) Selection and characterisation of RNA aptamers to the RNA-dependent RNA polymerase from foot and mouth disease virus. *RNA* **12**; 1970-1979.

Investigating the molecular interactions of the ϕ29 DNA packaging motor

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

Introduction

The bacteriophage ϕ29 of *Bacillus subtilis* packages its genomic dsDNA in an ATP-dependent fashion, with components of the virus forming a powerful molecular motor. The viral procapsid is composed of the capsid proteins with an associated connector (portal). Packaging of the viral genomic dsDNA (covalently complexed with gp3) involves the presence of RNA molecules (pRNA) and hydrolysis of ATP by an associated ATPase (gp16). The predicted secondary structure of pRNA is a three-helix junction with complementary bases in two of the loops. The possibility that base-pairing between adjacent pRNA molecules is involved in the tertiary structure of pRNA has been investigated through the mutation of these loop sequences and as a consequence, there have been many reports in the literature on the multimeric nature of pRNA in DNA packaging. The issue of pRNA stoichiometry is a matter of debate and it has been postulated that both pRNA hexamers and pentamers could play a role in DNA packaging. We have taken a novel approach in order to provide information essential for the rational elucidation of the mechanism of DNA packaging, based on determination of the affinity between components and the use of single molecule photobleaching experiments (SMPB).

Towards a model for motor function

The presence of solution state multimers of pRNA, formed by multimerization in the absence of other ϕ29 components, has been analysed by light scattering experiments. Further analysis of pRNA multimerization has been achieved using analytical ultracentrifugation (AUC), leading to data for the affinity of the solution state pRNA : pRNA interaction under a range of magnesium ion concentrations. The binding affinity of pRNA to connector gp10 has been investigated by two independent techniques (see previous Astbury reports).

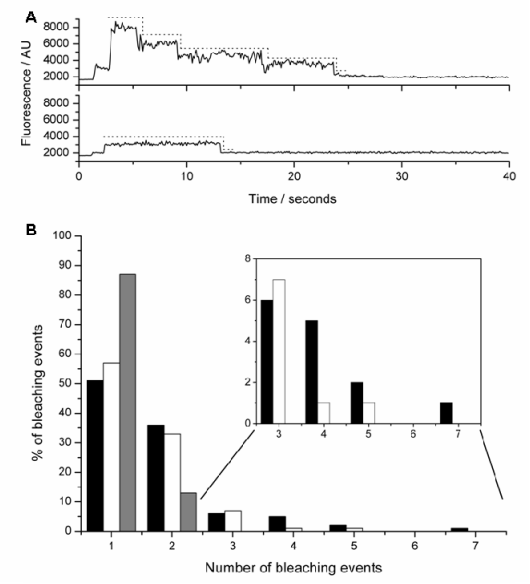


Figure 1 Example traces showing fluorophore photobleaching for labelled multimerisation-competent pRNA with 10mM Mg^{2+} + C-His gp10 connector (A, upper) and mutant pRNA + 10 mM Mg^{2+} (A, lower). The dashed line provides an indication of the steps identified in this particular trace and used in data analysis. Data from SMPB experiments were analysed by determining the number of photobleaching events observed in different experimental assemblies (B). In the case of the mutant pRNA + 10 mM Mg^{2+} (grey), approximately 90% of the photobleaching events represent monomeric species. When the number of photobleaching events for the multimerisation-competent pRNA + 10 mM Mg^{2+} is analyzed (white), there is an increase in the proportion of dimeric species observed and a significant number of pRNA trimers are also present. A stabilization of higher order pRNA species is observed as an increase in the occurrence of four or more photobleaching events (inset show enlarged view of higher order region of histogram) when this RNA + 10 mM Mg^{2+} is incubated with gp10 connector (black).

SMPB experiments, in which the photobleaching of individual fluorophore-labelled pRNA molecules is monitored, have been used to further confirm the ability of pRNA to form higher order assemblies and also probe the effect of the presence of connector on pRNA

multimerization. The role of intermolecular base-pairing of pRNA in the formation of higher order multimeric forms has been probed in all experiments by the comparison of wild type and multimerization-incompetent mutant pRNAs.

Our current studies are focusing on conformational changes in both the connector and pRNA. We believe that unlikely that there is a rotation between the connector and pRNA during DNA packaging and we propose that magnesium-induced conformational changes in both connector and pRNA could drive the packaging event.

Collaborators
Alistair Smith, University of Leeds

Funding
Funding from the BBSRC is gratefully acknowledged.

Publications
Robinson M.A., Wood J.P.A, Capaldi S.A., Baron A.J. Gell, C. Smith, D.A. and Stonehouse, N.J. (2006) Affinity of molecular interactions in the bacteriophage ϕ29 packaging motor. *Nucleic Acids Res*. **34**; 2698-2709.

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. Our work to date has been with the small, mobilisable, staphylococcal plasmid pC221 (4.6 kb), which offers a simple system that embodies the initial events in plasmid mobilisation.

We have recently focussed our attention on another small staphylococcal plasmid, pE194. This is a 3.7 kb, erythromycin resistance plasmid reported to be non-mobilisable, despite the *in vivo* identification of an RS_A/*oriT* nick site with significant identity to that of a mobilisable streptococcal plasmid, pMV158. Similarly, the formation of a *recA* independent cointegrate between pE194 and the related plasmid pT181 has been demonstrated at this site, and shown to require a plasmid encoded protein termed Pre (Plasmid recombination).

Initial observations in our laboratory have demonstrated the mobilisation of pE194-based plasmids, in the presence of the co-resident conjugative plasmid, pGO1; such mobilisation depends on the presence of an intact *pre* gene.

Recent findings

The pE194 Pre protein has been grouped within the Pre/Mob family of relaxases, which also include Pre of pT181 and MobM of pMV158, the latter being shown to mediate the nicking and subsequent mobilisation of pMV158 in the presence of a co-resident conjugative plasmid pIP501. This has prompted us to examine the potential relaxase activity of Pre.

As a nicked open-circular plasmid complex between Pre and pE194 has not been detectable *in vivo* using a traditional whole cell lysate methodology, we have cloned pE194 *pre* and purified it as an N-terminal hexahistidine-tagged fusion protein. The purified protein shows nicking activity against the RS_A/*oriT* of both pE194 and also that of plasmid pT181, confirming it has sequence-specific relaxase activity (Fig. 1).

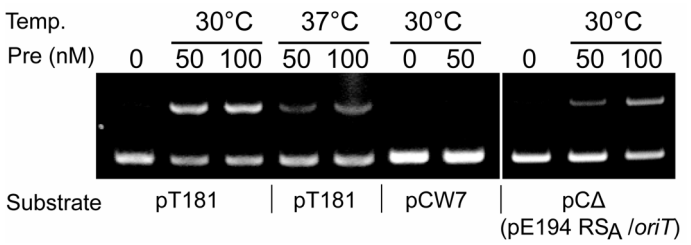


Fig. 1. Agarose gel showing nicking of supercoiled plasmid DNA substrates by purified pE194 Pre protein. The control plasmid, pCW7, contains an interrupted RS_A/*oriT* sequence.

This confirms the role of Pre as the relaxase likely to be involved in pE194 mobilisation. To dissect the active site of Pre, we have constructed a mutant pE194 plasmid that encodes a Y45F mutation within Pre. This is being assayed in mobilisation experiments to confirm that Y45 is the catalytic tyrosine, and concomitantly, the functionality of Pre that is required for mobilisation.

Publications

Caryl, J.A. & Thomas, C.D. (2006) Investigating the basis of substrate recognition in the pC221 relaxosome. *Mol. Microbiol*. **60**, 1302-1318.

Acknowledgements & Funding

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

X-ray crystallographic analysis of GehD lipase – a potential virulence factor *Staphylococcus epidermidis* infection

Stephen B. Carr, John D. Wright, Keith T. Holland and Simon E. V. Phillips

Background

Staphylococcus epidermidis is common resident of normal human skin especially in sebum rich areas. Nascent sebum is rich in triacylglycerides, which are hydrolysed by microbial lipases liberating free fatty acids onto the cutaneous surface. It has been suggested that the lipase may be important for the colonisation and persistence of the microflora on skin by liberating glycerol as a nutrient source. GehD has also been shown to bind collagen and may also play a role in the adherence of bacterial cells to the skin surface.

Staphylococcus epidermidis has assumed greater medical importance in recent years, as opportunistic pathogens of compromised hosts, causing infections associated with the insertion of medical devices such as prosthetic heart valves. GehD has been shown to be present in higher levels in infected patients than in healthy individuals suggesting it is an important factor in the colonisation and persistence of *S. epidermidis* at infection sites.

Recent findings

X-ray diffraction data were collected on crystals of GehD lipase at the SRS (Daresbury, UK) to a maximum resolution of 1.55 Å. The crystals belong to space-group $P2_12_12$ and contain one molecule per asymmetric unit. Molecular replacement using a fragment of lipase P1 from *B. stearothermophilus* resulted in interpretable electron density maps. The final model of GehD displays excellent stereochemistry and has R and R_{free} values of 16.1 and 20.1 % respectively.

GehD displays an αβ-hydrolase fold similar to other bacterial lipases, however, GehD has crystallised with the "lid" open exposing a hydrophobic cleft suitable for binding lipids. The catalytic triad (Ser 113, His 349 and Asp 307) sit at the base of this cleft (Figure 1). The protein also chelates calcium and zinc ions, the former has been shown to enhance catalytic activity, while the latter has been suggested to have a purely structural role. The structure of GehD also shows homology to the collagen binding domain of integrin α1β2 suggesting it may bind collagen in a similar manner.

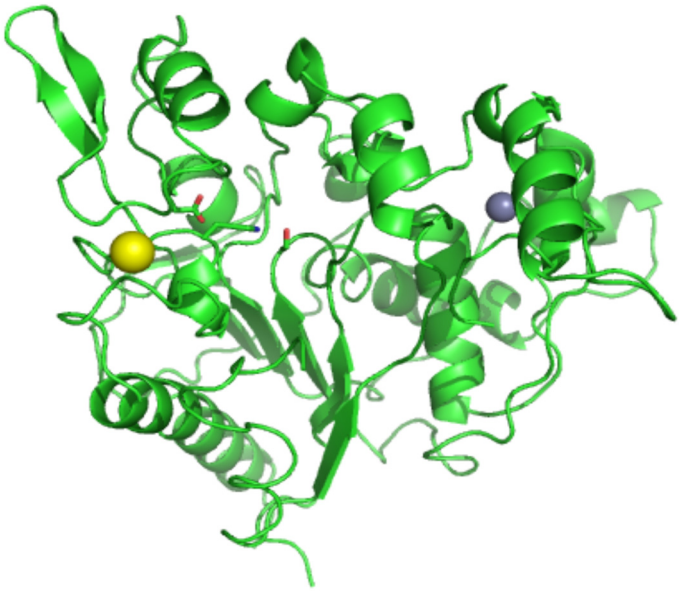


Figure 1 – GehD lipase displaying active site residues and bound calcium (yellow) and zinc (blue) ions

Acknowledgements

This work was funded by the Wellcome Trust

Structure determination of a 59 kDa fragment of the DNA-cleavage domain of topoisomerase IV from *Staphylococcus aureus*

Stephen B. Carr, Simon E. V. Phillips and Christopher D. Thomas

Background

DNA is prone to topological problems due to its double helical structure and the intertwining of its complementary strands. For example, during transcription or replication the unwinding of the chromosome forces adjacent regions of DNA to overwind and underwind, generating a superhelical strain. The removal of this intermolecular strain is vital for the prevention of DNA tangling and knotting, which produces structures that can lead to the formation of double-stranded breaks in the DNA and chromosome instability; furthermore, it is essential that circular daughter chromosomes of bacteria are disentangled at cell division. The resolution of such structures is performed by topoisomerase IV (GrlA), a type IIA DNA topoisomerase, which can remove these regions of unusual topology by passing one double-helical segment of DNA through another.

Staphylococcal topoisomerase IV is composed of two subunits GrlB and GrlA. Previously, a 56 kDa fragment of GrlA spanning the DNA cleavage domains (residues 25-491) had been crystallised and diffraction data collected. However, the refinement of the resulting structural model proved extremely difficult due to the presence of translational non-crystallographic symmetry (NCS) within the unit cell of the crystal. This led to the purification and crystallisation of an alternative 59 kDa construct (GrlA59, residues 1-491), and the collection of diffraction data to a resolution of 2.8 Å.

Recent findings

Crystals of GrlA59 belong to space-group $P2_12_12$ and contain one molecule per asymmetric unit, removing the problem of translational NCS. Molecular replacement using the 56 kDa fragment of GrlA resulted in interpretable maps, allowing the majority of the polypeptide chain to be traced (no electron density was observed for residues 1-29, but their presence was required for growth of well ordered crystals).

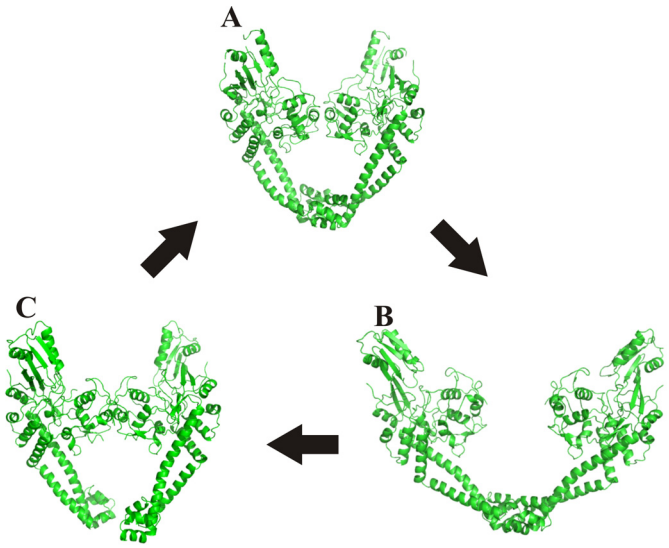


Figure 1 – Conformational changes within topoisomerase IV during DNA cleavage: A – Closed DNA-binding state. B – G-gate open, conformation adopted after DNA binding and cleavage to allow passage of one DNA strand through another. C – C-gate open, adopted after strand passage to allow relegation of cleaved strand and release of transported

The overall fold of GrlA59 is similar to that seen in other type II topoisomerase enzymes with a DNA binding groove at the "top" of the enzyme the active site at the central interface between the two monomers and the major dimerisation interface at the base of the protein. Comparison of the structure of GrlA59 with that of ParE (homologous subunit from topoisomerase IV of *E. coli*) has allowed the visualisation of the different conformational states accessed by the protein during its catalytic cycle (Figure 1).

Publications

Carr, S.B. Makris, G. Phillips, S.E.V. & Thomas, C.D. (2006) Crystallization and preliminary X-ray diffraction analysis of two N-terminal fragments of the DNA-cleavage domain of topoisomerase IV from *Staphylococcus aureus Acta Cryst.* ***F62*** 1164-1167.

Collaborators

George Makris, Omega Mediation Hellas Ltd, Clinical and Pharma Consulting, 11525 N. Psychiko, Athens, Greece.

Acknowledgements

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

Convergent transcription studied at the single molecule level by AFM

Neal Crampton, Jennifer Kirkham, Bill Bonass and Neil Thomson

We are using atomic force microscopy (AFM) to study what happens at the molecular level when two genes are simultaneously expressed. It is becoming increasingly evident that there are gene structures within the genomes of a wide variety of organisms, from viruses through to humans, where genes are "nested" one within another. A nested gene is defined as a gene that lies completely within the sequence of another gene and frequently in the opposite orientation. This raises the question: can these genes can be simultaneously transcribed from a single genetic locus? 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 as yet unclear. One example of a nested gene is associated with the development of tooth enamel. *AMEL*, the gene for amelogenin (the principle protein component of developing enamel) lies wholly within intron 1 of the *ARHGAP6* gene, which encodes RhoA, a monomeric G protein. *AMEL* is only transcribed for a specific time interval whereas *ARHGAP6* is widely transcribed throughout enamel development, raising questions regarding their reciprocal control via their nested gene structure.

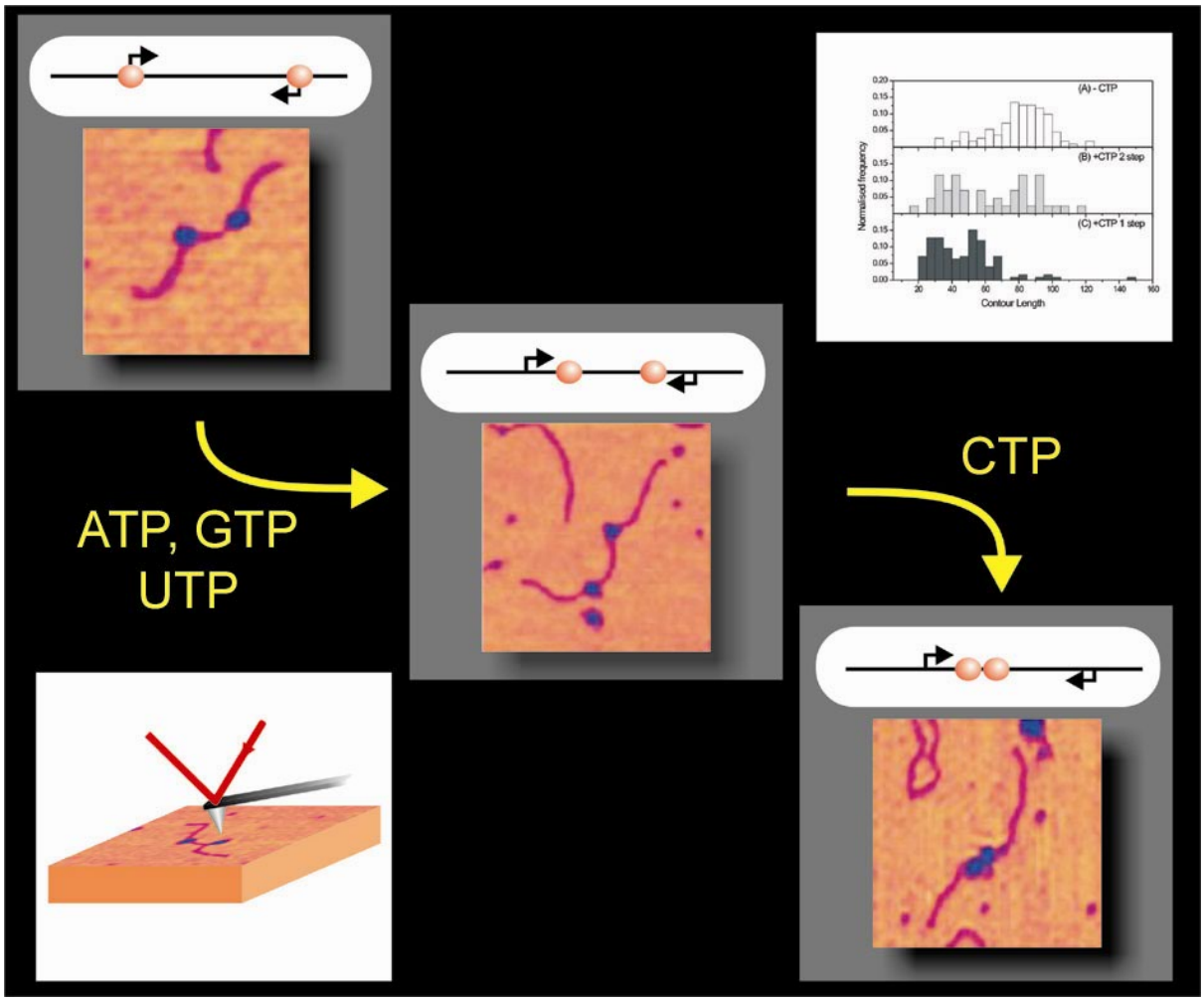


Fig. 1 – Summary of AFM data acquired in *ex-situ* imaging of air dried samples on mica using tapping-mode. The bottom left panel illustrates the principle of operation of the AFM. The images are single complexes of linear DNA (1074bp) with two *E. coli* RNA polymerases (RNAP) bound. They are in sequence: before addition of NTPs (ribonucleoside triphosphates) with the RNAP at the promoters; after addition of 3 NTPs when the RNAPs are stalled at specific positions along the template; after addition of the 4th NTP, CTP which allows the RNAPs to continue to transcribe and collide together. The panel in the top right shows histograms of the spacing between the RNAP before and after addition of NTPs.

For two genes that are lying in opposite orientations to be simultaneously expressed at a single locus, the two RNA polymerases (RNAPs) transcribing the DNA would have to be able to pass each other without losing their transcripts. While this scenario may seem unlikely from current structural knowledge of RNAPs and the current understanding of the mechanics of transcription, there is no existing experimental data at the single molecule level to judge the outcomes of head-on "collisions" between RNAPs. When the converging RNAPs encounter each other on a single DNA template there are a number of possibilities that could occur: 1) Both RNAPs stall against each other preventing expression of both genes; 2) One RNAP preferentially displaces the other enabling one gene to "silence" the other; 3) Both RNAPs are displaced during collision, again preventing expression of both genes; 4) The RNAPs pass each other, by some as yet unknown mechanism, allowing simultaneous expression of both genes.

In our studies, a model linear DNA template with two convergently-aligned λ_{pr} promoters 338bp apart has been used to observe what happens when two *E. coli* RNAPs "collide". Imaging of the complexes *ex-situ* before and after collision indicates that for a significant proportion of complexes the RNAPs stall against each other. Interestingly, they stall at distances larger than that expected for close contact (> 20 nm separation) indicating that they may sense the presence of each other either through long range forces or through stress-mediated communication via the DNA linkage between them. Accurate measurements of the positions of the RNAPs on the DNA template show that one of the RNAPs "collides" with the second one before it has cleared the promoter or stall site at which it is paused. The active RNAP appears to cause the second RNAP to back-track by a distance of 50 to 60 bp.

These data may imply that RNAPs do not have the ability to pass each other on a single template and therefore that genes cannot be simultaneously transcribed. This question, however, cannot be completely resolved until an experiment that directly observes the collision process close-to real-time can be performed. Conventional AFM technology is currently too slow to reliably capture the collision event, but a number of groups worldwide are working on fast scan technologies that allow image acquisition times in the millisecond rather than seconds time regime. Application of these technologies to biomolecular systems in aqueous environments, while not straightforward, should allow the collision process to be studied in more detail.

Collaborators

Carolyn W. Gibson (University of Pennsylvania, USA)
Claudio Rivetti (University of Parma, Italy)

Publications

Crampton N., Bonass W.A., Kirkham J., Rivetti C. & Thomson N.H. (2006)
Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Research* **34**, 5416-5425.

Crampton N., Bonass W.A., Kirkham J., Gibson, C.W. & Thomson N.H. (2006)
Imaging RNA Polymerase-Amelogenin Gene Complexes with Single Molecule Resolution using Atomic Force Microscopy. *European Journal of Oral Science,* **114** (Suppl. 1) 133-138.

Funding

We gratefully acknowledge the EPSRC and University of Leeds for financial support. Neil Thomson is a Leeds University Advanced Fellow.

Titin

Larissa Tskhovrebova, Tanniemola Liverpool, Andy Baron, Ahmed Houmeida,
Nasir Khan and John Trinick

Titin consists of the largest known polypeptide (3.7 MDa) and is the third most abundant protein of muscle. A large part of the molecule forms elastic connections between the end of thick filaments and the Z-line. These connections are the main source of elasticity of relaxed muscle. This elasticity also ensures that the myosin filaments stay in the middle of the sarcomere, which ensures that even forces are developed by opposing halves of the filament during active contraction.

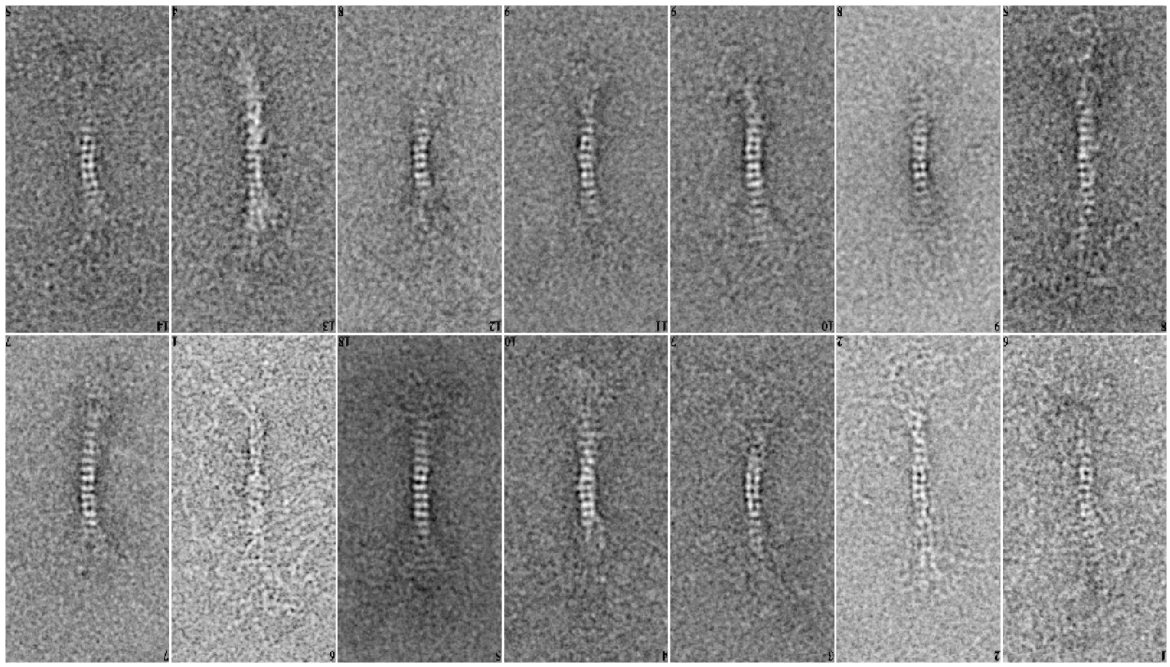


Fig. 1 Titin end filaments

The mechanism of titin elasticity has been pursued in many single molecule studies (including some by ourselves) using optical tweezers and AFM. Such studies have also led to claims that the single molecule data can be extrapolated directly to explain the passive elasticity of muscle. However, we have shown that for ~100 nm beyond the end of the myosin filament, the titin molecules form hexameric aggregates called end-filaments (Fig. 1). We have recently isolated a large proteolytic fragment from the appropriate region of the molecule, which, in reduced salt concentrations, forms structures very similar to end-filaments. The molecular mass of this fragment was measured as 290 kDa by analytical ultracentrifugation. Titin molecules in this region therefore do not act independently, as in the single molecule experiments, and the elastic behaviour must be more complex. We are also exploring the effects on titin of molecular crowding and confinement within the lattice of actin filaments, which are also likely to modulate the behaviour of the molecule.

Funding
British Heart Foundation

Electron microscopy of intact tissues and cells

Kasim Sader and John Trinick

There is an urgent need across biology and in medicine to obtain more detailed images of intact cells and tissues. An important goal is to reach ~2 nm resolution where protein molecule shapes can be recognised. Electron microscopy of thin sections cut from material embedded in plastic remains by far the most widely used method for such *in situ* studies. However, details smaller than ~ 50Å are generally not preserved by this method. This precludes easy integration of higher resolution structural information from component macromolecules and complexes into the cellular context, which is a major limitation to understanding cellular function. There have been few studies exploring how detail is lost during embedding and sectioning; however, it is often assumed that fixation and dehydration are the most damaging steps. Our collaborator, Michael Reedy, has monitored the preservation of insect flight muscle through fixation, dehydration and embedding by X-ray fibre diffraction. Remarkably, the best embedded blocks diffracted to 1.2 nm resolution (see Figure). However, Fourier transforms of sections cut from these blocks extended to only ~5 nm. This suggests that the cutting is the stage where much damage can occur. A major artefact of sectioning is compression in the direction of the cut, often by ~30%. Thus, if compression could be eliminated, preservation might be greatly improved. An oscillating knife ultramicrotome eliminates the compression, but the effect of this on resolution had not been measured. We cut sections, with and without, the knife oscillating, from the embedded insect flight muscle that give X-ray transforms extending to 1.2 nm. Sections cut with the knife oscillating did not show improved preservation over those cut without. Thus compression during cutting does not appear to be the major source of damage, which leaves unexplained the 50Å versus 13Å discrepancy between block and section preservation. The results nevertheless suggest that improvements in ultramicrotomy will be important for bringing thin-sectioning of plastic-embedded cells and tissues to the point where macromolecule shapes can be resolved.

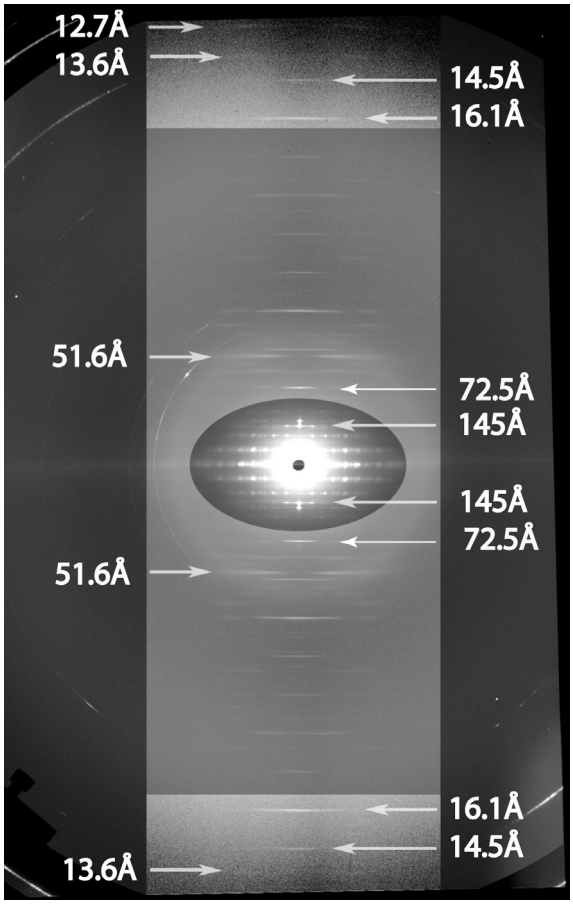


Figure 1. X-ray diffraction data from an embedded block

Collaborator

Michael Reedy (Duke University) and Daniel Studer (Berne University)

Publications

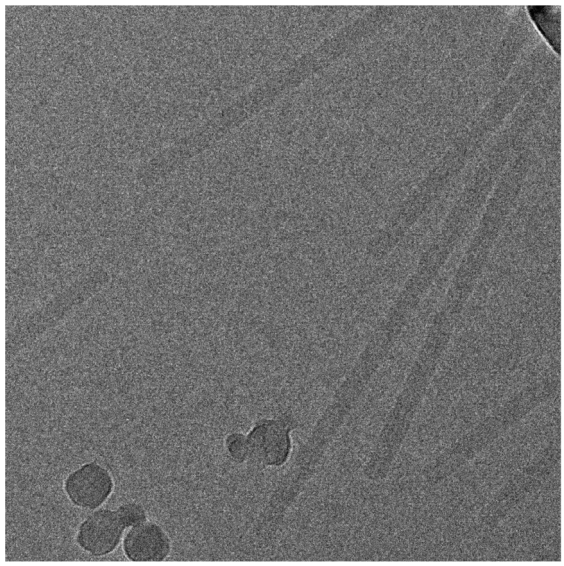
K. Sader, M. Reedy, D. Popp, C. Lucaveche and J. Trinick (2007) Measuring the resolution of uncompressed plastic sections cut using an oscillating knife ultramicrotome.. *J Struct. Biol*. Published online. Feb. 2007.

Funding: Wellcome Trust

Automated electron microscopy with Leginon

William Nicholson and John Trinick

Most proteins do not function as separate molecules, but as larger complexes and assemblies, often in what are called macromolecular machines. Electron cryo-microscopy of unstained frozen-hydrated specimens is an excellent way to study the structure and mechanisms of such complexes. Sub-1 nm resolution is frequently possible, which resolves internal helices, and the method has good time-resolution (~5 msec). Radiation damage is the main factor limiting information recovery. The method is therefore dependent on the use of single particle image processing to integrate the image data from many weakly irradiated objects. Frequently this also leads to 3D models. Generally, 10^4 - 10^5 particle images are needed and acquisition of these is laborious, requiring considerable skill and time. This aspect of electron microscopy is therefore ripe for the introduction of high-throughput techniques analogous to those that have greatly accelerated other aspects of structural biology, such as crystallography. To this end, colleagues at the Scripps Institute (La Jolla, CA) have developed Leginon software for completely automated image acquisition, including recognition of promising specimen areas, focussing and low-dose image recording.



The figure shows a micrograph of frozen-hydrated tobacco mosaic virus recorded with under automatic control

Using Leginon, they recently recorded 1100 micrographs in a 22 hour session, from which 280,000 particle images were windowed. We are using Leginon to control our FEI Tecnai F20 field emission gun (feg) microscope, which has a coherent electron beam for phase contrast imaging of unstained, frozen specimens. Leginon runs on an external laptop computer and the goal is to run the microscope unattended for extended periods, eg overnight. .

Funding
BBSRC and NIH (USA)

Bioinformatics applied to protein structure and function analysis, and systems biology

James Bradford, Matthew Care, Andrew Garrow, Binbin Liu, Sally Mardikian, Archana Sharma-Oates, Philip Tedder, Elizabeth Webb and David Westhead

Bayesian networks

Much of our work involves applying machine learning methods to solve biological problems. One such method is a Bayesian network. Bayesian networks are probabilistic graphical models which provide compact representations for expressing joint probability distributions and for inference. This representation and use of probability theory makes Bayesian networks suitable for learning from incomplete datasets, expressing causal relationships, combining domain knowledge and data, and avoid over-fitting a model to data. Consequently, a host of applications in computational biology have used Bayesian networks and Bayesian learning methodologies.

Protein-protein interface prediction

Identifying the interface between two interacting proteins provides important clues to the function of a protein, and is becoming increasing relevant to drug discovery. In this work, surface patch analysis was combined with a Bayesian network to predict protein-protein binding sites with a success rate of 82% on a benchmark dataset of 180 proteins, improving by 6% on previous work that used a support vector machine, and well above the 36% that would be achieved by a random method. A comparable success rate was achieved even when evolutionary information was missing, a further improvement on our previous method which was unable to handle incomplete data automatically. We demonstrated the applicability of our method to drug discovery efforts by successfully locating a number of binding sites involved in the protein-protein interaction network of papilloma virus infection (Figure 1).

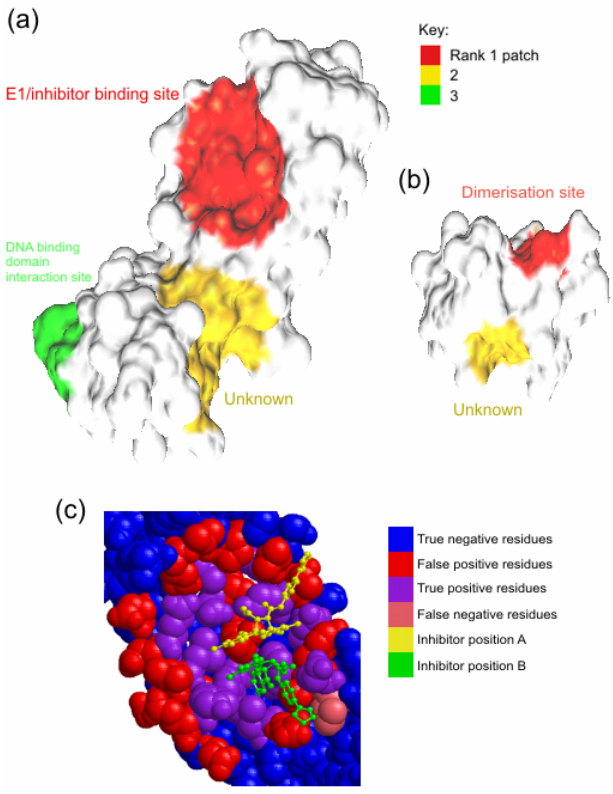


Figure 1: Predicted patches on papilloma virus E2 protein. (a) Trans-activation domain (PDB code: 1r6k), (b) DNA binding domain (PDB code: 1jj4), (c) detail of the inhibitor peptide inandione interaction site.

Predicting the effect of missense mutations on protein function

We have also used Bayesian networks to predict the functional consequences of missense mutations on proteins. Exploiting the ability of the Bayesian network to handle missing data automatically, we found that structural information is significantly more discriminatory than evolutionary information in this classification task and on the dataset used. Indeed, the top three strongest connections with the class node in the network all involved structural nodes. We therefore derived a simplified Bayesian network that used just these three structural descriptors, with comparable performance to that of an all node network.

In related work, we carried out a comprehensive analysis of datasets used in single nucleotide polymorphism (SNP) prediction methods. Such methods require data sets of mutations classified as 'deleterious' or 'neutral' for training and/or validation. While different workers have used different data sets there has been no study of which is best. Here, the three most commonly used data sets were analysed. We examined their contents and related this to classifiers, with the aims of revealing the strengths and pitfalls of each data set, and recommending a best approach for future studies. The data sets examined were shown to be substantially different in content, particularly with regard to amino acid substitutions, reflecting the different ways in which they are derived. This could account for some of the differences seen between classifiers and reveals some serious pitfalls in some data sets, making them less than ideal for nsSNP prediction.

Arabidopsis Co-expression Tool (ACT)

The Arabidopsis Co-expression Tool, ACT, ranks genes across a large microarray dataset according to how closely their expression follows the expression of a query gene. A database stores pre-calculated co-expression results for approximately 21,800 genes based on data from over 300 arrays. These results can be corroborated by calculation of co-expression results for user-defined sub-sets of arrays or experiments from the NASC/GARNet array dataset. A wealth of tools is available on the ACT website (http://www.arabidopsis.leeds.ac.uk/act/) to analyse the various outputs from the tool.

Reconstruction and analysis of biochemical networks

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. 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 or the factors controlling them are likely to be very effective. In particular, our group has focused on *Plasmodium falciparum*, the parasite causing the most virulent type of malaria.

With these goals in mind, we have developed a comprehensive suite of programs for the representation and analysis of metabolic networks. The metabolic SearcH And Reconstruction Kit (metaSHARK) includes an object-oriented database to store knowledge about networks of chemicals and reactions, as well as an automated system to search an un-annotated 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 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 metabolic enzyme is expressed under a particular condition or at a particular time point in an organism's life cycle. Pathways in the network can then be ranked according to their biological relevance based on the combined expression levels of each gene in the pathway. This data has recently been incorporated into metaSHARK so that gene expression can be visualised on top of metabolic network diagrams, thus allowing the expression levels of multiple genes to be viewed simultaneously. Statistical analysis can be performed to identify particular pathways which are significantly perturbed in a particular experiment.

Searching genomes for trans-membrane barrel proteins

The ability to search sequence datasets for membrane spanning proteins is an important requirement for genome annotation. However, the development of algorithms to identify novel types of transmembrane beta-barrel (TMB) protein has proven substantially harder than for transmembrane helical proteins, owing to a shorter TM domain in which only alternate residues are hydrophobic. Although recent reports have described important improvements in the development of such algorithms, there is still concern over their ability to confidently screen genomes. We have recently developed a new algorithm combining composition and Hidden Markov Model (HMM) topology (Figure 2) based classifiers that achieves a cross-validation accuracy greater than 95%, with 96.7% precision and 94.2% recall. Interestingly, TMB/extracellular protein discrimination is significantly more difficult than TMB/cytoplasmic protein discrimination, with the predictor correctly rejecting just 74% of extracellular proteins, in comparison to 98% of cytoplasmic proteins. Tools and datasets are available through a website called TMB-Web (http://www.bioinformatics.leeds.ac.uk/TMB-Web/TMB-Hunt2).

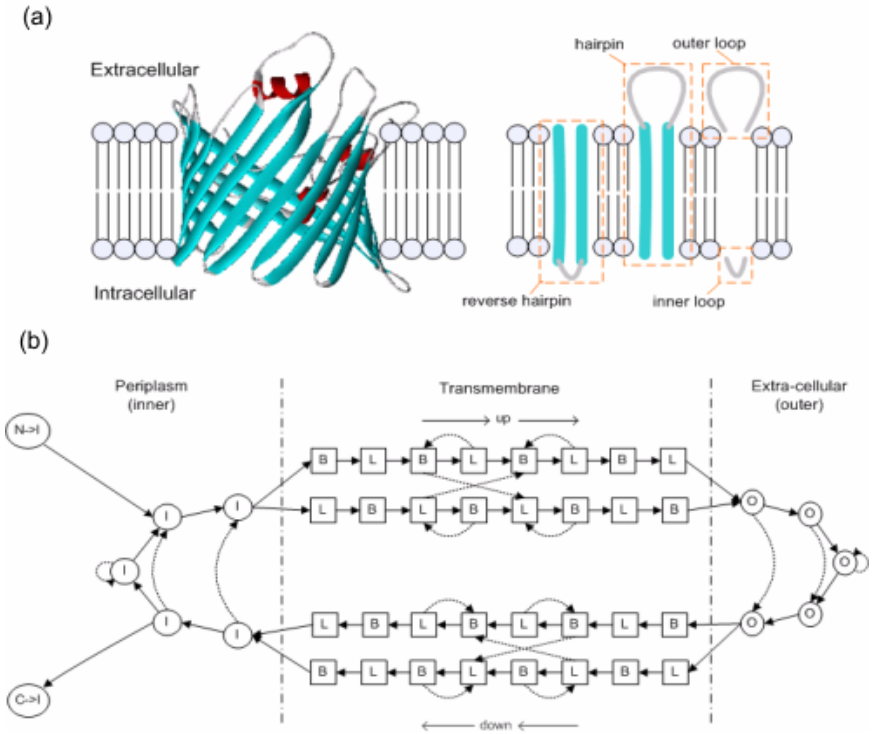


Figure 2: (a) Some of the topological features used in TMB-Hunt2 classification and (b) TMB-HMM architecture, showing different state types and possible transitions.

Collaborators

Professor Mark Boyett (University of Manchester)
Dr.Val Gillet (University of Sheffield)
Dr Richard Jackson (Institute of Molecular and Cellular Biology, University of Leeds)
Drs Chris Needham and Andy Bulpitt (School of Computing, University of Leeds)
Dr Glenn McConkey (Institute of Molecular and Cellular Biology, University of Leeds)

Publications

Bradford, J.R., Needham, C.J., Bulpitt, A.J. & Westhead, D.R. (2006) Insights into protein-protein interfaces using a Bayesian network prediction method. *J. Mol. Biol*. **362**, 365-386.

Care, M.A., Needham, C.J., Bulpitt, A.J. & Westhead, D.R. (2007) Deleterious SNP prediction: Be mindful of your training data! *Bioinformatics*. (in press)

Garrow, A.G. & Westhead, D.R. (2007) A consensus algorithm to screen genomes for novel families of transmembrane beta barrel proteins. *Proteins: Structure, Function and Bioinformatics*. (in press)

Hyland, C., Pinney, J.W., McConkey, G.A. & Westhead, D.R. (2006) metaSHARK: a WWW platform for interactive exploration of metabolic networks. *Nucleic Acids Res.* **34**, W725-W728.

Jen, C.-H., Manfield, I.W., Michalopoulos, I., Pinney, J.W., Willats, W.G.T., Gilmartin, P.M. & Westhead, D.R. (2006) The Arabidopsis co-expression tool (ACT): a WWW-based tool and database for microarray-based gene expression analysis. *Plant Journal*. **46**: 336-348.

Manfield, I.W., Jen, C.-H., Pinney, J.W., Michalopoulos, I., Bradford, J.R., Gilmartin, P.M. & Westhead, D.R. (2006) Arabidopsis Co-expression Tool (ACT): Web server tools for microarray-based gene expression analysis. *Nucleic Acids Res.* **34**, W504-W509.

Needham, C.J., Bradford, J.R., Bulpitt, A.J. & Westhead, D.R. (2006) Inference in Bayesian networks. *Nature Biotechnology*. **24**, 51-53.

Needham, C.J., Bradford, J.R., Bulpitt, A.J., Care, M.A. & Westhead, D.R. (2006) Predicting the effect of missense mutations on protein function: analysis with Bayesian networks. *BMC Bioinformatics*. **7**, 405.

Funding

BBSRC (James Bradford), Yorkshire Cancer Research (Archana Sharma-Oates), University of Manchester (Binbin Liu) and MRC (Sally Mardikian and Andrew Garrow)

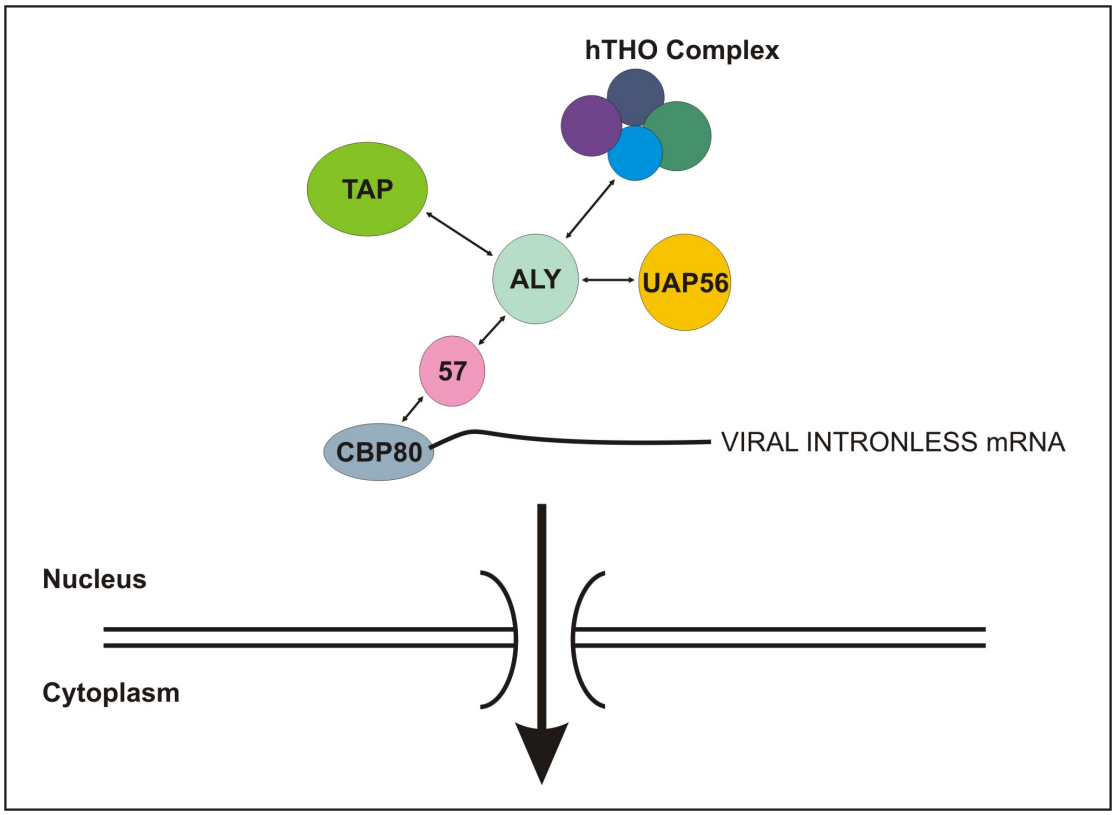
Identification of the ribonucleoprotein complex required for efficient export of herpesvirus intronless mRNAs

James Boyne, Kevin Colgan and Adrian Whitehouse

Introduction

The nuclear export of mRNA composes one part of a larger network of molecular events that begin with transcription of the mRNA in the nucleus and end with its translation and degradation in the cytoplasm. During trafficking to the cytoplasm, a nascent mRNA undergoes numerous co-transcriptional processing steps, including 5'-capping, splicing to remove introns and 3'-polyadenylation. Of these events, it has become clear that splicing is particularly important for mRNA nuclear export, as recruitment of multiprotein complexes required for mRNA export are bound to mRNA in a splicing dependent manner. Two multiple protein complexes, namely, hTREX and the EJC bind at separate locations on spliced mRNA. hTREX, which comprises Aly, UAP56 and the multiprotein ThoI complex, is recruited exclusively to the 5'-end of the first exon, providing 5'-polarity and therefore directionality observed in mRNA export.

However, in contrast to the majority of mammalian genes, analysis of herpesvirus genomes has highlighted that most lytically expressed viral genes lack introns. Herpesviruses replicate in the nucleus of the host mammalian cell, and therefore require their intronless mRNAs to be exported out of the nucleus to allow viral mRNA translation in the cytoplasm. This therefore leads to an intriguing question concerning the mechanism by which the viral intronless mRNAs are exported out of the nucleus in the absence of splicing. To circumvent this problem, and to facilitate viral mRNA export, γ-2 herpes viruses encode the ORF 57 protein. ORF 57 interacts with Aly, binds viral RNA, shuttles between the nucleus and the cytoplasm and promotes the nuclear export of viral mRNA.



Assembly of the ribonucleoprotein complex required for efficient export of herpesvirus intronless mRNA

We are currently investigating how an intronless viral mRNP is assembled in KSHV and what role ORF57 plays in that process. We have shown that ORF57 interacts with hTREX and is essential for the recruitment of hTREX onto intronless viral mRNA transcripts. Importantly, ORF57 does not recruit the EJC to intronless viral transcripts. Moreover, we are currently determining how ORF 57 recognises the viral mRNA and allows recruitment of hTREX. This is the first system that has distinguished between hTREX and EJC *in vivo* and demonstrates that recruitment of hTREX alone to mRNA transcripts is sufficient for their nuclear export. Therefore, we believe this viral system is an exciting model to further study mRNA export mechanisms. We propose a model for herpesvirus mRNA export, whereby ORF57 mimics splicing in order to recruit the mRNA export machinery to intronless viral mRNAs.

Publications

Williams, B., Boyne, J.R., Goodwin, D.J., Roaden, L.R., Wilson, S.A. & Whitehouse, A. (2005). The prototype gamma-2 herpesvirus nucleocytoplasmic shuttle protein, ORF 57, transports viral RNA via the cellular mRNA export pathway. *Biochemical Journal,* **387**, 295-308.

Boyne, J.R. & Whitehouse, A. (2006). Gamma-2 herpesvirus post-transcriptional gene regulation. *Clinical Microbiology and Infection,* **12**, 110-117.

Boyne, J.R. & Whitehouse, A. (2006). Nucleolar trafficking is essential for nuclear export of intronless herpesvirus mRNAs. *Proc. Natl. Acad. Sciences, USA,* **103**, 15190-15195.

Acknowledgements

This project is funded by the BBSRC and we wish to thank Professor Reed, Harvard University for providing hTREX reagents.

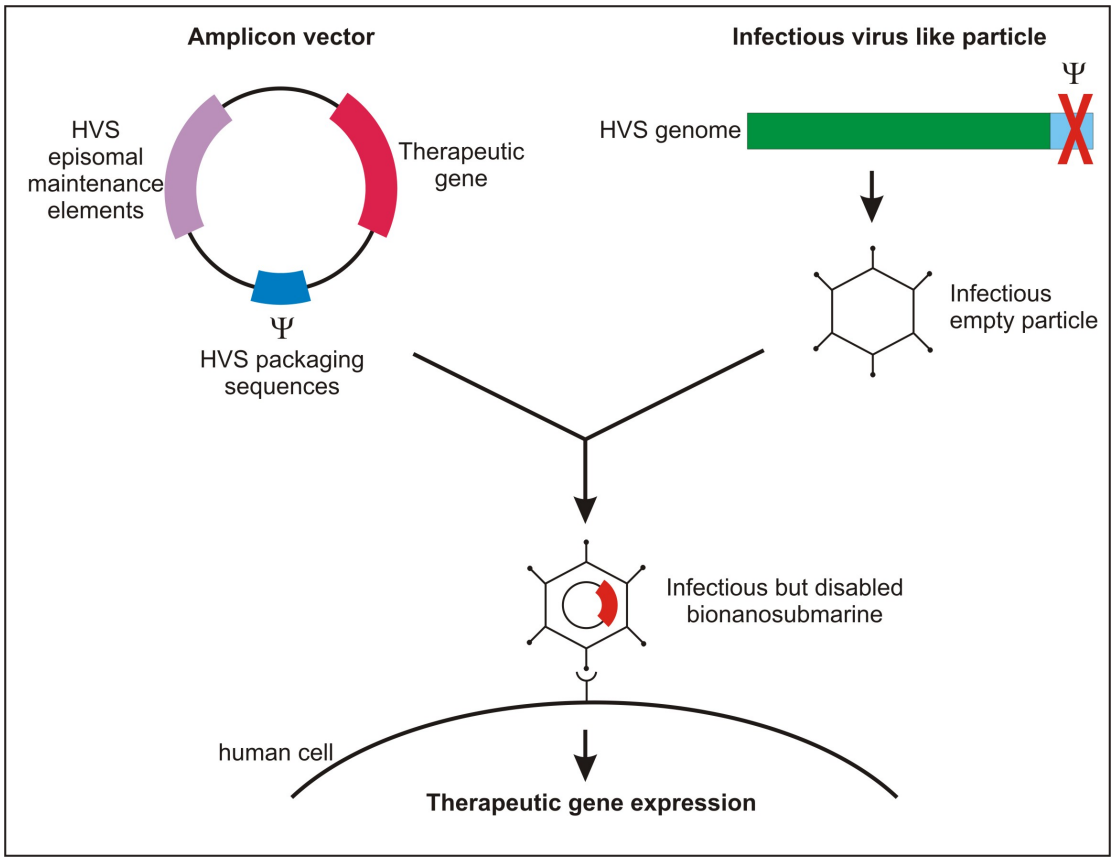
Development of a herpesvirus-based bionanosubmarine

Stuart McNab, Julian Hiscox and Adrian Whitehouse

Introduction

The continuing advances and understanding in the field of molecular and cellular biology, and the underpinning genetic control mechanisms, has allowed the creation and development of non-traditional means for treating disease to be pursued such as gene therapy and bionanomedicine. We are currently developing a bionanosubmarine which delivers therapeutic genes to the correct cell types. The bionanosubmarine is based on elements of *Herpesvirus saimiri* (HVS). We have previously shown that HVS can infect a variety of human cell types and upon infection the viral genome can persist as a non-integrated episome in both *in vitro and in vivo* studies. Therefore, HVS has great potential to be developed as a bionanosubmarine. To achieve this, the infectious delivery system of HVS needs to be retained, but the viral genes need to be deleted and replaced with therapeutic genes.

The bionanosubmarine comprises a two-tier system created in tissue culture. It utilises the biosafety of an amplicon vector plasmid coupled with the natural infectivity of the wild-type virus. The system creates a virus like particle (VLP) containing the transgene of interest expressed from an amplicon vector. The VLP particle is essentially the wild-type viral coat preferably lacking, or with minimal immuno-stimulatory antigens exposed. The VLP is generated by a helper-virus genome which contains all the necessary structural genes while remaining replication deficient.



Methodology to produce a HVS-based bionanosubmarine

An amplicon is a gutless vector derived from a viral genome. The amplicon contains a transgene, the related expression sequences, and the *cis*-acting sequences required for replication, cleavage, and packaging into the VLP. The amplicon carries no transacting virus genes and consequently does not induce synthesis of virus proteins. Therefore, these vectors

are non-toxic for the infected cells and non-pathogenic for the inoculated organisms. Another major advantage of the amplicon system is the removal of most of the virus genome, this consequently creates a transgene capacity equivalent in size to the wild-type virus genome.

To date, we have produced a bionanosubmarine, namely the VLP containing the amplicon. We have demonstrated that the bionanosubmarine is exported from the culture cell in the normal manner of the wild-type virus and can then be harvested from the culture system and is still infectious. In particular, we have delivered the transgene to multiple human carcinoma cell types. We are now improving the bionanosubmarine production system to produce submarines in higher quantities.

Publications

Griffiths, R.A., Boyne, J.R. & Whitehouse, A. (2006). Herpesvirus saimiri-based gene delivery vectors. *Current Gene Therapy,* **6**, 1-15.

Whitehouse, A. (2006). Gamma-2 Herpesvirus saimiri-based vectors, p 255-264. In Gene Transfer: Delivery and expression of DNA and RNA, a laboratory manual. *Cold Spring Harbor Laboratory Press.*

Acknowledgements

This project is funded by the University of Leeds Interdisciplinary Institute of Bionanoscience.

Repressosome formation and disruption regulates the KSHV latent-lytic switch

Faye Gould and Adrian Whitehouse

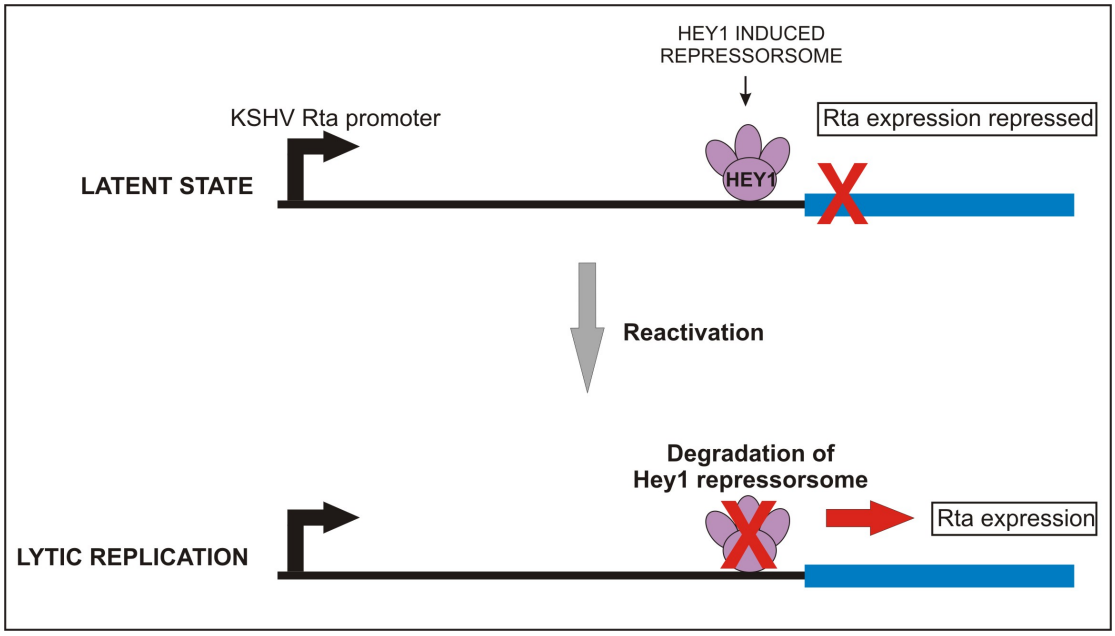
Introduction

The etiological agent of Kaposi's Sarcoma, Kaposi's sarcoma associated herpesvirus virus (KSHV), is the most recently identified human tumour virus. KSHV has two distinct forms of infection, latent persistence and lytic replication. The switch between these phases is important, as lytic replication plays an essential part in the pathogenesis and spread of KSHV infection. The KSHV ORF 50 protein is the key gene product which regulates viral lytic gene expression as sustained transient expression of ORF 50 in a KSHV-latently infected cell line leads to the stimulation of its own expression and consequently viral lytic replication. This implicates the ORF 50 protein as the molecular switch for reactivation and initiation of the KSHV lytic replication cycle.

We are currently investigating the host cell-ORF 50 interactions to further understand the role of KSHV ORF 50 in the latent-lytic switch. We have demonstrated that KSHV ORF 50 interacts with the cellular protein, Hey-1. Hey-1 functions as a transcriptional repressor, acting as adapter protein that binds to specific DNA binding sites within gene promoters, and subsequently recruits transcriptional repressosome complexes.

The interaction between KSHV ORF 50 and the transcriptional repressor protein Hey-1 is a particular intriguing one. Why would ORF 50 interact with a transcriptional repressor protein, given the role of ORF 50 in transcriptional activation and initiating the lytic replication cycle? However, we have shown that the Hey-1-ORF 50 interaction is an essential interaction playing a pivotal role in regulating the KSHV latent-lytic switch.

We have shown that Hey-1 specifically represses ORF 50 expression by binding the ORF 50 promoter and recruiting a functional repressosome, thus helping KSHV to remain in the latent state. However, we have also demonstrated that KSHV ORF 50 can act as an ubiquitin E3 ligase, which results in the degradation of Hey-1 via a proteasome-degradation pathway. This disrupts the repressosome and allows ORF 50 expression leading to KSHV lytic replication.



Repressosome assembly and disassembly regulates the KSHV latent-lytic switch

We now aim to further characterise the DNA/protein and protein/protein interactions involved in KSHV reactivation and in the regulation of lytic gene expression. This analysis will also give further insights into the role of the cellular Hey-1 protein in transcriptional repression. Moreover, this project will provide valuable information on KSHV reactivation that may ultimately lead to the identification of specific antiviral targets to inhibit ORF 50–host cell interactions which may be developed as a novel treatment for this important human pathogen.

Acknowledgements
This project is funded by the Wellcome Trust and BBSRC.

Synthesis of α-helix mimetics: inhibitors of protein-protein interactions

Fred Campbell, Jeff Plante, Bara Malkova and Andrew Wilson

Introduction

Protein-protein interactions are fundamental to the operation of many and varied cellular processes, yet, unlike the wealth of enzyme-substrate inhibition studies, the disruption of these larger binding interactions has, to date, been rarely investigated. This is because it is difficult to design small molecules that match the ill-defined presentation of functional groups across a large (600-1500Å^2) protein surface. We have initiated a program directed towards developing iterative syntheses of rigid scaffolds that project multiple functional groups in a well defined manner, such that they mimic key amino acid residues involved in α- helix mediated protein-protein interactions.

Current work

Synthetic strategies towards two different scaffolds have been optimised, allowing for inclusion of a multitude of functionality that mimics α-amino acid side chains (Fig. 1). Both scaffolds project R groups such that they spatially match reasonably well to the side chains in the α-position of the *i, i + 4* and *i + 7* amino acid residues of an α-helix. Compound **1** possesses an internal H-bond conformational lock, which is able to maintain a linear conformation as illustrated by the crystal structure of **1a** (Fig. 2a). In contrast, the *N*-alkylated scaffold prefers a *cis* orientation at the amide bond. We have exploited this in a different way to prepare macrocycles. The crystal structure of **2a** highlights how these macrocycles could be used to project functionality over a large flat surface and achieve a more generalised protein surface recognition (Fig. 2b).

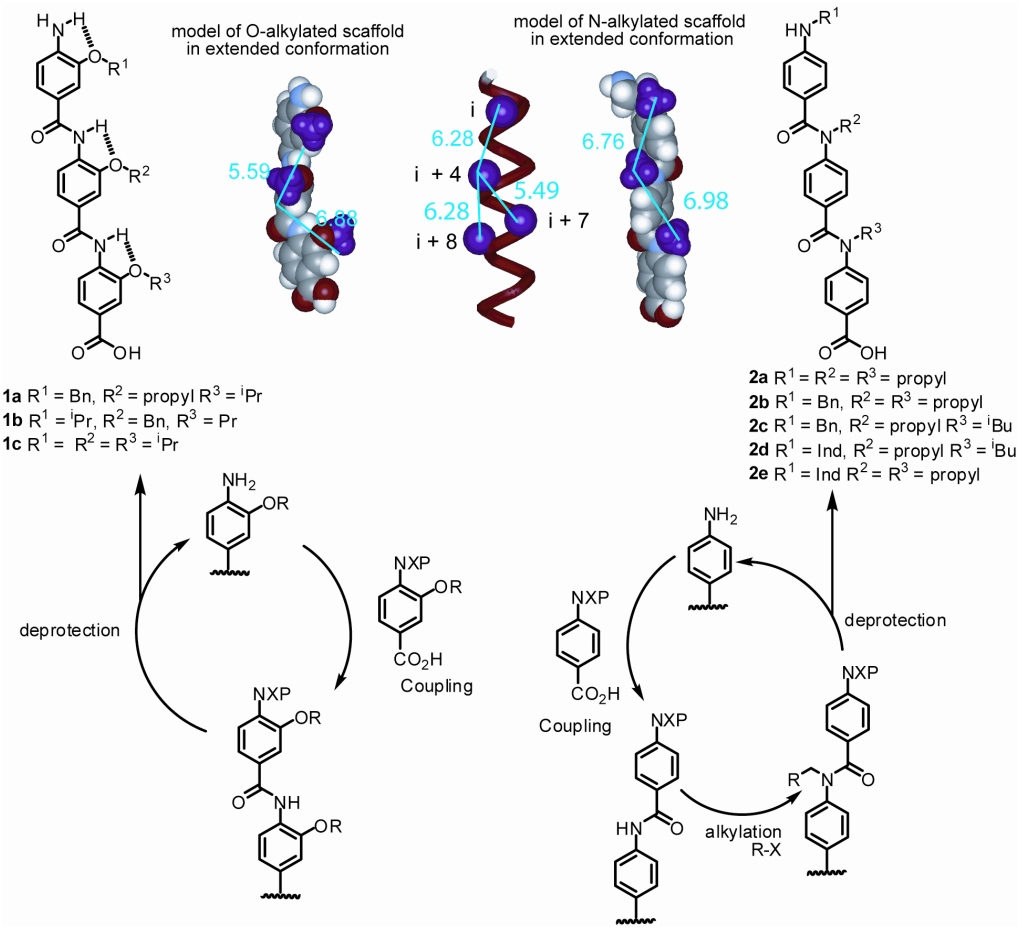
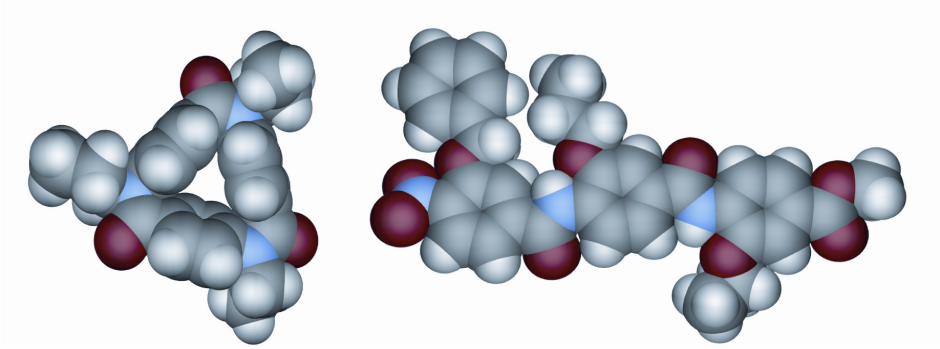


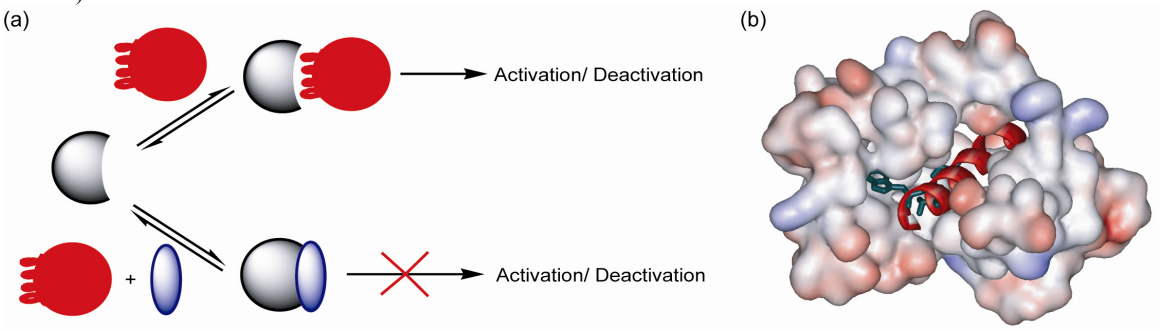
Figure 1. Synthetic route towards the two target scaffolds and successfully synthesised trimers.

Figure 2. Crystal structure of propyl-appended macrocycle derived from **2a** and the protected *O*-alkylated compound **1a** (with Bn, ^{n}Pr and ^{i}Pr side chains).



To date, five fully deprotected *N*-alkylated scaffolds (compounds **2a-e)** and three *O*-alkylated scaffolds (compounds **1a-c)** have been isolated (see Fig 1). The R groups on these scaffolds have been selected to match closely to those in smooth muscle myosin light chain kinase (smMLCK) a protein that binds in an α-helical conformation to calmodulin, through three key residues spaced at *i, i + 4* and *i +7*: Trp-800, Thr-803 and Val-807 (Fig. 3).

Fig. 3. (a) Cartoon representation showing how our synthetic ligands will be used to inhibit protein-protein interactions (α-helix binding protein represented in black, α-helix presenting protein in red and inhibitor in blue) (b) X-ray crystal structure of the α-helical sMLCK peptide (red) bound to calmodulin (Connolly surface).



Future work

Calmodulin activates the enzyme phosphodiesterase: we will test these compounds as inhibitors of this interaction by following the rate of hydrolysis of a fluorescent substrate in the presence of our inhibitor compounds. We will also develop our synthetic methods so that large libraries of inhibitor candidates can be prepared using solid phase synthesis.

Funding

We gratefully acknowledge EPSRC, the University of Leeds and The Wellcome Trust for financial support of this research.

Theoretical and computational studies of mechanical unfolding of proteins

Daniel West, Emanuele Paci and Peter D Olmsted

Over the past year we have continued our use of theory and molecular dynamics simulations to study the nature of free energies in proteins. Our motivation has been to try and understand which features govern the potential that describes forced unfolding of proteins. We have been motivated by work in Leeds using Atomic Force Microscopy to measure the mechanical properties of proteins. The study of forced unfolding of proteins using theoretical tools is part of an integrated study of the mechanical response of proteins using both experimental and simulation/theoretical methods. Our work is driven by frequent discussion with the Radford/Brockwell/Smith experimental groups within ACSMB.

Multidimensional aspects of the free energy landscape of a forced protein

Mechanical unfolding of polyproteins by force spectroscopy provides valuable insight into their free energy landscapes. Most experiments have been interpreted in terms of the Bell/Evans model for forced activation over a barrier. This physically motivated model yields, when fitting to unfolding data, a distance between the transition and native state, in the pulling direction, and a zero-force unfolding rate, analogous to the zero-denaturant unfolding rate in chemical denaturation experiments. However, proteins are inherently multidimensional objects, with up to several hundred fluctuating degrees of freedom. An open question in the field is how to reduce these many degrees of freedom to the effective one-dimensional model which fits experiments. We have shown, by considering the entire phase space of a model protein under a constant force, that the transition state displacement contains a sizeable entropic contribution from exploring the full multidimensional energy landscape. Proteins with more degrees of freedom are expected to have larger effective transition state displacements. We have tested this computationally by adding extra linker sections to unfolding proteins, and showing that even these external degrees of freedom, because they couple directly to the force, can renormalise the effective one-dimensional parameters.

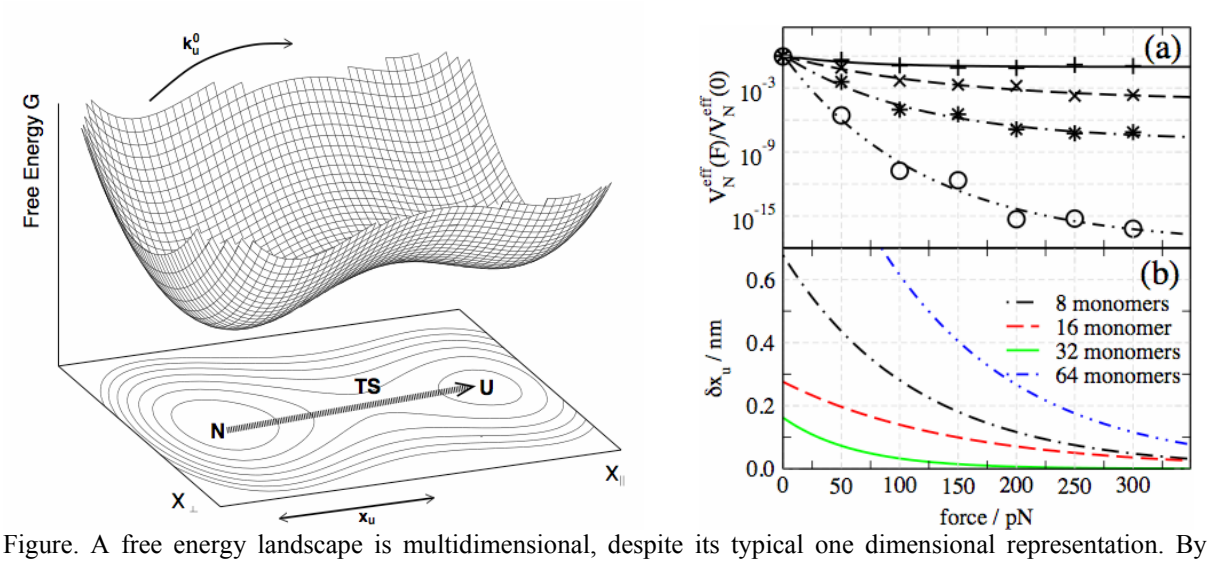
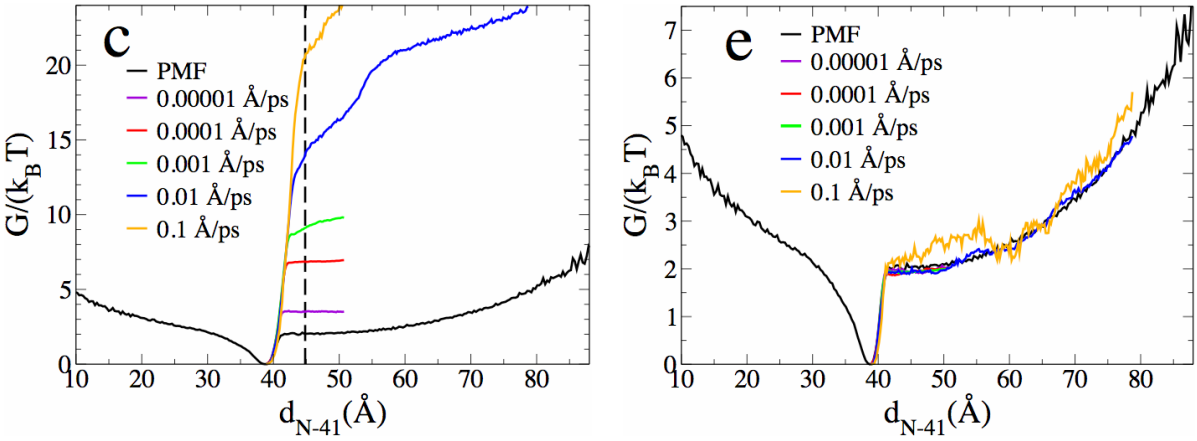


Figure. A free energy landscape is multidimensional, despite its typical one dimensional representation. By considering the fluctuation around these transverse dimensions, we have shown that the **effective** distance to the transition state x_u changes! The bottom right hand figure shows how the transition state distance changes as a function of the length of linker strands attached to the protein. These theoretical calculations match well with simulations of the same protein being pulled by these additional linker strands.

Calculating equilibrium potentials from non-equilibrium experiments and simulations

One of the most exciting new results in physics is the **Jarzynski Equality (JR)**, with which

one can use non-equilibrium experiments (or simulations) to obtain equilibrium free energies. This has raised the tantalizing possibility of using fast mechanical unfolding experiments to measure equilibrium free energy differences with respect to the native state. Hence we have been exploring the utility of this new result for proteins. We have tested this for a simple compact protein, which is small enough to calculate equilibrium potential exactly by performing many unfolding and refolding simulations (E2Lip3). We have indeed shown that the free energy can be exactly and effectively estimated from non-equilibrium simulations. However, in doing so we have found a couple of sometimes-neglected aspects of the Jarzynski Equality. First, one must carefully and correctly determine the very small fraction of denatured states which exist even at zero force; even neglecting a fraction as small as one part in 10^6 can contaminate the answer! Secondly, the potential of mean force that is found is not actually the free energy that determines mechanical stability, since it includes information from the unfolded state. However, we have outlined circumstances under which the Jarzynski Equality may be used.



Left: Free energies determined using the Jarzynski relation for different pulling speeds, which unfolding from only native configurations. In this case the reaction coordinate is the distance between the N terminus and the 41st residue. Notice that the potential is different for different pulling speeds, which is unphysical. The reason for this error is that these potentials to not include any information from unfolding the small fraction of states (in this case 14%) which are unfolded in the absence of a force. Right: By incorporating these neglected states, the extracted potential is now independent of pulling speed, and also equal to the potential of mean force (free energy) calculated from the unforced equilibrium simulation

Collaborators

Sheena Radford and David Brockwell (Institute for Molecular Biophysics), D. Alastair Smith (School of Physics and Astronomy, Institute of Molecular Biophysics

Publications

West, D.K., Paci, E. & Olmsted, P.D. (2006) Internal protein dynamics shifts the distance to the mechanical transition state, *Physical Review E* **74** 061912.

West, D.K., Olmsted, P.D. & Paci, E. (2006) Free energy of protein folding using the Jarzynski equality, *Journal of Chemical Physics* **125** (2006) 204910.

Funding

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

Molecular mechanisms of dextran and pectin extension

Bhavin Khatri, Masaru Kawakami, Katherine Burne, Igor Neelov David Adolf, Emanuele Paci, Sheena Radford, Alastair Smith and Tom McLeish.

Introduction

Polysaccharides are fundamental components of cells and have many potential applications in the pharmaceutical industry and material science. Mechanical properties of polysaccharides are important because they constitute cell walls in plants and bacteria and take part in cell interactions and adhesion. The mechanical properties of single molecules, extended by an imposed force or subjected to a constant rate of extension, can be measured in single molecule Atomic Force Microscopy (AFM). Several polysaccharides, such as dextran (1-6-polyglucose), linear pectin (1-4 polygalacturonic acid) and some others have been investigated computationally, theoretically and experimentally.

Mechanisms of polysaccharides extension

For some polysaccharides, including dextran, there is a plateau in force-extension dependence. For dextran it was suggested first that this plateau occurs due to rotation around C5-C6 bond. Later the plateau was related with force-induced structural transition of monomer rings from chair (circles, in Fig.1) to boat (squares) state. We have shown using simulation with full atomic details that there is an additional intermediate state (inverted chair conformation (triangles)) of dextran monomer rings at plateau forces (800-1200pN).

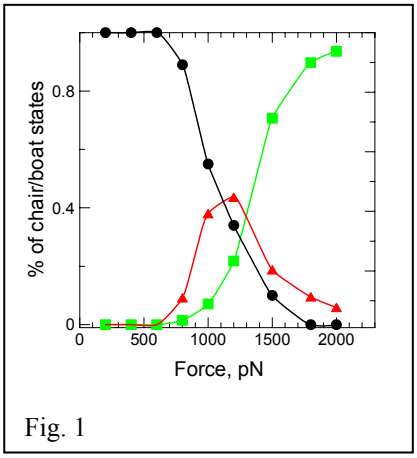


Fig. 1

This transition energy landscape was compared with an experimentally-derived one from single-molecule AFM experiments. The spectrum of thermal noise in the signal was analysed using new theoretical tools to give information on dissipation and elasticity that arise from transitions in the biopolymer. Figure 2 gives examples of noise spectra fitted to models for the dextran polymer at different degrees of extension. It was possible to deduce the separation of the two states of the monomer, their width, the position and shape of the barrier and the roughness of the energy landscape.

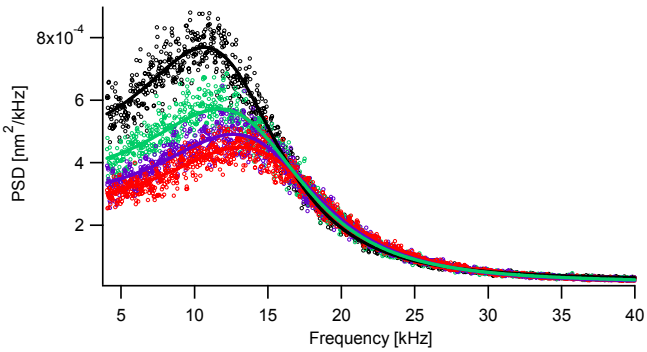


Fig. 2

Publications

Neelov I.M., Adolf D.B., McLeish, T.C.B. & Paci E. (2006) Molecular dynamics simulation of dextran extension by constant force in single molecule AFM, *Biophys. J.,* 91:3579-3588

Khatri, B.S., Kawakami, M., Byrne, K., Smith, D.A.M. & McLeish, T.C.B. (2007) "Entropy and barrier-controlled fluctuations determine conformational viscoelasticity of single biomolecules", *Biophys. J.*, **92**, 1825-1835.

Funding and Acknowledgements

This work was funded by the EPSRC.

Bacteriophage and virus-like particles as scaffolds for the fabrication of nano-scale devices.

Peter G. Stockley, Zulfiqar Hasan, Atul Mohan, Simon White, Malgorzata Wnek, Marcin Grozny, Christoph Wälti, Steve Evans and Giles Davies.

The desire to fit ever more electronic components onto microchips has, for the past thirty years, been driven by Moore's Law, which states: for silicon based integrated circuits, the number of transistors per square centimetre doubles every twelve months. Our ability to keep up with this aim has almost reached its limits with regard to current technology. Nanowires are an important component for extending current limits by building extremely small circuits in future nanoelectronic devices, such as biomolecular nanosensors. At present they are fabricated in the laboratory from inorganic components using techniques of lithography, chemical etching, deposition etc. The 'top-down' approach of fabrication lithography is, however, fast approaching its fundamental limits as the dimensions are scaled down into the nanometer range. The area where nanoscale science meets biology is known as bionanotechnology and encompasses both biomedical applications (sensors, coatings, drug delivery, imaging etc.) and bioinspired devices and processes. Nature uses a rich variety of molecules that self-assemble to perform structural and functional tasks and form the inspiration of self-assembly based nanotechnology. The properties of self assembly, simple manipulation of the viral coat proteins using genetic modification, and ease of production are what make rod-shaped viruses and virus-like particles, such as the bacteriophage M13 and the Tobacco Mosaic Virus (TMV) such inviting targets as ways to overcome the current limitations.

Figure 1. A platinum-coated nanowire templated onto TMV virus (left) and the four-probe nanoman used to probe its conductive properties (middle and right -in top close-up).

M13 is a filamentous virus that infects *Escherichia coli* expressing the F pilus. It is 1 μm long and 6 nm in diameter and is therefore already in the nanoscale size. Its major coat protein, pVIII (around 2700 copies), covers the entire length of the M13 genome. It is possible, by using phage display technology, to express a small peptide at the N-terminal of pVIII by manipulation of the pVIII gene. Using a method known as "bio-panning" it is possible to select for a peptide that can bind to, or nucleate, an inorganic material (biomineralisation). We are currently using M13 as a template for the creation of cobalt oxide and gold nanowires. Tobacco mosaic virus is another example of a virus, which can be used as a scaffold for nanowire fabrication. The TMV particle is a 300 nm long rod-shaped structure, which consists of the protein capsid and single-stranded (ss) RNA. 2130 identical copies of the coat protein subunits self-assemble on the 6400 nucleotides long ssRNA molecule to form a right-handed, single helix, with the RNA intercalated between the helix turns. In order to create conducting TMV nanowires, nickel or cobalt ions can be introduced

to the outer surface of the virus particle. For this, TMV coat protein is genetically modified and metal binding peptides presented on the outer surface. Preliminary measurements of the electrical conductivity of these phage nanowires demonstrate the potential of the technology, whilst elsewhere across the University we are developing techniques to characterise and interrogate these novel bio-templated structures.

Funding:
We thank the EPSRC and The University of Leeds for funding.

Viruses and virus-like particles as nanoscale delivery systems on a "molecular railroad".

Andrew Roche, Christoph Wälti and Peter G. Stockley.

Bionanotechnology utilises biological components for use in potential future nano-scale devices. Synthetic motors which are also capable of doing reliable mechanical work are difficult to fabricate. A convenient solution is to develop miniature electrical and mechanical systems (MEMS) based on existing biological molecular motors, e.g kinesin. Kinesins are linear molecular motors which can translocate directionally along microtubule tracks while carrying a wide range of large cellular cargoes.

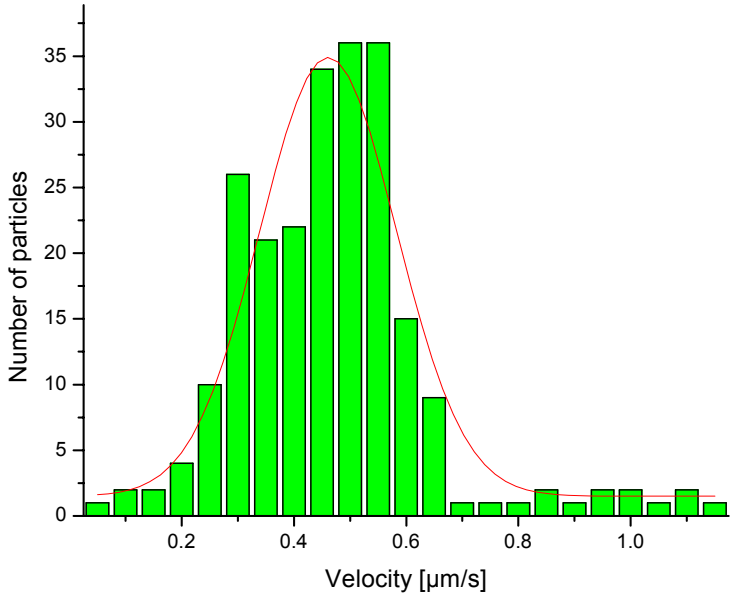


Figure 1. Distribution of the velocities of microtubules propelled by Drosophila heavy chain kinesin construct. Mean velocity = 0.46 µm / s (N=232)

This project aims to replicate such motility *in vitro* for the transport of synthetic and non-synthetic materials. Microtubule motility assays have been developed using a commercially sourced kinesin. This has allowed systematic development of the assay by introduction of a faster Drosophila heavy chain kinesin, as well as the movement of biotinylated or streptavidin-labelled microtubules which will form the connectors to other cargoes. Lithographically patterned nano-grooves are being fabricated on silicon wafers which will enable unidirectional motion of microtubules attached via streptavidin to cargoes. Magnetic or highly charged cargoes will be "steered" electronically around the kinesin "tracks". It is also intended to extend this work to improve our understanding of natural systems such as the movement of large eukaryotic viruses along microtubules.

Collaborators:
This is a joint project with Prof. Tom Wileman (University of East Anglia)

Funding:
We thank the EPSRC and The University of Leeds for funding.

The design and construction of magnetic tweezers, and the investigation of fibrin clots

Rob Harrand, Alastair Smith, Neil Thomson, Peter Grant, Ramzi Ajjan and Robert Ariens.

Introduction

In 1949, Crick reported on the motion of magnetic particles within a living cell. Today, Crick's basic idea of moving a magnetic particle using an external magnetic field is still used, but in conjunction with sophisticated optics and computer control systems, and using multiple electromagnets rather than a single permanent magnet.

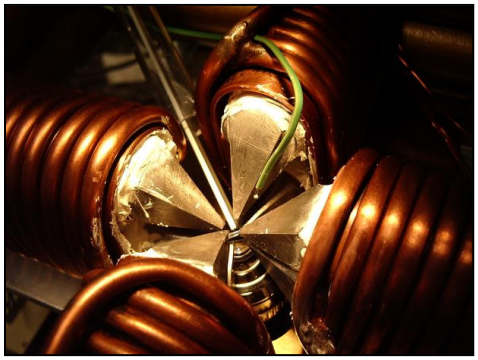


Fig 1. Custom built Magnetic Tweezers

At their most fundamental level, magnetic tweezers work by exerting a force upon a magnetic particle via an external field gradient. The magnetic particles act like small, individual bar magnets, and are tracked via an optical microscope with custom written software. When in the presence of a magnetic field gradient, they move towards the region of higher magnetic flux. Force calibration is achieved by placing the magnetic particles in a fluid of known viscosity and applying several different voltages. By measuring the velocity of the particle and using Stokes' law, the force can be calculated. These voltage levels (for a given size and type of magnetic particle) can then be used to allow force control.

Mechanical behaviour of fibrin clots

A custom-built magnetic tweezer device (Fig 1) has been used with fibrinogen samples to measure their mechanical properties over time. Fibrinogen is a protein made up of three globular regions, connected by alpha helical coils. The central globular domain possesses negatively charged fibrinopeptides, preventing fibrinogen molecules from aggregating. The enzyme thrombin cleaves these peptides, enabling the outer domain of one fibrinogen molecule to interact with the central domain of another. This creates a staggered protofibril which then laterally aggregates into a fibrin fibre. Another molecule, Factor XIII, then crosslinks the molecules to further increase their stiffness. Fibrin clots are part of the natural response to damage and injury in vascular systems.

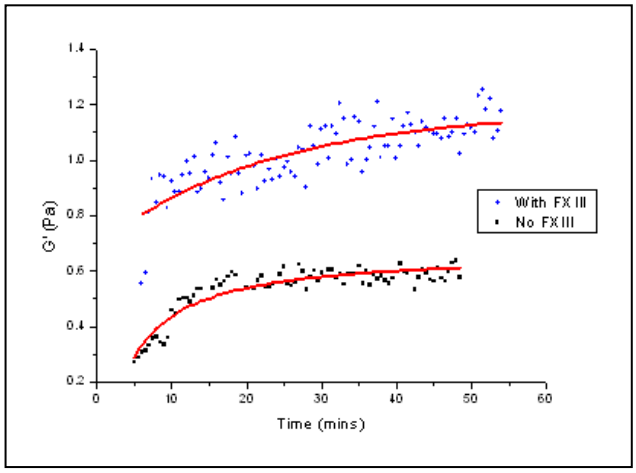


Fig 2. Increasing stiffness of fibrin clot over time

As the clot rapidly forms in the presence of magnetic particles, some are caught within the protein network. Having located a suitable particle (one that is in the centre of the sample, away from the sample holder walls), force pulses are applied and the resulting motion of the particle is tracked. In a typical clot of normal fibrinogen, a force of 100 pN may displace the particle by a few microns. This motion distorts and stretches the fibrin fibres, which forces the particle back to its original position once the external field is removed.

Results have been obtained for normal fibrinogen with and without treatment with FXIII (Fig 2), as-well-as for fibrinogen variants, such as those formed from snake venom and genetic

abnormalities. The behaviour and clot stiffness for these variants have been shown to be significantly different when compared to normal fibrinogen.

Funding:
We thank the University of Leeds for funding.

Assembly of ssRNA viruses: The role(s) of the package during packaging.

Peter G. Stockley, Alison E. Ashcroft, Gabriela Basnak, Simon E.V. Phillips, Neil Ranson, Ottar Rolfsson, Nicola J. Stonehouse, Gary S. Thompson and Katerina Toropova.

Previously we have been using the RNA bacteriophage MS2, a *T*=3 virus, as a model in which to answer fundamental questions about assembly mechanisms of ssRNA viruses. This has allowed us to define the molecular basis of genome sequence specific recognition by coat protein and in the past year this has been extended to an understanding of how discrimination between closely related viral RNAs is achieved. We have also shown for the first time in any system how the coat protein conformation is switched between quasi-equivalent conformers during the assembly process. *T*=3 capsid assembly is only efficient when two distinct conformers of the capsomer – a coat protein dimer in this case – are present. Binding the RNA stem-loop packaging signal leads to an allosteric conformational change forming a conformer of the CP_2 distinct from the one in the absence of RNA. We were also able to show that the unit of capsid growth is a coat protein (CP) dimer and that the initial higher order complexes formed are based around the particle three-fold axis rather than the five-fold. These data lead naturally to a model of assembly (Fig. 1) in which specific RNA binding causes a switch in quasi-equivalent conformer in the CP_2, thus providing both types of species needed for the final *T*=3 capsid. This begs the question of how quasi-equivalent switching is achieved beyond the initiation complex? Does the RNA folding present appropriately positioned stem-loop mimics that can function in this way, or is the initial conformational switch caused by the initiator stem loop binding sufficient to 'template' further coat protein dimers as they bind? Cryo-electron microscopy of wild-type phage particles is currently being used to address this question.

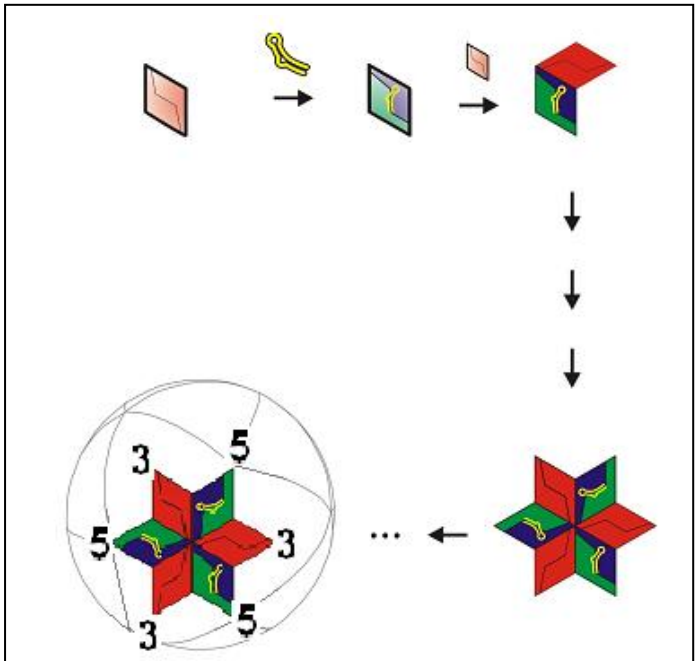


Figure 1 shows the model for MS2 assembly based on mass spectrometry and NMR spectroscopy; the allosteric RNA-CP interaction leads to formation of the initial three-fold axis in the particle, via CP dimer addition, ensuring that *T*=3 rather than *T*=1 shells are assembled.

Publications:

Horn, W.T., Tars, K., Heigstrand, C., Baron, A.J., Lago, H., Adams, C.J., Peabody, D.S., Phillips, S.E.V., Liljas, L. and Stockley, P.G.(2006). Structural basis of RNA binding discrimination between bacteriophages Q beta and MS2. *Structure* **14**, 487-495.

Stockley, P.G., Rolfsson, O., Thompson, G.S., Basnak, G., Francese, S., Stonehouse, N.J., Homans, S.W. and Ashcroft, A.E. (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a *T*=3 viral capsid. *J. Mol. Biol.* **368**, 541-552.

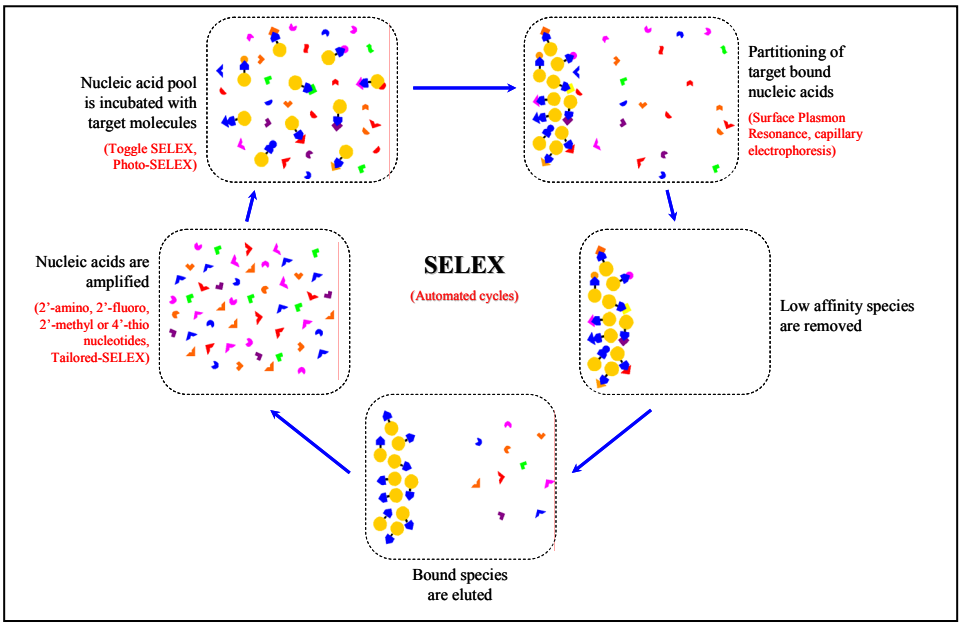
Acknowledgements:

This project is a collaboration with Dr Reidun Twarock and her PhD students Mr Tom Keef & Nick Grayson, from the Departments of Biology and Mathematics, University of York, York YO10 5DD, UK

Uses of RNA aptamers in molecular biology

David Bunka, Benjamin Mantle, Tamara Belyaeva, Bo Meng, Claire Lane, Adam Nelson, Simon Phillips, David Rowlands, Sheena Radford and Peter Stockley.

The use of DNA and RNA aptamers is becoming more and more prevalent in modern molecular biology. These highly adaptable molecules have a wide variety of applications, including identification of recognition motifs, novel inhibitors of protein function, diagnostic or therapeutic reagents and even as novel gene regulatory elements. Aptamers are isolated from complex degenerate libraries of approximately 10^{15} sequences, in a process referred to as SELEX (Systematic Evolution of Ligands by Exponential Enrichment) which is undergoing rapid technical advances (see Figure). We have successfully used our automated *in vitro* selection platform to raise aptamers against many targets, a few of which will be mentioned here.



In one project, in collaboration with Adam Nelson's laboratory, we aim to create novel regulatory elements, "riboswitches", which can be used in both prokaryotic and eukaryotic gene expression studies. Our initial target for allosteric regulation of function was the *E. coli* methionine repressor, Met J. We have previously reported the isolation of RNA aptamers which bound to Met J and were able to interfere with binding to promoter DNA *in vitro*. The next step of the project was to test these aptamers for *in vivo* activity. Using a specially designed tester strain of *E. coli,* we demonstrated that activity of a reporter gene is increased 1.8-2.5 times in presence of aptamers. Expression of aptamers in the cells was confirmed by Northern blot. Future work is directed at creating allosterically regulated versions of these aptamers.

In another project, in collaboration with Simon Phillips' laboratory, we raised aptamers against satellite tobacco necrosis virus (STNV) capsids, with the aim of understanding some of the RNA-protein interactions involved in capsid assembly. Our selections lead to the isolation of several aptamers with sequence and structural similarities to parts of the viral genome. Preliminary results suggest that this structure may act as an assembly initiation signal. In a related project, in collaboration with David Rowlands, we isolated aptamers against the poliovirus capsid protein pentamer. As with STNV, the aim of this project was to identify RNA-protein interactions important for viral assembly/RNA packaging and to determine how the viral capsid specifically packages its genome amongst all the other cellular RNAs. Our results show that the isolated aptamers specifically interact with capsid assembly intermediates. Sequence and structure comparisons have identified matches

between the poliovirus genome and isolated aptamers, again suggesting that we have identified an assembly initiation signal or recognition element. Studies on the effects of aptamers on viral assembly *in vitro* and *in vivo* are ongoing.

A separate project in collaboration with Sheena Radford had the aim of raising aptamers which specifically recognise either the fibrillar or monomeric forms of the amyloidogenic protein, beta-2 microglobulin (β_2m). We were successful in isolating aptamers that preferentially bind to β_2m fibrils. Interestingly, these aptamers also recognised β_2m fibril morphologies other than the selection target. This suggests some degree of structural similarity between these morphologies. Strikingly we also noted that these aptamers were also able to specifically recognise fibrils formed from a few other amyloidogenic proteins. This suggests that there are structural similarities between fibrils formed by these different proteins.

Publications:

Bunka, D. H. J. & Stockley, P. G. (2006). Aptamers come of age - at last! *Nature Reviews Microbiology* **4**, 588-596.

Funding:

We would like to thank the Wellcome Trust, BBSRC, MRC and EPSRC for funding these various projects.

Mapping ATP-Dependent Activation at a Sigma54 Promoter

Rob N. Leach, A.Miah, Simon Baumberg♦, Chris Gell,
D. Alastair Smith and Peter G. Stockley

The σ^{54} promoter specificity factor is distinct from other bacterial RNA polymerase (RNAP) sigma factors in that it forms a transcriptionally silent closed complex upon promoter binding. Transcriptional activation occurs through a nucleotide-dependent isomerisation of σ^{54}, mediated via its interactions with an enhancer-binding activator protein that utilises the energy released in ATP hydrolysis to effect structural changes in σ^{54} and core RNA polymerase. The organization of σ^{54}-promoter and σ^{54}-RNAP-promoter complexes was investigated by FRET assays using σ^{54} single cysteine-mutants labelled with an acceptor fluorophore and donor fluorophore-labelled DNA sequences containing mismatches that mimic *nifH* early- and late-melted promoters.

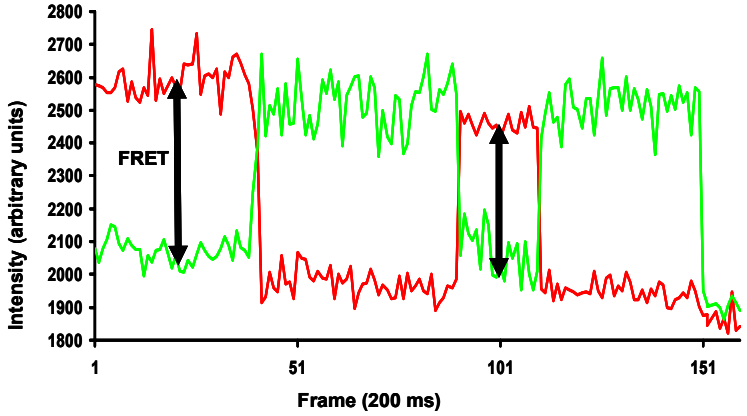


Figure. Time-dependent FRET signals observed from single DNA heteroduplexes immobilised on a glass slide and observed via TIRF.

The results show that σ^{54} undergoes spatial re-arrangements of functionally important domains upon closed complex formation. σ^{54} and σ^{54}-RNAP promoter complexes reconstituted with the different mismatched DNA constructs were assayed by addition of the activator PspF in the presence or absence of ATP and of non-hydrolysable analogues. Nucleotide-dependent alterations in FRET efficiencies identify different functional states of the activator-σ^{54}-RNAP-promoter complex that exist throughout the mechano-chemical transduction pathway of transcriptional activation, i.e. from closed to open promoter complexes. The results suggest that open complex formation only occurs efficiently on replacement of a repressive fork junction with downstream melted DNA.

In order to deconvolute the complex interactions and conformational changes occurring at such promoters, and to extend the work to include the arginine catabolic operators from *Bacillus subtilis* which also operate via a similar mechanism, we have used single molecule TIRF to analyse changes in the FRET signals described above. This allows us to use novel assays for the conformational and functional states of this complex transcriptional machinery.

Publications:

Leach, R., Gell, C., Wigneshveraraj, S., Buck, M., Smith, D. A. M. & Stockley, P. G. (2006). Mapping ATP-Dependent Activation at a Sigma54 Promoter. *J. Biol. Chem.* **281**, 33717-33726.

Acknowledgements:

This is a collaborative project with Prof M. Buck & Dr S. Romesh Wigneshweraraj from the Division of Biology, Faculty of Natural Sciences, Sir Alexander Fleming Building, Imperial College, London, SW72AZ, UK.

Funding:

We thank the BBSRC for funding.

♦ deceased 11th April, 2007.

Astbury Seminars 2006

2nd February 2006
Dr Sarah Perrett, Beijing Institute of Biophysics, University of Cambridge
"The yeast prion protein Ure2: Function, folding and amyloid formation"

9th February 2006
Prof Michael Sundstrom and Prof Stefan Knapp, Genomics Consortium, University of Oxford
"*Directed structural genomics: Focus on human protein families*" and "*New insights for the development of specific kinas inhibitors by targetted structural genomics*"

16th March 2006
Dr Jim Warwicker, Bioinformatics Department of Biomolecular Sciences, University of Manchester
"Modelling charge interactions and pH dependence in biology"

27th April 2006
Dr David Dryden, School of Chemistry, University of Edinburgh
"Restriction and anti-restriction"

18th May 2006
Dr Jose Pereira-Leal, Instituto Gulbenkian de Ciencia, Portugal
"Evolution of functional modules in protein networks"

15th June 2006
Dr Matthias Geyer, Max Planck Institute for Molecular Physiology, Dortmund
"Biochemical analysis of membrane-binding properties in HIV-1 Nef: Insights to curvature, composition and signal motifs"

6th July 2006
Prof Steve Chapman, School of Chemistry, University of Edinburgh
"The magic of haem: advice to young scientists from a wise man"

12th September 2006
Prof Satoshi Murakami, Department of Cell Membrane Biology, Osaka University
"A functionally rotating mechanism revealed in crystal structures of the multidrug transporter, AcrB, with bound substrates"

14th September 2006
Prof David Towers, School of Mechanical Engineering, University of Leeds
"BioPhotonics: Quantitative optical sensor for microscopy"

12th October 2006
Prof Steve Homans, Astbury Centre, University of Leeds
"What Drives Ligand-Protein Interactions?"

19th October 2006
Dr Reidun Twarock, The University of York
"A Mathematical Physicist's View of Virus Structure And Assembly"

16th November 2006
Dr Ruth Brenk, School of Life Sciences, University of Dundee
"Probing molecular docking in a charged model binding site"

20th November 2006
Prof Bentley Fane, Department of Veterinary Science & Microbiology, University of Arizona
"Structure, function and evolution of viral scaffolding proteins in a two scaffolding protein system"

23rd November 2006
Dr Jun Liu, National Institutes of Health, Bethesda, USA
"Pushing the envelope of cryo electron tomography: 3-D structure of myosin-V inhibited state"

Publications by Astbury Centre Members 2006

Anstiss, M. & Nelson, A. (2006). Catalytic and stoichiometric approaches to the desymmetrisation of centrosymmetric piperazines by enantioselective acylation: a total synthesis of Dragmacidin A. *Org. Biomol. Chem.* 4, 4135-4143.

Arscott, S., Gaudet, M., Brinkmann, M., Ashcroft, A. E. & Blossey, R. (2006). Capillary filling of miniaturized sources for electrospray mass spectrometry. *J. Phys.-Condes. Matter* 18, S677-S690.

Barnes, K., Dobrzynski, H., Foppolo, S., Beal, P. R., Ismat, F., Scullion, E. R., Sun, L. J., Tellez, J., Ritzel, M. W. L., Claycomb, W. C., Cass, C. E., Young, J. D., Billeter-Clark, R., Boyett, M. R. & Baldwin, S. A. (2006). Distribution and functional characterization of equilibrative nucleoside transporter-4, a novel cardiac adenosine transporter activated at acidic pH. *Circulation Research* 99, 510-519.

Barratt, E., Bronowska, A., Vondrasek, J., Cerny, J., Bingham, R., Phillips, S. & Homans, S. W. (2006). Thermodynamic penalty arising from burial of a ligand polar group within a hydrophobic pocket of a protein receptor. *J. Mol. Biol.* 362, 994-1003.

Bentham, M., Mazaleyrat, S. & Harris, M. (2006). A cluster of arginine residues near the N-terminus of the HIV-1 Nef protein is required both for membrane association and CD4 down-modulation. *J. Gen. Virol.* 87, 563-571.

Bentham, M., Mazaleyrat, S. & Harris, M. (2006). Role of myristoylation and N-terminal basic residues in membrane association of the human immunodeficiency virus type 1 Nef protein. *J. Gen. Virol.* 87, 563-571.

Besong, G., Jarowicki, K., Kocienski, P. J., Sliwinski, E. & Boyle, F. T. (2006). Synthesis of (S)-(-)-N-acetylcolchinol using intramolecular biaryl oxidative coupling. *Org. Biomol. Chem.* 4, 2193-2207.

Bingham, R. J., Dive, V., Phillips, S. E. V., Shirras, A. D. & Isaac, R. E. (2006). Structural diversity of angiotensin-converting enzyme - Insights from structure-activity comparisons of two Drosophila enzymes. *Febs J.* 273, 362-373.

Boda, K. & Johnson, A. P. (2006). Molecular complexity analysis of *de novo* designed ligands. *Journal of Medicinal Chemistry* 49, 5869-5879.

Boyne, J. R. & Whitehouse, A. (2006). gamma-2 Herpes virus post-transcriptional gene regulation. *Clin. Microbiol. Infect.* 12, 110-117.

Boyne, J. R. & Whitehouse, A. (2006). Nucleolar trafficking is essential for nuclear export of intronless herpesvirus mRNA. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15190-15195.

Bracey, K. & Wray, D. (2006). Tubulin as a possible binding partner of the heag2 potassium channel. *Acta Pharmacologica Sinica* 27, 186-186.

Bradford, J. R., Needham, C. J., Bulpitt, A. J. & Westhead, D. R. (2006). Insights into protein-protein interfaces using a Bayesian network prediction method. *J. Mol. Biol.* 362, 365-386.

Bunka, D. H. J. & Stockley, P. G. (2006). Aptamers come of age - at last. *Nature Reviews Microbiology* 4, 588-596.

Burgoyne, N. J. & Jackson, R. M. (2006). Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* 22, 1335-1342.

Carr, S. B., Makris, G., Phillips, S. E. V. & Thomas, C. D. (2006). Crystallization and preliminary X-ray diffraction analysis of two N-terminal fragments of the DNA-cleavage domain of topoisomerase IV from *Staphylococcus aureus*. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications* 62, 1164-1167.

Caryl, J. A. & Thomas, C. D. (2006). Investigating the basis of substrate recognition in the pC221 relaxosome. *Molecular Microbiology* 60, 1302-1318.

Clare, D. K., Orlova, E. V., Finbow, M. A., Harrison, M. A., Findlay, J. B. C. & Saibil, H. R. (2006). An expanded and flexible form of the vacuolar ATPase membrane sector. *Structure* 14, 1149-1156.

Clarke, D., Griffin, S., Beales, L., Gelais, C. S., Burgess, S., Harris, M. & Rowlands, D. (2006). Evidence for the formation of a heptameric ion channel complex by the hepatitis C virus p7 protein *in vitro*. *J. Biol. Chem.* 281, 37057-37068.

Cobos, E. S. & Radford, S. E. (2006). Sulfate-induced effects in the on-pathway intermediate of the bacterial immunity protein Im7. *Biochemistry* 45, 2274-2282.

Connell, S. D. & Smith, D. A. (2006). The atomic force microscope as a tool for studying phase separation in lipid membranes. *Mol. Membr. Biol.* 23, 17-28.

Cooksey, J. P., Kocienski, P. J., Li, Y. F., Schunk, S. & Snaddon, T. N. (2006). Synthesis of the C1-C16 fragment of ionomycin using a neutral (eta(3)-allyl)iron complex. *Org. Biomol. Chem.* 4, 3325-3336.

Cordy, J. M., Hooper, N. M. & Turner, A. J. (2006). The involvement of lipid rafts in Alzheimer's disease. *Mol. Membr. Biol.* 23, 111-122.

Crampton, N., Bonass, W. A., Kirkham, J., Rivetti, C. & Thomson, N. H. (2006). Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Res.* 34, 5416-5425.

Crampton, N., Bonass, W. A., Kirkham, J. & Thomson, N. H. (2006). Studying silane mobility on hydrated mica using ambient AFM. *Ultramicroscopy* 106, 765-770.

Crampton, N., Thomson, N. H., Kirkham, J., Gibson, C. W. & Bonass, W. A. (2006). Imaging RNA polymerase-amelogenin gene complexes with single molecule resolution using atomic force microscopy. *European Journal of Oral Sciences* 114, 133-138.

Crawford, C., Nelson, A. & Patel, I. (2006). Development of an approach to the synthesis of the ABC ring system of Hemibrevetoxin B. *Organic Letters* 8, 4231-4234.

Das, C., Inkson, N. J., Read, D. J., Kelmanson, M. A. & McLeish, T. C. B. (2006). Computational linear rheology of general branch-on-branch polymers. *Journal of Rheology* 50, 207-235.

Das, C., Read, D. J., Kelmanson, M. A. & McLeish, T. C. B. (2006). Dynamic scaling in entangled mean-field gelation polymers. *Phys. Rev. E* 74.

Davies, R. P. W., Aggeli, A., Beevers, A. J., Boden, N., Carrick, L. M., Fishwick, C. W. G., McLeish, T. C. B., Nyrkova, I. & Semenov, A. N. (2006). Self-assembling beta-sheet tape forming peptides. *Supramolecular Chemistry* 18, 435-443.

Dobrzynski, H., Tellez, J. O., Barnes, K., Chandler, N. J., Qureshi, M. A., Imrit, N., Greener, I. D., Baldwin, S. A., Billeter, R., Boyett, M. R. & Howarth, F. C. (2006). Effects of streptozotocin-induced diabetes on connexin43 mRNA and protein expression in ventricular muscle. *Journal of Molecular and Cellular Cardiology* 40, 1008-1008.

Dove, B., Brooks, G., Bicknell, K., Wurm, T. & Hiscox, J. A. (2006). Cell cycle perturbations induced by infection with the coronavirus infectious bronchitis virus and their effect on virus replication. *J. Virol.* 80, 4147-4156.

Dove, B. K., Bicknell, K., Brooks, G., Harrison, S. & Hiscox, J. A. (2006). Infectious bronchitis coronavirus induces cell-cycle perturbations. In *Nidoviruses: Toward Control of Sars and Other Nidovirus Diseases*, Vol. 581, pp. 357-362.

Dove, B. K., You, J. H., Reed, M. L., Emmett, S. R., Brooks, G. & Hiscox, J. A. (2006). Changes in nucleolar morphology and proteins during infection with the coronavirus infectious bronchitis virus. *Cellular Microbiology* 8, 1147-1157.

Dubuis, E., Rockliffe, N., Hussain, M., Boyett, M., Wray, D. & Gawler, D. (2006). Evidence for multiple Src binding sites on the alpha 1c L-type Ca^{2+} channel and their roles in activity regulation. *Cardiovascular Research* 69, 391-401.

Edwards, T. A., Butterwick, J. A., Zeng, L., Gupta, Y. K., Wang, X., Wharton, R. P., Palmer, A. G. & Aggarwal, A. K. (2006). Solution structure of the Vts1 SAM domain in the presence of RNA. *J. Mol. Biol.* 356, 1065-1072.

Ellingham, M., Bunka, D. H. J., Rowlands, D. J. & Stonehouse, N. J. (2006). Selection and characterization of RNA aptamers to the RNA-dependent RNA polymerase from foot-and-mouth disease virus. *RNA* 12, 1970-1979.

Ewan, L. C., Jopling, H. M., Jia, H. Y., Mittar, S., Bagherzadeh, A., Howell, G. J., Walker, J. H., Zachary, I. C. & Ponnambalam, S. (2006). Intrinsic tyrosine kinase activity is required for vascular endothelial growth factor receptor 2 ubiquitination, sorting and degradation in endothelial cells. *Traffic* 7, 1270-1282.

Geierhaas, C. D., Best, R. B., Paci, E., Vendruscolo, M. & Clarke, J. (2006). Structural comparison of the two alternative transition states for folding of TI I27. *Biophys. J.* 91, 263-275.

Gillings, A. S., Yang, J., Marley, A. E., Holman, G. D., Peckham, M., Kilgour, E. & Baldwin, S. A. (2006). Investigation of the role of Kv1.3 as a potential regulator of peripheral insulin sensitivity. *Diabetologia* 49, 366-366.

Gold, N. D. & Jackson, R. M. (2006). Fold independent structural comparisons of protein-ligand binding sites for exploring functional relationships. *J. Mol. Biol.* 355, 1112-1124.

Gold, N. D. & Jackson, R. M. (2006). A searchable database for comparing protein-ligand binding sites for the analysis of structure-function relationships. *Journal of Chemical Information and Modeling* 46, 736-742.

Gold, N. D. & Jackson, R. M. (2006). SitesBase: a database for structure-based protein-ligand binding site comparisons. *Nucleic Acids Res.* 34, D231-D234.

Gosal, W. S., Myers, S. L., Radford, S. E. & Thomson, N. H. (2006). Amyloid under the atomic force microscope. *Protein Pept. Lett.* 13, 261-270.

Graham, R. S., Bent, J., Hutchings, L. R., Richards, R. W., Groves, D. J., Embery, J., Nicholson, T. M., McLeish, T. C. B., Likhtman, A. E., Harlen, O. G., Read, D. J., Gough, T., Spares, R., Coates, P. D. & Grillo, I. (2006). Measuring and predicting the dynamics of linear monodisperse entangled polymers in rapid flow through an abrupt contraction. A small angle neutron scattering study. *Macromolecules* 39, 2700-2709.

Griffiths, R. A., Boyne, J. A. & Whitehouse, A. (2006). Herpesvirus saimiri-based gene delivery vectors. *Current Gene Therapy* 6, 1-15.

Gsponer, J., Hopearuoho, H., Whittaker, S. B. M., Spence, G. R., Moore, G. R., Paci, E., Radford, S. E. & Vendruscolo, M. (2006). Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl. Acad. Sci. U. S. A.* 103, 99-104.

Hawkins, R. J. & McLeish, T. C. B. (2006). Coupling of global and local vibrational modes in dynamic allostery of proteins. *Biophys. J.* 91, 2055-2062.

Hawkins, R. J. & McLeish, T. C. B. (2006). Dynamic allostery of protein alpha helical coiled-coils. *Journal of the Royal Society Interface* 3, 125-138.

Heeley, E. L., Fernyhough, C. M., Graham, R. S., Olmsted, P. D., Inkson, N. J., Embery, J., Groves, D. J., McLeish, T. C. B., Morgovan, A. C., Meneau, F., Bras, W. & Ryan, A. J. (2006). Shear-induced crystallization in blends of model linear and long-chain branched hydrogenated polybutadienes. *Macromolecules* 39, 5058-5071.

Heikkila, T., Thirumalairajan, S., Davies, M., Parsons, M. R., McConkey, A. G., Fishwick, C. W. G. & Johnson, A. P. (2006). The first *de novo* designed inhibitors of *Plasmodium falciparum* dihydroorotate dehydrogenase. *Bioorg. Med. Chem. Lett.* 16, 88-92.

Holyoake, J., Caulfeild, V., Baldwin, S. A. & Sansom, M. S. P. (2006). Modeling, docking, and simulation of the major facilitator superfamily. *Biophys. J.* 91, L84-L86.

Hooper, N. M. (2006). Foreword: lipid rafts/biophysics, cell signalling, trafficking and processing. *Mol. Membr. Biol.* 23, 1-3.

Horn, W. T., Tars, K., Grahn, E., Heigstrand, C., Baron, A. J., Lago, H., Adams, C. J., Peabody, D. S., Phillips, S. E. V., Stonehouse, N. J., Liljas, L. & Stockley, P. G.

(2006). Structural basis of RNA binding discrimination between bacteriophages Q beta and MS2. *Structure* 14, 487-495.

Hyland, C., Pinney, J. W., McConkey, G. A. & Westhead, D. R. (2006). metaSHARK: a WWW

platform for interactive exploration of metabolic networks. *Nucleic Acids Res.* 34, W725-

W728.

Inkson, N. J., Graham, R. S., McLeish, T. C. B., Groves, D. J. & Fernyhough, C. M. (2006). Viscoelasticity of monodisperse comb polymer melts. *Macromolecules* 39, 4217-4227.

Jack, E., Newsome, M., Stockley, P. G., Radford, S. E. & Middleton, D. A. (2006). The organization of aromatic side groups in an amyloid fibril probed by solid-state H-2 and F-19 NMR spectroscopy. *J. Am. Chem. Soc.* 128, 8098-8099.

Jahn, T. R., Parker, M. J., Homans, S. W. & Radford, S. E. (2006). Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Mol. Biol.* 13, 195-201.

Jen, C. H., Manfield, I. W., Michalopoulos, I., Pinney, J. W., Willats, W. G. T., Gilmartin, P. M. & Westhead, D. R. (2006). The Arabidopsis co-expression tool ((ACT)): a WWW-based tool and database for microarray-based gene expression analysis. *Plant Journal* 46, 336-348.

Jeuken, L. J. C., Connell, S. D., Henderson, P. J. F., Gennis, R. B., Evans, S. D. & Bushby, R. J. (2006). Redox enzymes in tethered membranes. *J. Am. Chem. Soc.* 128, 1711-1716.

Johnson, A. P., Boda, K., Weaver, S., Valko, A. & Valko, V. (2006). *De novo* design tools for the generation of synthetically accessible ligands. *Abstr. Pap. Am. Chem. Soc.* 231.

Johnson, R., Gamblin, R. J., Ooi, L., Bruce, A. W., Donaldson, I. J., Westhead, D. R., Wood, I. C., Jackson, R. M. & Buckley, N. J. (2006). Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. *Nucleic Acids Res.* 34, 3862-3877.

Ju, M. & Wray, D. (2006). Molecular regions responsible for differences in activation between heag channels. *Biochem. Biophys. Res. Commun.* 342, 1088-1097.

Kawakami, M., Byrne, K., Brockwell, D. J., Radford, S. E. & Smith, D. A. (2006). Viscoelastic study of the mechanical unfolding of a protein by AFM. *Biophys. J.* 91, L16-L18.

Kawakami, M., Byrne, K., Khatri, B. S., McLeish, T. C. B. & Smith, D. A. (2006). Viscoelastic properties of single poly(ethylene glycol) molecules. *Chemphyschem* 7, 1710-1716.

King, A. E., Ackley, M. A., Cass, C. E., Young, J. D. & Baldwin, S. A. (2006). Nucleoside transporters: from scavengers to novel therapeutic targets. *Trends Pharmacol. Sci.* 27, 416-425.

Kirkham, J., Andreev, I., Robinson, C., Brookes, S. J., Shore, R. C. & Smith, D. A. (2006). Evidence for direct amelogenin-target cell interactions using dynamic force spectroscopy. *European Journal of Oral Sciences* 114, 219-224.

Kobrinsky, E., Stevens, L., Kazmi, Y., Wray, D. & Soldatov, N. M. (2006). Molecular rearrangements of the K(v)2.1 potassium channel termini associated with voltage gating. *J. Biol. Chem.* 281, 19233-19240.

Laguri, C., Stenzel, R. A., Donohue, T. J., Phillips-Jones, M. K. & Williamson, M. P. (2006). Activation of the global gene regulator PrrA (RegA) from *Rhodobacter sphaeroides*. *Biochemistry* 45, 7872-7881.

Laurie, A. T. R. & Jackson, R. M. (2006). Methods for the prediction of protein-ligand binding sites for Structure-Based Drug Design and virtual ligand screening. *Current Protein & Peptide Science* 7, 395-406.

Le Duff, C. S., Whittaker, S. B. M., Radford, S. E. & Moore, G. R. (2006). Characterisation of the conformational properties of urea-unfolded Im7: Implications for the early stages of protein folding. *J. Mol. Biol.* 364, 824-835.

Leach, R. N., Gell, C., Wigneshweraraj, S., Buck, M., Smith, A. & Stockley, P. G. (2006). Mapping ATP-dependent activation at sigma(54) promoter. *J. Biol. Chem.* 281, 33717-33726.

Mankouri, J., Taneja, T. K., Smith, A. J., Ponnambalam, S. & Sivaprasadarao, A. (2006). Kir6.2 mutations causing neonatal diabetes prevent endocytosis of ATP-sensitive potassium channels. *Embo J.* 25, 4142-4151.

Marincs, F., Manfield, I. W., Stead, J. A., McDowall, K. J. & Stockley, P. G. (2006). Transcript analysis reveals an extended regulon and the importance of protein-protein co-operativity for the *Escherichia coli* methionine repressor. *Biochem. J.* 396, 227-234.

Masca, S. I., Rodriguez-Mendieta, I. R., Friel, C. T., Radford, S. E. & Smith, D. A. (2006). Detailed evaluation of the performance of microfluidic T mixers using fluorescence and ultraviolet resonance Raman spectroscopy. *Review of Scientific Instruments* 77.

McCormick, C. J., Brown, D., Griffin, S., Challinor, L., Rowlands, D. J. & Harris, M. (2006). A link between translation of the hepatitis C virus polyprotein and polymerase function; possible consequences for hyperphosphorylation of NS5A. *J. Gen. Virol.* 87, 93-102.

McCormick, C. J., Maucourant, S., Griffin, S., Rowlands, D. J. & Harris, M. (2006). Tagging of NS5A expressed from a functional hepatitis C virus replicon. *J. Gen. Virol.* 87, 635-640.

McLeish, T. C. B. (2006). Diffusive searches in high-dimensional spaces and apparent 'two-state' behaviour in protein folding. *J. Phys.-Condes. Matter* 18, 1861-1868.

Murphy, J. E., Tacon, D., Tedbury, P. R., Hadden, J. M., Knowling, S., Sawamura, T., Peckham, M., Phillips, S. E. V., Walker, J. H. & Ponnambalam, S. (2006). LOX-1 scavenger receptor mediates calcium-dependent recognition of phosphatidylserine and apoptotic cells. *Biochem. J.* 393, 107-115.

Myers, S. L., Jones, S., Jahn, T. R., Morten, I. J., Tennent, G. A., Hewitt, E. W. & Radford, S. E. (2006). A systematic study of the effect of physiological factors on beta 2-microglobulin amyloid formation at neutral pH. *Biochemistry* 45, 2311-2321.

Myers, S. L., Thomson, N. H., Radford, S. E. & Ashcroft, A. E. (2006). Investigating the structural properties of amyloid-like fibrils formed *in vitro* from beta 2-microglobulin using limited proteolysis and electrospray ionisation mass spectrometry. *Rapid Commun. Mass Spectrom.* 20, 1628-1636.

Needham, C. J., Bradford, J. R., Bulpitt, A. J., Care, M. A. & Westhead, D. R. (2006). Predicting the effect of missense mutations on protein function: analysis with Bayesian networks. *BMC Bioinformatics* 7.

Needham, C. J., Bradford, J. R., Bulpitt, A. J. & Westhead, D. R. (2006). Inference in Bayesian networks. *Nature Biotechnology* 24, 51-53.

Neelov, I. M., Adolf, D. B., McLeish, T. C. B. & Paci, E. (2006). Molecular dynamics simulation of dextran extension by constant force in single molecule AFM. *Biophys. J.* 91, 3579-3588.

Netherton, C. L., McCrossan, M. C., Denyer, M., Ponnambalam, S., Armstrong, J., Takamatsu, H. H. & Wileman, T. E. (2006). African swine fever virus causes microtubule-dependent dispersal of the trans-golgi network and slows delivery of membrane protein to the plasma membrane. *J. Virol.* 80, 11385-11392.

Omaetxebarria, M. J., Hagglund, P., Elortza, F., Hooper, N. M., Arizmendi, J. M. & Jensen, O. N. (2006). Isolation and characterization of glycosylphosphatidylinositol-anchored peptides by hydrophilic interaction chromatography and MALDI tandem mass spectrometry. *Analytical Chemistry* 78, 3335-3341.

Phillips, S. E. V. (2006). Turning up the HEAT on translation. *Structure* 14, 806-807.

Potter, C. A., Jeong, E. L., Williamson, M. P., Henderson, P. J. F. & Phillips-Jones, M. K. (2006). Redox-responsive *in vitro* modulation of the signalling state of the isolated PrrB sensor kinase of *Rhodobacter sphaeroides* NCIB 8253. *FEBS Lett.* 580, 3206-3210.

Protopapa, E., Aggeli, A., Boden, N., Knowles, P. F., Salay, L. C. & Nelson, A. (2006). Electrochemical screening of self-assembling beta-sheet peptides using supported phospholipid monolayers. *Medical Engineering & Physics* 28, 944-955.

Radford, S. E. (2006). GroEL: More than just a folding cage. *Cell* 125, 831-833.

Ranson, N., Stromer, T., Bousset, L., Melki, R. & Serpell, L. C. (2006). Insights into the architecture of the Ure2p yeast protein assemblies from helical twisted fibrils. *Protein Sci.* 15, 2481-2487.

Ranson, N. A., Clare, D. K., Farr, G. W., Houldershaw, D., Horwich, A. L. & Saibil, H. R. (2006). Allosteric signaling of ATP hydrolysis in GroEL-GroES complexes. *Nat. Struct. Mol. Biol.* 13, 147-152.

Redondo, C., Burke, B. J. & Findlay, J. B. C. (2006). The retinol-binding protein system: A potential paradigm for steroid-binding globulins? *Hormone and Metabolic Research* 38, 269-278.

Reed, M. L., Dove, B. K., Jackson, R. M., Collins, R., Brooks, G. & Hiscox, J. A. (2006). Delineation and modelling of a nucleolar retention signal in the coronavirus nucleocapsid protein. *Traffic* 7, 833-848.

Rella, M. & Jackson, R. M. (2006). Identification of novel ACE2 inhibitors by structure-based pharmacophore modeling and virtual screening. *Abstr. Pap. Am. Chem. Soc.* 231.

Rella, M., Rushworth, C. A., Guy, J. L., Turner, A. J., Langer, T. & Jackson, R. M. (2006). Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. *Journal of Chemical Information and Modeling* 46, 708-716.

Remaut, H., Rose, R. J., Hannan, T. J., Hultgren, S. J., Radford, S. E., Ashcroft, A. E. & Waksman, G. (2006). Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol. Cell* 22, 831-842.

Robinson, M. A., Wood, J. P. A., Capaldi, S. A., Baron, A. J., Gell, C., Smith, D. A. & Stonehouse, N. J. (2006). Affinity of molecular interactions in the bacteriophage phi 29 DNA packaging motor. *Nucleic Acids Res.* 34, 2698-2709.

Saidijam, M., Benedetti, G., Ren, Q. H., Xu, Z. Q., Hoyle, C. J., Palmer, S. L., Ward, A., Bettaney, K. E., Szakonyi, G., Meuller, J., Morrison, S., Pos, M. K., Butaye, P., Walravens, K., Langton, K., Herbert, R. B., Skurray, R. A., Paulsen, I. T., O'Reilly, J., Rutherford, N. G., Brown, M. H., Bill, R. M. & Henderson, P. J. F. (2006). Microbial drug efflux proteins of the major facilitator superfamily. *Current Drug Targets* 7, 793-811.

Seeber, M., Fanelli, F., Paci, E. & Caflisch, A. (2006). Sequential unfolding of individual helices of bacterioopsin observed in molecular dynamics simulations of extraction from the purple membrane. *Biophys. J.* 91, 3276-3284.

Shimokhina, N., Bronowska, A. & Homans, S. W. (2006). Contribution of ligand desolvation to binding thermodynamics in a ligand-protein interaction. *Angew. Chem.-Int. Edit.* 45, 6374-6376.

Smith, A. M., Jahn, T. R., Ashcroft, A. E. & Radford, S. E. (2006). Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry. *J. Mol. Biol.* 364, 9-19.

Spencer, K. A. & Hiscox, J. A. (2006). Characterisation of the RNA binding properties of the coronavirus infectious bronchitis virus nucleocapsid protein amino-terminal region. *FEBS Lett.* 580, 5993-5998.

Spencer, K. A. & Hiscox, J. A. (2006). Expression and structural analysis of infectious bronchitis virus nucleoprotein. In *Nidoviruses: Toward Control of Sars and Other Nidovirus Diseases*, Vol. 581, pp. 133-138.

Stead, J. A., Keen, J. N. & McDowall, K. J. (2006). The identification of nucleic acid-interacting proteins using a simple proteomics-based approach that directly incorporates the electrophoretic mobility shift assay. *Molecular & Cellular Proteomics* 5, 1697-1702.

Suzuki, S. & Henderson, P. J. F. (2006). The hydantoin transport protein from *Microbacterium liquefaciens*. *J. Bacteriol.* 188, 3329-3336.

Taylor, D. R. & Hooper, N. M. (2006). The prion protein and lipid rafts. *Mol. Membr. Biol.* 23, 89-99.

Tezuka-Kawakami, T., Gell, C., Brockwell, D. J., Radford, S. E. & Smith, D. A. (2006). Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys. J.* 91, L42-L44.

Thirumurugan, K., Sakamoto, T., Hammer, J. A., Sellers, J. R. & Knight, P. (2006). The cargo-binding domain regulates structure and activity of myosin 5. *Nature* 442, 212-215.

Thomas, D. A., Francis, P., Smith, C., Ratcliffe, S., Ede, N. J., Kay, C., Wayne, G., Martin, S. L., Moore, K., Amour, A. & Hooper, N. M. (2006). A broad-spectrum fluorescence-based peptide library for the rapid identification of protease substrates. *Proteomics* 6, 2112-2120.

Tuthill, T. J., Bubeck, D., Rowlands, D. J. & Hogle, J. M. (2006). Characterization of early steps in the poliovirus infection process: Receptor-decorated liposomes induce conversion of the virus to membrane-anchored entry-intermediate particles. *J. Virol.* 80, 172-180.

Unwin, R. D., Smith, D. L., Blinco, D., Wilson, C. L., Miller, C. J., Evans, C. A., Jaworska, E., Baldwin, S. A., Barnes, K., Pierce, A., Spooncer, E. & Whetton, A. D. (2006). Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells. *Blood* 107, 4687-4694.

Vohra, R. S., Murphy, J. E., Walker, J. H., Ponnambalam, S. & Homer-Vanniasinkam, S. (2006). Atherosclerosis and the lectin-like oxidized low-density lipoprotein scavenger receptor. *Trends in Cardiovascular Medicine* 16, 60-64.

Wallace, E. J., Hooper, N. M. & Olmstedz, P. D. (2006). Effect of hydrophobic mismatch on phase behavior of lipid membranes. *Biophys. J.* 90, 4104-4118.

West, D. K., Brockwell, D. J., Olmsted, P. D., Radford, S. E. & Paci, E. (2006). Mechanical resistance of proteins explained using simple molecular models. *Biophys. J.* 90, 287-297.

West, D. K., Brockwell, D. J. & Paci, E. (2006). Prediction of the translocation kinetics of a protein from its mechanical properties. *Biophys. J.* 91, L51-L53.

West, D. K., Olmsted, P. D. & Paci, E. (2006). Free energy for protein folding from nonequilibrium simulations using the Jarzynski equality. *Journal of Chemical Physics* 125.

West, D. K., Olmsted, P. D. & Paci, E. (2006). Mechanical unfolding revisited through a simple but realistic model. *Journal of Chemical Physics* 124.

West, D. K., Paci, E. & Olmsted, P. D. (2006). Internal protein dynamics shifts the distance to the mechanical transition state. *Phys. Rev. E* 74.

Wijma, H. J., Jeuken, L. J. C., Verbeet, M. P., Armstrong, F. A. & Canters, G. W. (2006). A random-sequential mechanism for nitrite binding and active site reduction in copper-containing nitrite reductase. *J. Biol. Chem.* 281, 16340-16346.

Williams, G. J., Woodhall, T., Farnsworth, L. M., Nelson, A. & Berry, A. (2006). Creation of a pair of stereochemically complementary biocatalysts. *J. Am. Chem. Soc.* 128, 16238-16247.

Wilson, A. J., van Gestel, J., Sijbesma, R. P. & Meijer, E. W. (2006). Amplification of chirality in benzene tricarboxamide helical supramolecular polymers. *Chem. Commun.*, 4404-4406.

You, J. H., Reed, M. L., Dove, B. K. & Hiscox, J. A. (2006). Three-dimensional reconstruction of the nucleolus using meta-confocal microscopy in cells expressing the coronavirus nucleoprotein. In *Nidoviruses: Toward Control of Sars and Other Nidovirus Diseases*, Vol. 581, pp. 313-318.

Zhang, J., Smith, K. M., Tackaberry, T., Sun, X. J., Carpenter, P., Slugoski, M. D., Robins, M. J., Nielsen, L. P. C., Nowak, I., Baldwin, S. A., Young, J. D. & Cass, C. E. (2006). Characterization of the transport mechanism and permeant binding profile of the uridine permease Fui1p of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 281, 28210-28221.

Zhang, J., Sun, X. J., Smith, K. M., Visser, F., Carpenter, P., Barron, G., Peng, Y. S., Robins, M. J., Baldwin, S. A., Young, J. D. & Cass, C. E. (2006). Studies of nucleoside transporters using novel autofluorescent nucleoside probes. *Biochemistry* 45, 1087-1098.

Zhang, J., Tackaberry, T., Ritzel, M. W. L., Raborn, T., Barron, G., Baldwin, S. A., Young, J. D. & Cass, C. E. (2006). Cysteine-accessibility analysis of transmembrane domains 11-13 of human concentrative nucleoside transporter 3. *Biochem. J.* 394, 389-398.

Zsoldos, Z., Reid, D., Simon, A., Sadjad, B. S. & Johnson, A. P. (2006). eHITS: An innovative approach to the docking and scoring function problems. *Current Protein & Peptide Science* 7, 421-435.

UNIVERSITY OF LEEDS

Leeds, United Kingdom
LS2 9JT
Tel. 0113 243 1751
www.leeds.ac.uk