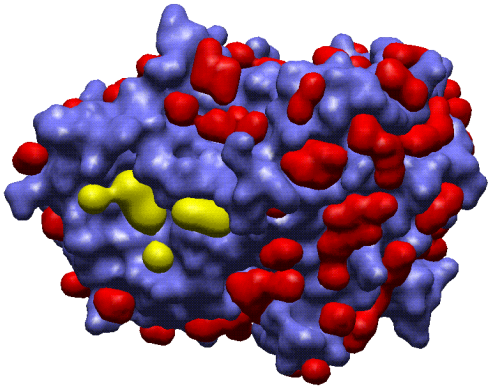
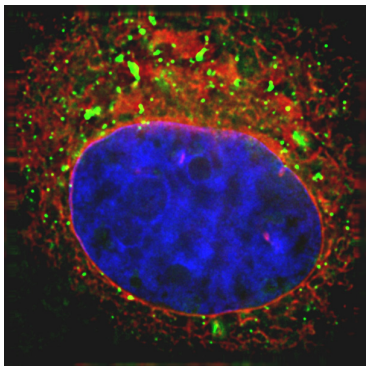
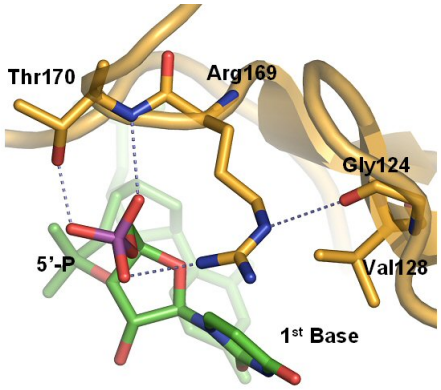
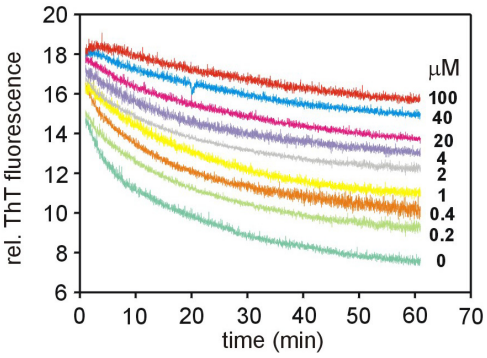
# Astbury Centre for Structural Molecular Biology

## University of Leeds








# Annual Report 2005

***Front cover illustration:*** A collage of pictures illustrating the work of the Astbury Centre. Upper left: Predicting protein-protein interfaces. Probes (red and yellow) fill the identified clefts on the protein surface of Ran (blue). Those clefts that are occupied during the interaction with nuclear transport factor-2 are coloured yellow; these are more easily desolvated than many of the other clefts on the protein surface (see page 34); Upper right: Visualisation of an NS5A-EGFP fusion protein (green fluorescence) in a human hepatoma cell harbouring an autonomously replicating HCV sub-genomic replicon. Replication complexes are evident adjacent to the ER (stained with concanavalin A: red) but distinct from the ER. Cell counterstained with Hoechst (nucleus: blue) (see page 15). Image taken in the Wellcome Trust Bioimaging suite within the Astbury building; Lower left: **.** The 5' monophosphate-binding pocket of *E. coli* RNaseE. The phosphate is engaged by a semicircular ring of hydrogen-bonding donors from the side chain and peptide amide of Thr 170 and the guanidino group of Arg 169. The interaction of Arg 169 is consolidated by a hydrogen bond to the peptide backbone of Gly 124 in the neighbouring strand (see page 44); Lower right: The time-course of depolymerisation of $\beta_2$m amyloid fibrils at neutral pH is significantly reduced by the addition of heparin, as monitored by the fluorescent dye Thioflavin-T. (see page 80).

**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, and privilege, to write this Introduction to the Annual Report on the activities within the Astbury Centre once again. 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 Vice-Chancellor's drive to ensure that Leeds remains a research-led institution at the forefront of research has led to the creation of designated Centres of Research Excellence and I'm glad to report that we were selected as one of these during 2005. This formalises the University's commitment to inter-disciplinary working across our constituent Schools and Faculties, and recognises our unique character that other institutions are only just beginning to try and mimic. In order to fulfil our part of this bargain we have also been busy, recruiting an outstanding external Scientific Advisory Board consisting of world leading authorities in their respective fields whom I look forward to introducing to you all in these pages next year. We very much look forward to their honest assessment of the qualities of our outputs and their suggestions for future developments.

Our internal organisation is also changing. The two new University Research Institutes I told you about last year, Molecular Biophysics (IMB) headed by Alastair Smith and Sheena Radford, and the Institute of Bionanosciences (IoB) headed by myself, are being rolled formally into the Astbury management structure, as are the new developments in Chemical Biology headed by Adam Nelson. These will form three sub-themes for our research, joining Structural Molecular Biology which of course has long been established. With these structural/administrative changes in hand we are now focusing on creating a greater sense of community across the various staff at all levels who work in the Centre. Our postdoctoral and postgraduate colleagues have formed an Astbury Society and are busy organising social events, and contributing to management and our research activities. We are also just beginning a series of research retreats allowing our staff to meet their colleagues from across campus and hear about "Astbury" science from outside their direct area of expertise. We hope that these activities will act as further spurs to inter-disciplinary working.

Astbury members continue to be very successful in raising external grant income. A special mention here must go the bid to the UK BBSRC for PhD studentships under the auspices of a Departmental Training Award. This was funded at a level of up to 12 4-year studentships a year – a fantastic result! As with our previous inter-disciplinary bid for a Wellcome Trust JIF Centre this result shows the power of integrated non-traditional research groupings. I would like to thank Alan Berry and Sheena Radford particularly for the efforts they put in during the year in the production of a highly professional application, assisted to various degrees by Stuart Warriner and Simon Phillips.

The pages that follow describe some of the highlights of our work over the last year. These reports have largely been written by our younger researchers. Their tremendous enthusiasm for this kind of interdisciplinary work augurs well for our future. As always I am particularly struck by the breadth of activity in the Centre, ranging from the sophisticated applications of synthetic organic chemistry to the developments in single molecule biophysics. In between these extremes you will find groundbreaking activity in many traditional areas for structural biology. The Astbury Centre has always been outward looking and this tradition continues with the many external collaborations acknowledged in these pages, from both within the UK and beyond. We would welcome discussions with anyone wishing to collaborate or simply to make use of our facilities, the details of which can be found via our web page (http://www.astbury.leeds.ac.uk). These brief summaries, however, only scratch the surface of the work of the Centre. I hope you enjoy reading them, and if you wish to learn more

please visit our website or contact the Director. The Centre also continues to host a very successful seminar programme that illustrates aspects of work within the Centre.

Finally, I must bring you up to date with some changes in personnel. Steve Homans has been appointed as Head of one of the new Institutes within the Faculty of Biological Sciences and following the precedent established by Simon Phillips immediately stepped down from his Deputy position in the Centre. I'd like to take this opportunity to congratulate Steve and thank him for all the hard work he has put into helping the Centre get established and the various things he did to support me as Director. We look forward to working closely with him in his new role. I'm delighted to announce that Adam Nelson has agreed to be our new Deputy Director and he took on this position in the middle of 2005. We intend to establish a new precedent with Adam's appointment, namely to have a Director and Deputy from our different constituent Faculties to emphasise our commitment to inter-disciplinary working over the long term. I also need to thank Rachael Taylor, who provided administrative support to the Centre on a temporary basis after Jenny Gilmartin was promoted to head the FBS Research Office. We wish her well in her new role in the Leeds Institute of Genetics, Health and Therapy. Her activities have now been taken over by Mrs Donna Fletcher who has taken on the role of Administrative Assistant to the Director after working in the engineering departments. Donna has already made her mark by producing a regular Newsletter describing Centre 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 14MB PDF document that can be downloaded from our web site.

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

Peter G. Stockley
*Director, Astbury Centre for Structural Molecular Biology*
*Leeds, May 2006*

# Contents

## Contributions indexed by Astbury Centre Principal Investigator

# Mass spectrometry facility

Victoria L. Homer and Alison E. Ashcroft

**Overview of facility**

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

The Mass Spectrometry Facility runs an analytical service as well as being actively involved in several research areas within the Astbury Centre for Structural Molecular Biology and the Faculty of Biological Sciences, and also with other groups and external collaborators.

**Research**

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

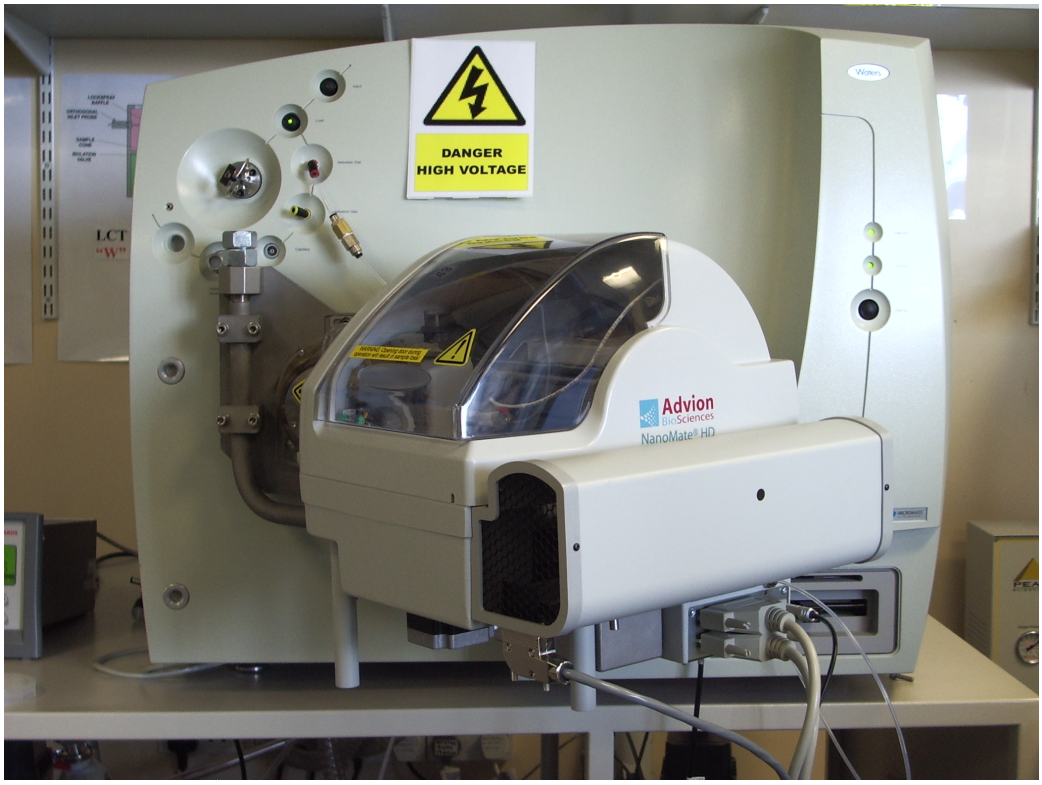
**i). Protein folding.** Protein folding is an intriguing area of biochemistry and protein mis-folding is thought to be a contributing factor to several diseases. Working with Prof. Sheena Radford's group, ESI-MS is being used to investigate $\beta_2$-microglobulin conformation using charge state distribution 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 $\beta_2$-microglobulin amyloid studies with Prof. Sheena Radford.

**iii). Reaction monitoring.** We use mass spectrometry to measure the uptake of ATP by the muscle protein myosin (with Prof. Howard White, Eastern Virginia Medical School, USA). Myosin has a motor domain that interacts with filaments of F-actin and splits ATP to generate force and movement. In a collaborative project with Prof. Gabriel Waksman and Dr Han Remaut of Birkbeck College, London, we are studying Saf fibre formation 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 an investigation into the functional analysis of preproneuropeptide genes from the Drosophila genome with Prof. Elwyn Isaac, and studies into the peroxisomal protein import machinery with Dr Alison Baker.

**v). Mass spectrometry method development.** To further our structural molecular biology studies, we are investigating the use of Ion Mobility Spectrometry coupled to mass spectrometry as a potential method to separate co-populated protein conformers and for the structural characterisation of small peptides.



The new LCT Premier mass spectrometer equipped with a NanoMate sample delivery and nano-electrospray ionisation interface. The LCT Premier has a range of m/z 60,000 in addition to collisional cooling facilities for the analysis of non-covalently bound macromolecular complexes.

## Publications

Ashcroft, A.E. (2005) Recent developments in electrospray ionisation mass spectrometry: noncovalently bound protein complexes, *Natural Products Reports,* **22**, 452-464.

Khan, A., Ashcroft, A.E., Higenell, V., Korchazhkina, O.V. & Exley, C. (2005) Metals accelerate the formation and direct the structure of amyloid fibrils of NAC, *Journal of Inorganic Biochemistry,* **99,** 1920-1927.

Ashcroft, A.E. Lago, H. Macedo, J.M.B. Horn, W.T. Stonehouse, N.J. & Stockley P. G. (2005) Engineering thermal stability in RNA phage capsids via disulphide bonds, *J. Nanosci. & Nanotechnol.,* **5,** 2034-2041.

Stockley, P.G., Ashcroft, A.E., Francese, S., Thompson, G.S., Ranson, N., Smith, A.M., Homans, S.W. & Stonehouse, N.J. (2005) Dissecting the fine details of assembly of a *T*=3 capsid, *J. Theor. Med.,* **6**, 119-125.

## Funding

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

# Functional characterisation of the novel human equilibrative nucleoside transporter (hENT3) located in intracellular membranes

Kay Barnes, Ralph Hyde, Sophie Foppolo and Stephen A. Baldwin

## Introduction

Both hENT3 and mENT3 are members of the equilibrative nucleoside transporter family. These transporters play key roles in the uptake of precursors for nucleotide synthesis by salvage pathways in a number of cell types, for example, bone marrow and brain. They also regulate the concentration of adenosine available to cell surface purinoreceptors, thus influencing coronary blood flow, inflammation and neurotransmission. Family members are predicted to share a common topology of 11 transmembrane (TM) α-helices, with a cytoplasmic N-terminus and a large cytoplasmic loop linking TM6 and TM7. Unlike other family members, hENT3 has a long (51 residue) hydrophilic N-terminal region preceding TM1 that possesses a putative dileucine-based endosomal/lysosomal targeting motif (DE)XXXL(LI). We propose that hENT3, unlike other family members located at the cell surface, functions intracellularly.

## Subcellular distribution of hENT3

To examine the role of the dileucine motif in the subcellular distribution of the transporter, a comparison was made of green fluorescent protein (GFP) fusion proteins bearing wild type hENT3 (GFP-hENT3), hENT3 lacking the first 36 residues of the N-terminal region (GFP-hENT3ΔN), or hENT3 in which the dileucine motif at positions 31 and 32 had been replaced by alanine residues (GFP-hENT3AA)(Fig.1).

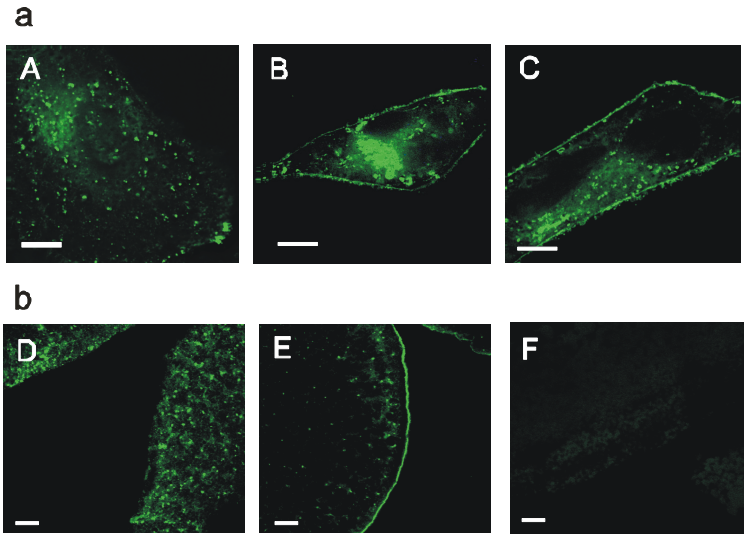



**Fig. 1 Role of the N-terminal dileucine motif in determining the subcellular distribution of hENT3** A-C, distribution of fluorescence in HeLa cells expressing GFP fusion proteins bearing the wild-type hENT3 (GFP-hENT3) (A), hENT3 lacking the first 36 residues of the N-terminal region (GFP-hENT3ΔN) (B), or hENT3 in which the dileucine motif at positions 31 and 32 had been replaced by alanine residues (GFP-hENT3AA) (C). Each image corresponds to one representative deconvolved optical section. Scale bars 10 μm. D-F, distribution of fluorescence in cryosections of oocytes injected with RNA transcripts encoding GFP-hENT3 (D) or GFP-hENT3AA (E) or injected with water alone (F). Scale bars, 20 μm.

The GFP protein-tagged forms of the full-length hENT3 protein were found to be predominantly intracellular proteins that co-localised, in part, with lysosomal markers in cultured human HeLa cells. Truncation of the hydrophilic N-terminal region or mutation of the dileucine motif to alanine caused the protein to be relocated to the cell surface in human cells and in *Xenopus* oocytes.

## Characterisation of transport activity

By using mRNA transcripts of the dileucine mutant hENT3AA injected into *Xenopus* oocytes, the protein was shown to be a low affinity sodium-ion independent nucleoside transporter that could also transport adenine. Its activity was strongly dependent upon pH, and the optimum pH value of 5.5 probably reflected the location of the transporter in acidic, intracellular compartments (Fig. 2).

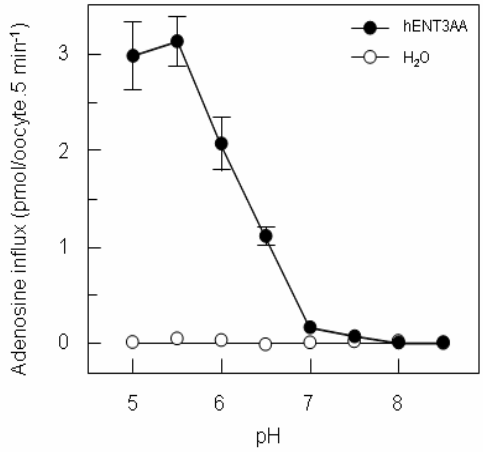



**Fig. 2 pH dependence of hENT3-mediated adenosine transport.** Uptake of $^{14}$C-labelled adenosine (20 μM, 20 ºC, 5 min) in oocytes injected with the hENT3AA RNA transcripts or water alone was measured in transport medium containing 100 mM sodium chloride and buffered at pH values ranging from 5.0 to 8.5.

## Collaborators

We thank Sylvia Y.M.Yao, Amy M.L. Ng, Mabel W.L. Ritzel and James D.Young of the Department of Physiology, University of Alberta and Carol E. Cass of the Department of Oncology, University of Alberta and the Cross Cancer Institute, Edmonton, Alberta T6G 2H7, Canada.

## Publications

Baldwin, S.A., Yao, S.Y.M., Hyde, R.J., Ng, A.M.L., Foppolo, S., Barnes, K., Ritzel, M.W.L., Cass, C.E. and Young, J.D. (2005) Functional characterisation of novel human and mouse equilibrative nucleoside transporters (hENT3 and mENT3) located in intracellular membranes. *J. Biol. Chem*. **280**, 15880-15887.

## Funding

We gratefully acknowledge the British Heart Foundation, Medical Research Council, the Wellcome trust, Science Research Investment Fund, UK and the Canadian Institutes of Health and Research and the Alberta Cancer board for financial support.

# Chronic myeloid leukaemia: the role of Bcr-Abl-induced abnormalities in glucose transport regulation.

Kay Barnes, Elizabeth McIntosh and Stephen A. Baldwin

## Introduction

In chronic myeloid leukaemia (CML) expression of the chimeric tyrosine kinase, Bcr-Abl, promotes the inappropriate survival of haemopoietic stem cells by a non-autocrine mechanism in the absence of the growth factor, IL-3. Stimulation of glucose uptake appears to play an important role in the suppression of apoptosis by this cytokine in normal haemopoietic cells. To investigate whether the cell survival mechanisms mediated by the oncogene and cytokine showed any similarities, we employed a haemopoietic cell line, TonB210, engineered for tetracycline-inducible expression of Bcr-Abl.

## Role of the glucose transporter, GLUT1, in glucose transport regulation in TonB210 cells

Bcr-Abl tyrosine kinase expression in cytokine-deprived cells was found to mimic the effect of IL-3 in maintaining a higher $V_{max}$ for hexose uptake (Table 1).

**Table 1**. Kinetic analysis of 2-deoxy-D-glucose transport in TonB210 cells

| | Treatment | $K_m$ (mM) | $V_{max}$ (nmol/min/$10^6$cells) |
|---|---|---|---|
| Non-induced, IL-3-dependent cells | IL-3 (10 ng/ml) 1 h | 1.7 ± 0.4 (3) | 5.8 ± 1.7 (3) |
| | No IL-3 | 1.9 ± 0.4 (3) | 2.7 ± 0.2 (3) |
| Bcr-Abl-induced cells | Glivec (10 μM) 6 h | 1.4 ± 0.7 (4) | 1.1 ± 0.4 (4) |
| | No Glivec | 2.2 ± 0.6 (4) | 4.8 ± 1.0 (4) |

Transporter-mediated 2-deoxy-D-glucose uptake was measured over a range of concentrations following the indicated treatments of cells. The kinetic parameters of transport are shown as mean ± SEM (n). n = number of separate experiments, within which each uptake measurement was performed in triplicate.

In both IL-3-treated cells and those expressing Bcr-Abl, high rates of hexose uptake were associated with the retention at the cell surface of approximately 80% of the total cellular content of the GLUT1 glucose transporter (Fig. 1). In contrast, treatment of Bcr-Abl-expressing cells for 6 h with the Bcr-Abl kinase inhibitor Glivec (10 μM), in the absence of IL-3, led to the internalization of approximately 90% of the cell-surface transporters (Fig.1) and decreased the $V_{max}$ for hexose uptake more than 4-fold without significant effect on the $K_m$ (Table 1) for this process or on the total cellular transporter content. These effects were not the result of any loss in cell viability. Both IL-3 treatment and expression of Bcr-Abl led to enhanced phosphorylation of Akt (protein kinase B).

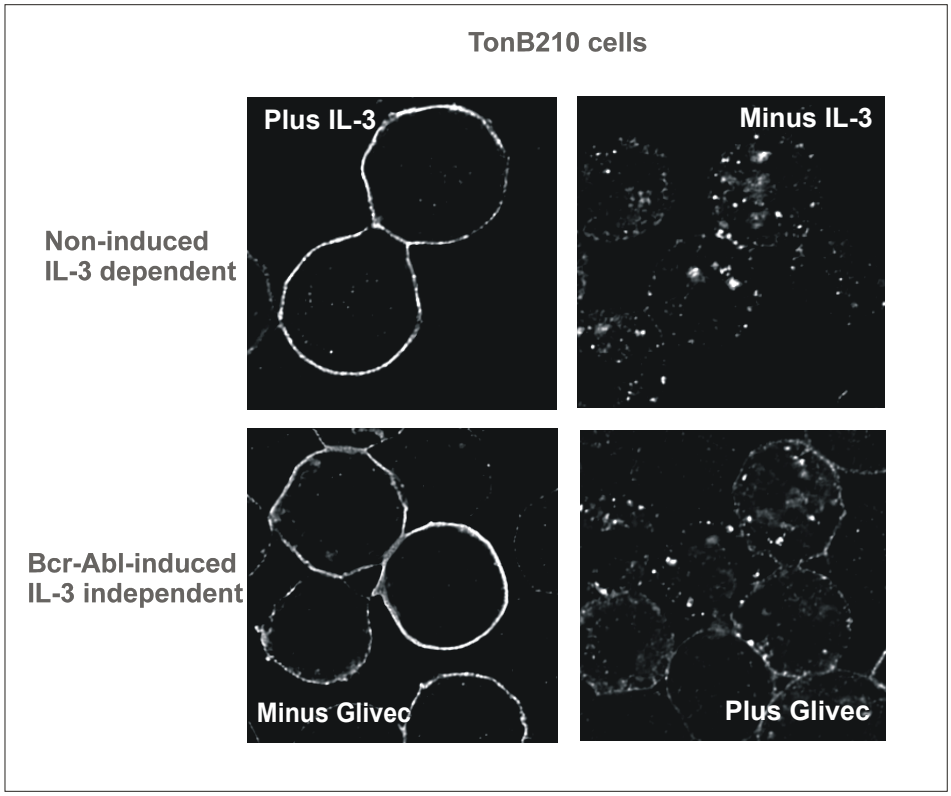



**Fig. 1** (A) Effect of IL-3 on the subcellular distribution of GLUT1 in TonB210 cells. Cells were washed for 3 h at 37 ºC to remove IL-3-conditioned culture medium. IL-3-depleted cells were incubated for 1 h with or without 10 ng/ml IL-3. (B) Induced, cytokine-deprived cells expressing Bcr-Abl were suspended in serum-free medium containing doxycycline (1 μg/ml). The cells were either left untreated or incubated with 10 μM Glivec for 6 h.

We have shown, therefore, that inhibition of glucose transport may play an important role in the therapeutic action of Glivec, and that the signal transduction pathways involved in transport stimulation by Bcr-Abl may offer novel therapeutic targets for CML.

**Collaborators** We thank Anthony D. Whetton (University of Manchester, UK), George Q. Daley (Whitehead Institute, Cambridge, MA, USA) and Johanne Bentley (Cancer Research UK Clinical Centre, St James's University Hospital, Leeds, UK) for their contributions to this research.

**Publications**

Barnes K., Mcintosh E., Whetton A.D., Daley, G.Q., Bentley J. & Baldwin S.A. (2005) Chronic myeloid leukaemia: an investigation into the role of Bcr-Abl-induced abnormalities in glucose transport regulation Oncogene, **24**, 3257-3267.

**Funding**

We thank Yorkshire Cancer Research, the Wellcome Trust and the Leukaemia Research Fund for financial support.

# Directed evolution of enzymes with synthetically useful activities

Matthew Edmundson, Lorna Farnsworth, Bernardo Perez-Zamorano,
Adam Nelson and Alan Berry

## Enzymes in synthesis

Enzymes are very active catalysts that exhibit remarkable regio-, chemo- and stereo-selectivity, but unfortunately the use of enzymes in synthesis is currently very limited. This is mainly due to their inherent substrate specificity and the fact that complementary enzymes, which provided access to all possible stereoisomers, are not always available. Directed evolution is being used to tackle these problems and vastly to increase the potential use of enzymes in organic synthesis

## Directed evolution

Directed evolution is very similar to "natural" evolution; in both cases selection pressures act on a diverse population to determine which individuals survive. While the selection pressures in natural evolution are random and determined by the environment in which an organism lives, the pressures in directed evolution are deliberately chosen by the experimenter via the screening assay, allowing those mutants with the desired activity to be selected and taken through further rounds of mutation and screening, thereby increasing the level of activity. The screening assay is thus vital to the success of a project. We are using directed evolution approaches to alter the properties of a range of enzymes as summarised below.

## Aldolases

Carbon-carbon bond formation is of great importance to organic chemists as they strive to generate molecules of increasingly complexity. One of the routes used by Nature to form new carbon-carbon bonds is the aldol condensation performed by enzymes called aldolases. One area of research focuses on sialic acid aldolase, an enzyme which catalyses the reversible condensation of pyruvate and *N*-acetyl-*D*-mannosamine to form *N*-acetyl-*D*-neuraminic acid (a sialic acid) (Fig. 1).

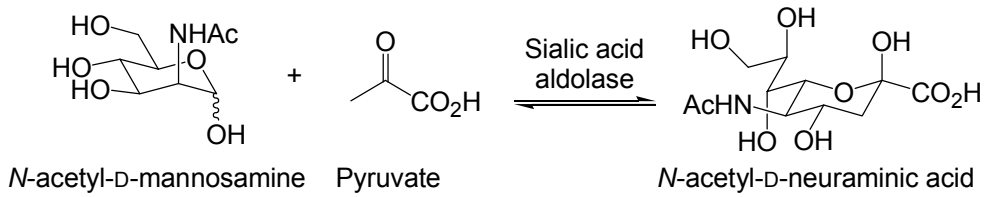



**Fig. 1** The sialic acid aldolase catalysed reaction of pyruvate and *N*-acetyl-*D*-mannosamine to form *N*-acetyl-*D*-neuraminic acid.

Analogues of sialic acid, such as Relenza™ (Fig. 2) and Tamiflu™, are sialidase inhibitors currently used in the treatment of influenza. These molecules are complex and difficult to synthesise, but we were able to show that directed evolution of sialic acid aldolase could be used to produce enzymes capable of the synthesis of sialic acid analogues.

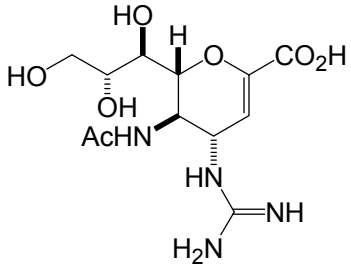



**Fig. 2** Relenza, a commercially availiable inhibitor of influenza, is an analogue of sialic acid.

The wild-type sialic acid aldolase shows little stereochemical discrimination for the configuration produced at C-4 during the reaction. We have used a combination of error-prone PCR and saturation mutagenesis to produce two

complementary enzymes for the specific synthesis of either 4-*R* or 4-*S* analogues of sialic acid. These enzymes have been identified, expressed, purified and used in the synthesis of a wide variety of tertiary amide substrates (Fig.3) all with excellent stereochemical control.

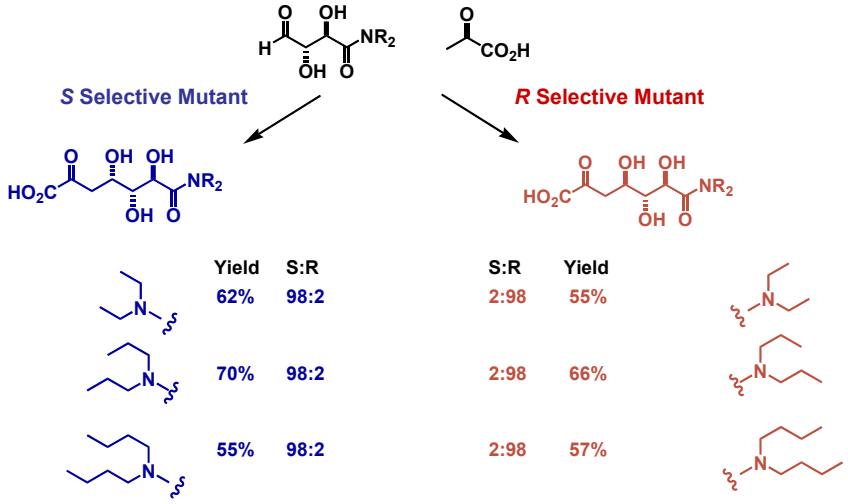



**Fig. 3.** High diastereoselectivity, greater than 98:2, and good yields were achieved in all cases, when the mutant enzymes were reacted with a variety of tertiary amides.

The same approach has been used to identify variants of sialic acid aldolase capable of the synthesis of a range of secondary amides. A number of active mutants have been identified and are currently being fully characterised.

**Obtaining D-ASA activity in dihydrodipicolinate synthase (DHDPS) by random mutagenesis**

Dihydrodipicolinate synthase also catalyses an aldol condensation reaction and we have been using directed evolution to produce an enzyme that would function with D-aspartate semialddehyde (D-ASA), the enantiomer of its natural substrate, L-ASA. A randomly evolved DHDPS library was screened for activity with D-ASA, and after screening approximately 500 clones, a promising candidate was found and sequenced. The protein encoded by this mutant bears the mutation N93K. (Fig. 4)

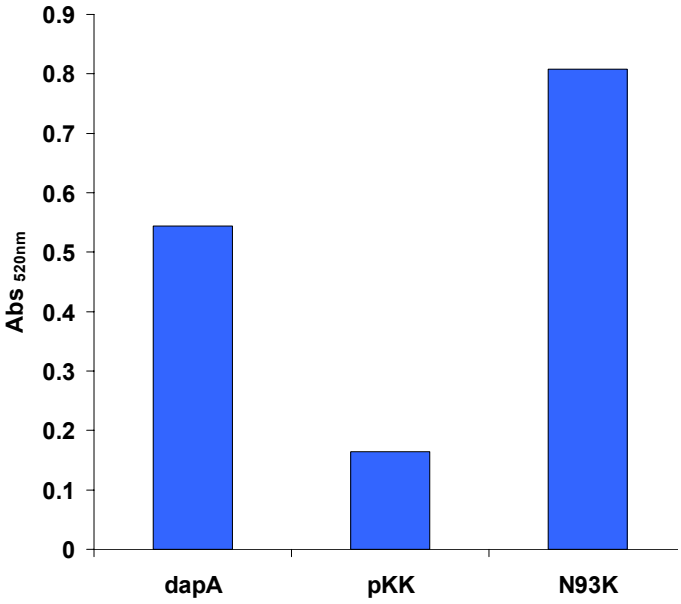



**Fig. 4.** $Abs_{520nm}$ signal in the o-aminobenzaldehyde colorimetric assay of the N93K mutant in comparison to the wild type gene (*dap*A) and the empty vector (pKK). The production of the purple cromophore in this assay is directly linked to the presence of DHDPS activity; and can be registered by spectrophotometry at $Abs_{520nm}$ . dapA= WT DHDPS, pKK= cells transformed with an pKK223-3

The N93K mutant showed a 45% increase in the activity as determined in the colorimetric assay for DHDPS. The N93K mutation is located at the beginning of the loop between α

helix 3 and β strand 4; on the external surface of the tetramer and is distant from the catalytic site of the enzyme. Further characterisation of the enzyme is underway to determine how its mutation produces the change in specificity.

**Additional work**

Work is also being carried out into the development of screening assays for other enzymes, with the focus being on linking the reaction being studied to a chromogenic screen via a cascade of coupling reactions. It is hoped that this type of screen can be applied directly to colonies on an agar plate, thereby allowing desirable mutants to be spotted immediately, greatly increasing the process of directed evolution.

**Publications**

Williams, G.J., Woodhall, T., Nelson, A. & Berry, A. (2005) Structure-guided saturation mutagenesis of *N*-acetylneuraminic acid lyase for the synthesis of sialic acid mimetics. *Protein Eng. Des. Sel.*, **18**, 239-246.

Woodhall, T., Williams, G.J., Berry, A. & Nelson, A. (2005) Synthesis of screening substrates for the directed evolution of sialic acid aldolase: Towards tailored enzymes for the preparation of influenza A inhibitor analogues. *Org. Biomol. Chem*, **3**, 1795-1800.

Woodhall, T., Williams, G.J., Berry, A. & Nelson, A.S. (2005) Directed evolution of an aldolase for the parallel synthesis of sialic acid mimetics. *Ang. Chem. Int Ed*, **44**, 2109-2112

**Funding**

We gratefully acknowledge the BBSRC, EPSRC and The Wellcome Trust for financial support.

# Conformation dynamics and catalysis of aldolase studied by NMR

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

## Introduction

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

## Backbone amide resonance assignment

FBP-aldolase forms a 78kDa dimer and as such represents a significant challenge to study by NMR. To date there only is one example of a larger polypeptide chain that has been fully assigned. A combination of HNCA, HN(CO)CA, HNCO, HN(CA)CO and HSQC TROSY-based assignment experiments using an isotopically enriched [$^{2}$H, $^{15}$N, $^{13}$C]-FBP-aldolase sample have been completed (Fig. 1). To assist backbone amide resonance assignment a number of selectively labeled ([$^{15}$N-Lys]-FBP-aldolase, [$^{15}$N-Glu]-FBP-aldolase, [$^{15}$N-Tyr]-FBP-aldolase, [$^{15}$N-Val]-FBP-aldolase, [$^{15}$N-Ala]-FBP-aldolase, [$^{15}$N-Leu]-FBP-aldolase and [$^{15}$N-Phe]-FBP-aldolase) samples have been produced. This has resulted in 35% of the backbone amide resonances being assigned to date, which includes residues from the major β5-α7 loop flexible loop.

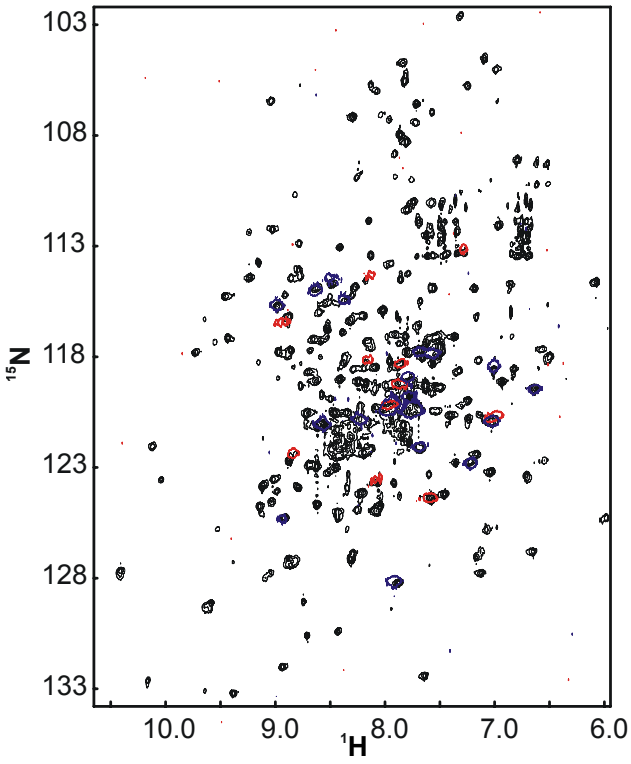



**Fig 1.** HSQC spectra of the amide resonances in [$^{2}$H, $^{15}$N, $^{13}$C]-FBP-aldolase measured at 900MHz are shown in black. Overlaying resonances from [$^{15}$N-Lys]-FBP-aldolase and [$^{15}$N-Tyr]-FBP-aldolase are shown in blue and red respectively (measured at 750MHz).

At present there are no assigned resonances from the highly solvent shielded core of the protein, but unfolding/refolding conditions have been developed which will allow complete hydrogen exchange of all backbone amide resonances. This will allow the backbone assignment to be completed.

**Characterization of nanosecond motions**

T1, T2, and hetero-nuclear NOE relaxation rates have been measured and, for assigned residues, nanosecond motions have been quantified using the Lipari-Szabo model-free formalism. This has confirmed the increased mobility of the β5-α7 loop flexible loop with respect to other more stable regions of protein's architecture. Relaxation data has also been obtained in presence of a substrate analogue and inhibitor phosphoglycolohydroxamate (PGH), resulting in significant changes in both chemical shift and relaxation times (Fig. 2). A range of site directed mutations have been produced to characterize residues whose amide resonances display large chemical shift changes upon inhibitor binding.

Relaxation data will also be obtained for natural substrate complexes, dihydroxyacetone phosphate, glyceraldyhde phosphate, and FBP. This coupled to a full assignment will allow a complete description of enzyme motions during the catalytic cycle.

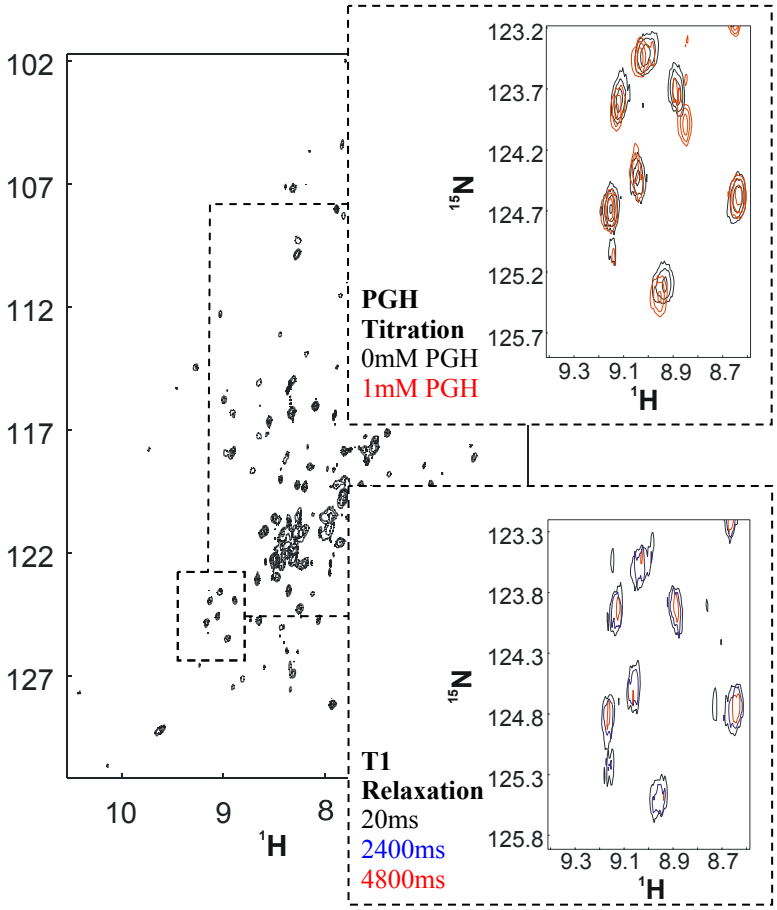



**Fig 2.** Example region of [$^2$H, $^{15}$N, $^{13}$C]-FBP-aldolase HSQC spectrum which shows, top, changes in amide resonance chemical shift in the presence and absence of PGH and, bottom, differing longitudinal (T1) relaxation rates for the uncomplexed form. All data was collected at 750MHz.

**Funding**

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

# Structural studies of thermostable FBP-aldolases created by directed evolution.

Liz Bennett, Chi Trinh, Alan Berry and Simon Phillips

## Introduction

The catalytic efficiency of enzymes has long been coveted by synthetic chemists keen to exploit the exquisite specificity and selectivity of these natural molecules. Enzymes exhibit stereo-, regio- and enatio-selective integrity throughout substrate binding and catalysis, but are often limited in their applicability beyond the natural environment by their poor stability at the high temperatures, pressures and solvent concentrations commonly employed in synthetic chemistry. However thermophilic organisms do exist naturally, functioning at temperatures exceeding 60°C (thermophilic), or even 90°C (hyperthermophilic). Most mesophilic-derived enzymes lose activity at temperatures above 40°. Many studies have attempted to rationalise stability and functionality under extreme conditions by comparing factors such as sequence, structural motifs, hydrophobic packing, loop length, salt bridges and residue bias. Conclusions generally point to no single factor alone being culpable. Comparison of natural thermophiles and mesophiles requires like enzymes from different species to be compared and must therefore take into account the underlying organismal variations. By evolving a novel thermophilic enzyme from high sequence identity mesophilic "parents" (>87% sequence identity) this variable baseline is reduced, and evolutionary steps towards thermostability may be revealed.

Fructose-1,6-bisphosphate aldolase catalyses the formation of a carbon-carbon bond during the reversible condensation of glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP) to form fructose 1,6-bisphosphate (FBP). Carbon-carbon bond formation is a prime tool in synthesis and this, together with the ubiquitous α/ß-barrel fold displayed by Class II FBP-aldolases, makes the enzyme an attractive target for directed evolution with the aim of improving functional thermostabilty. Four successive generations of FBP-aldolase variants were created by family-DNA shuffling and error-prone PCR. Each round of mutagenesis produced enzymes with increased thermostability and the best were carried through as parents of the next generation. The fourth generation mutant, 4-43D6, exhibited a 90-fold increase in half-life at 53°C and a $T_{50}$ increased by 12.2°C over the wild-type *E.coli* parent enzyme.

## X-ray crystallography

Crystallisation conditions for the four variants (the best of each generation) were identified through extensive screening of conditions and viable protein crystals obtained for each (Fig. 1). Cryoconditions for each crystal were also optimised, thus allowing for data collection at 100K. Diffraction data were collected at Daresbury SRS resulting in the four enzyme structures being solved to between 1.8 and 2.0Å.

Sequence data has already shown that there are no extra disulphide bridges, and the overall chain length varies only by one residue in relation to the *E.coli* parent due to a deletion of A82 in the *Ed.ictaluri* sequence; a deletion which is carried through to the fourth generation.

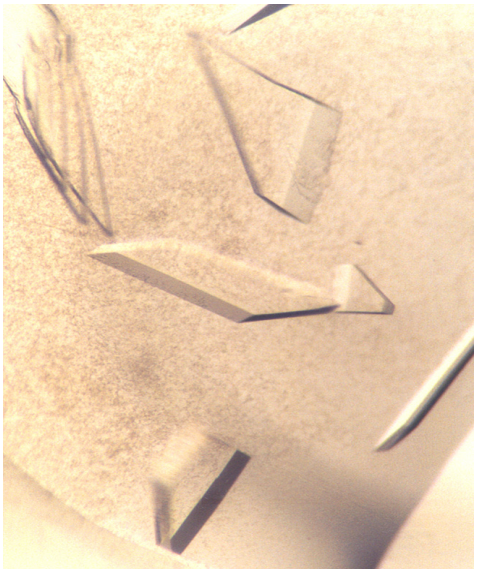


**Fig. 1** Crystals of third generation mutant, 3-4C10 grown in 0.05M $KH_2PO_4$, 17% PEG8000.

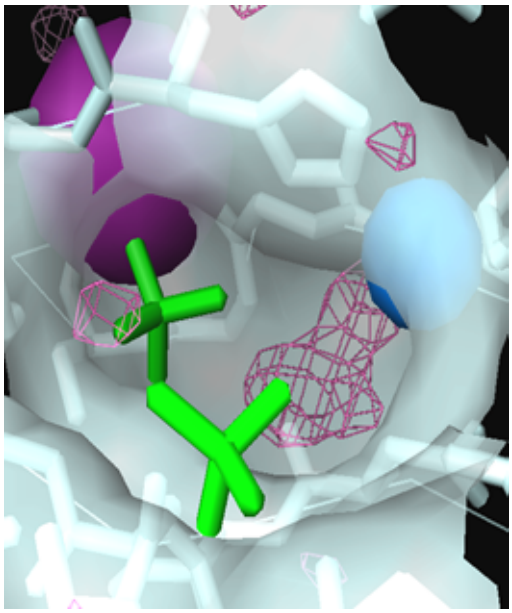


**Fig.2** The active site – the catalytic zinc ion is shown as a blue sphere with the postulated glycerol electron density adjacent in magenta chicken wire. Phosphate ions are shown in green and potassium is the large purple sphere at the back.

Preliminary comparisons of the structures of the four generations of thermostable enzymes show that tertiary and quaternary structure is maintained throughout the generations. However there appears to be slight variation in secondary structure which might result in altered packing and hence might affect thermostability. It was also noted that as well as binding two phosphates per monomer from the mother liquor into the active site (Fig. 2) (probably mimicking the phosphates in the natural substrate); each active site has a patch of, as yet unassigned, positive electron density adjacent to the zincs and between the phosphates. This density has been tentatively modelled as a glycerol molecule, sequestered from the cryoprotectant. This would correspond with a 3-carbon fragment of the backbone of the target substrate, FBP. Further structural analysis and comparative studies are underway.

**Funding**
This work was funded by the BBSRC.

# Structural biology of Pumilio and Puf family RNA binding proteins

Thomas Edwards

Translation regulation plays an essential role in the differentiation and development of animal cells. One well-studied case is the control of *hunchback (hb)* mRNA during early Drosophila embryogenesis. We are investigating structural details of the interactions in the "hunchback repression complex", composed of Pumilio, *hb* mRNA, Nanos, and Brat, which defines the posterior of the embryonic anterior-posterior axis in invertebrates.

In the *Drosophila* embryo, Pumilio binds to the 3' UTR of *hb* mRNA and recruits Nanos into a complex which represses translation. Pumilio is a founder member of the Puf family of RNA binding domains, and we have solved the structure of the Puf domain, a novel RNA binding fold. The Pumilio:RNA:Nanos complex recruits Brat to repress translation, and we have solved the crystal structure of the domain of Brat which mediates interactions with Pumilio, the NHL domain. We have also recently solved the structure of Pumilio bound to its RNA target.

We have initiated structural studies of yeast Puf proteins in order to address the specificity of RNA target sequence specificity, and of mammalian RNA binding proteins that bind Pumilio and are co-factors involved in the development of gametes in order to investigate how translational control plays a role in the differentiation of germ cells..

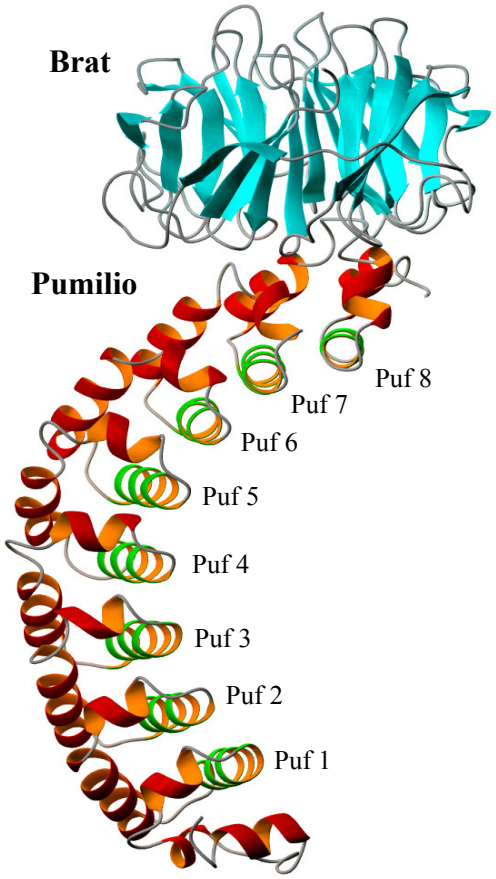



**Fig. 1**. Structure of the Brat NHL domain bound to the Pumilio Puf domain predicted by *in silico* docking, using the program BiGGER. The RNA binding surface of Pumilio is lined by the helices highlighted in green. This model is supported by mutagenesis and yeast 4-hybrid assays. Puf repeats are labeled Puf1-8.

**Collaborators**

Prof. Aneel Aggarwal, Structural Biology Program, Mount Sinai Med. School, New York.
Prof. Robin Wharton, HHMI, Molecular Genetics & Microbiology, Duke University Medical Centre, NC.
Prof. Wendy Olivas, Biology, University of Missouri-St. Louis. MO,
Prof. Howard Cooke, Chromosome Biology, MRC HGU Edinburgh.

**Funding**

This work was funded by the European Union and the University of Leeds

# Studies on the functions of hepatitis C virus proteins.

Steve Griffin, Corine StGelais, Holly Shelton, Philip Tedbury, Dave Rowlands and Mark Harris.

Hepatitis C virus (HCV) infects ~170 million individuals and is a major cause of chronic liver disease, including fibrosis, cirrhosis and hepatocellular carcinoma. The virus has a single stranded positive sense RNA genome of 9.5Kb that contains a long open reading frame encoding a single polyprotein of 3000 amino acids which is cleaved into 10 individual polypeptides by a combination of host cell and virus specific proteases. The molecular mechanisms of pathogenesis remain to be elucidated. To this end our laboratory is interested in understanding how viral protein products function within the viral lifecycle.

Current projects include the study of the p7 protein – we previously demonstrated that this protein oligomerises and forms a cation channel. Our work currently involves analysing its role in the virus lifecycle and establishing *in vitro* assays to measure p7 function. We believe that the protein plays a role in virus entry or exit – study of these events has until recently been hampered by the lack of a cell culture infectious clone of HCV. However this obstacle was removed in 2005 and we are currently working with the new infectious clone of HCV in the Category III containment facility available within the Faculty. We are also analysing a novel autoproteolytic event in the virus lifecycle – cleavage between the NS2 and NS3 proteins. We are attempting to characterise the protease activity and address the question of the role of this cleavage event in viral replication.

Thirdly we are studying the role of the NS5A protein, firstly in terms of perturbation of cellular signalling pathways - we have shown that NS5A perturbs two key mitogenic pathways within the cell; firstly NS5A blocks the Ras-ERK MAP kinase pathway and, secondly, NS5A stimulates the activity of phosphatidylinositol 3-kinase (PI3K) resulting in the activation of PI3K signalling pathways. Recently we have shown that this promotes cell survival and activitates the proto-oncogene β-catenin. The latter event has implications for the link between HCV and hepatocellular carcinoma. We are currently probing the interactions between NS5A and the SH3 domains of members of the Src family of protein tyrosine kinases using a range of techniques including bioinformatics and surface plasmon resonance, coupled with extensive mutagenesis of both NS5A and SH3 domains. Secondly we are also investigating the role of NS5A in virus replication. In particular 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 (see Fig. 1) or purify the viral RNA replication complexes with a view to undertaking a proteomic analysis of this complex.

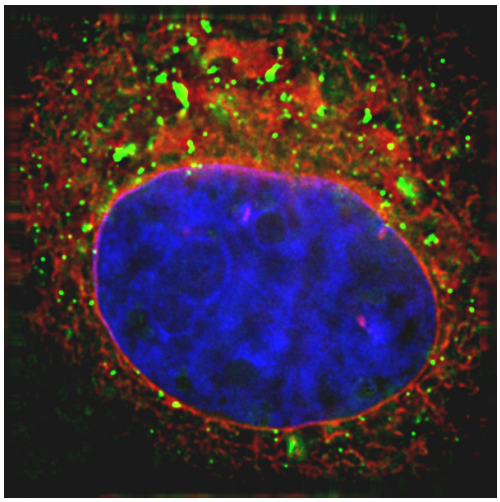


**Fig. 1:** Visualisation of an NS5A-EGFP fusion protein (green fluorescence) in a human hepatoma cell harbouring an autonomously replicating HCV sub-genomic replicon. Replication complexes are evident adjacent to the ER (stained with concanavalin A: red) but distinct from the ER. Cell counterstained with Hoechst (nucleus: blue). Image taken in the Wellcome Trust Bioimaging suite within the Astbury building.

**Publications:**

Macdonald, A., Chan, J.K.Y. & Harris, M. (2005) Perturbation of EGFR complex formation and Ras signalling in hepatoma cells harbouring the hepatitis C virus subgenomic replicon. *Journal of General Virology*, **86**, 1027-1033.

Macdonald, A., Mazaleyrat, S., McCormick, C., Street, A., Burgoyne, N., Jackson, R. M., Cazeaux, V., Saksela, K. & Harris, M. (2005) Further studies on hepatitis C virus NS5A-SH3 domain interactions: Identification of residues critical for binding, and implications for viral RNA replication and modulation of cell signalling. *Journal of General Virology*, **86**, 1035-1044

Street, A., Macdonald, A., McCormick C. & Harris, M. (2005) Hepatitis C virus NS5A-mediated activation of phosphoinositide-3-kinase results in stabilisation of cellular β-catenin and stimulation of β-catenin responsive transcription. *Journal of Virology*, **79**, 5006-5016

Griffin, S.D.C., McCormick, C.J., Rowlands, D.J. & Harris, M. (2005) Sub-cellular localisation of native p7 protein in cells expressing the full-length hepatitis C virus polyprotein. *Journal of Virology,* **79**, 15525-15536.

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. *Journal of General Virology,* **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. *Journal of General Virology,* **87**, 635-640.

**Collaborators:**

Kalle Saksela, University of Tampere, 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).

**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.

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 out pools of randomised RNA aptamers and we are currently in the process of characterising these aptamers. They will be tested out 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. *Journal of General Virology,* **87**, 563-571.

**Collaborators:**

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

**Funding:**
This work is funded by the MRC.

# The hydantoin transport protein from *Microbacterium liquefaciens*

Shun'ichi Suzuki and Peter J. F. Henderson

## Introduction

The stereo-selective hydrolyses of D,L-5-monosubstituted hydantoin compounds have been widely studied for the production of optically pure D- or L-amino acids, which are important intermediates for production of various drugs and pharmaceuticals. In spite of intensive studies for their commercialisation, the physiological roles of these hydantoin-related enzymes still remain obscure. The *hyu* genes encoding the enzymes often form gene clusters, indicating coherent roles for the combination of hydantoinase, N-carbamoylase, and HRase. Amongst these genes for the three hydantoin metabolic enzymes, we found genes encoding putative transporters, for example in the *hyu* gene clusters from *Pseudomonas sp.* NS671, *Arthrobacter aurescens* DSM 3747, and *Microbacterium liquefaciens* AJ 3912.

In order to establish, for the first time, the function of these transporters encoded by the genes located in *hyu* gene clusters, we have cloned and expressed in *Escherichia coli* one of them 'Mhp' from *M. liquefaciens* AJ3912. This is found to transport primarily 5-indolyl-methyl-hydantoin (IMH) and 5-benzyl-hydantoin (BH). The $(His)_6$-tagged protein, $MhpH_6$, was purified to an homogeneity sufficient for crystallisation trials.

## Cloning and heterologous expression of transport activity associated with the *hyuP* gene

A general strategy for the amplified expression in *Escherichia coli* of membrane transport proteins from other bacteria has been devised in the Henderson laboratory, and Shun'Ichi Suzuki of Ajinmoto Inc. in Japan spent two years in the Astbury Centre to exploit this. The *hyuP* gene from *M. liquefaciens* was ligated into the *Eco*RI+*Pst*I-cut vector pTTQ18. The resulting plasmid designated pSHP11 containing *hyuP* downstream of the IPTG-inducible *tac* promoter was transformed into the *E. coli* BLR host, which was grown plus or minus IPTG. The washed cells of the recombinant strain (*E. coli*/pSHP11) were examined for the uptake of radiolabelled L-BH or L-IMH made with the help of Simon Patching and Richard Herbert.

When using either L-BH or L-IMH as the substrate, higher uptake was observed in the cells of the IPTG-induced culture than in those of the uninduced culture (Fig. 1). Using induced cells of *E. coli* BLR/pTTQ18 without any inserts, the uptake of substrates was less than that observed in uninduced cells of *E. coli*BLR/pSHP11 (data not shown), demonstrating that the host strain did not significantly accumulate these hydantoins. The extents of uptake kept increasing even after 300 sec, but the rates of uptake for both the substrates progressively decreased as the reactions proceeded. Judging from both the initial uptake rates and the extents of uptake in 300 sec, L-IMH was a better substrate for $MhpH_6$ than L-BH.

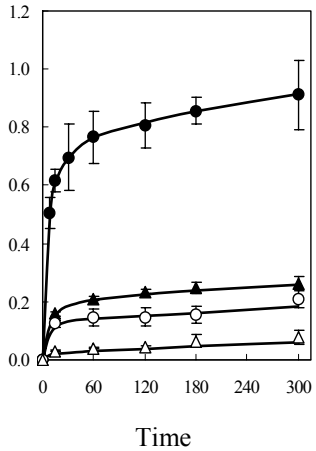



**Figure 1. Time course of uptakes of 5-substituted-hydantoins by *E. coli*/pSHP11.** $^3$H-L-IMH (circles) and $^3$H-L-BH (triangles) were used as the substrates. The washed cells of *E. coli*/pSHP11 obtained by growth on M9 casamino acid medium with (closed symbols) or without (open symbols) induction by IPTG were used for the uptake assay carried out in standard conditions (5 mM MES buffer (pH 6.6), 150 mM KCl, and 25 □M substrate, at 25°C).

**Amplified expression and identification of the $MhpH_6$ protein.**
The Mhp protein containing a $His_6$ tag at its C-terminus, was heterologously expressed by the construct pTTQ18 in an *E. coli* host. The level of expression was sufficient for the extra protein to stain with Coomassie Blue in membranes from induced, but not uninduced cells (Fig 2A, lanes 2, 3). The protein was not observed in the soluble fraction (Fig 2A, lane 4). After further separation of inner and outer membrane fractions from total membrane fraction, the localisation of $MhpH_6$ was mainly observed in the inner membrane fraction (Fig 2A, lane 5, 6). The apparent molecular mass of the produced protein was approximately 36 kDa on SDS-PAGE, although its predicted molecular mass is 54, 580, an anomaly common for membrane transport proteins. Its identification was authenticated by the Western blot to the $RGSH_6$ epitope (Fig. 2B), in which a positive reaction correlated with the molecular mass and appearance/disappearance of the putative MhpH6 in the Coomassie-stained gel (Fig. 2A).

**Solubilization and purification of $MhpH_6$.**
The solubilization of $MhpH_6$ was efficiently promoted by treatment with 1% (w/v) DDM (Fig 2A, B, lanes 7, 8). Finally, the purification of $MhpH_6$ to electrophoretic homogeneity was achieved by using Ni-NTA resin (Fig 2A, lane 9). Through all the purification steps, the molecular mass of $MhpH_6$ was observed at 36 kDa (Fig 2A and B, lanes 5,7,8,9) and no other proteins were observed in the Western-blot analysis (Fig. 2B).

The N-terminal amino acid sequence of the purified $MhpH_6$ was determined for ten residues as MNSTPIEEAR, corresponding exactly to what was expected from the nucleotide sequence in the constructed plasmid. Coupled with the integrity of the $RGSH_6$ C-terminal end of the protein shown by the Western blot this verifies that the purified polypeptide chain of $MhpH_6$ is intact.

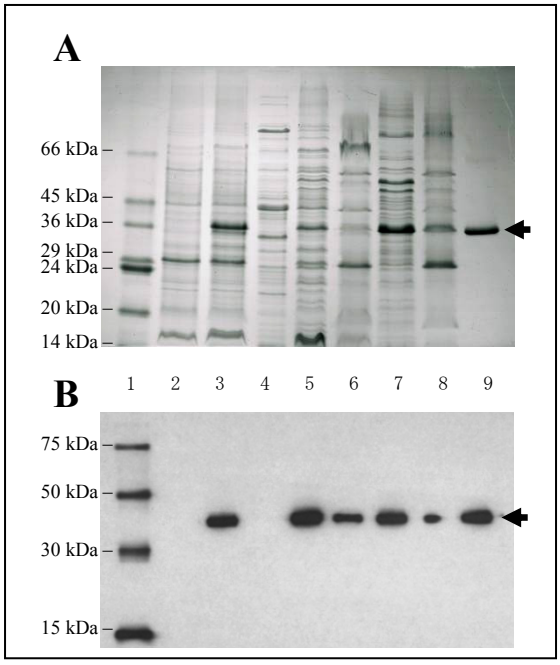



**Fig. 2. Localization, solubilization, and purification of $MHP_6$ heterologously expressed in *E. coli*.** The expression of $MhpH_6$ in *E. coli* was analyzed by SDS-PAGE (A, stained with Coomassie Brilliant Blue R) and western blotting (B, detected with anti-RGS·His antibody). Lane 1, marker proteins indicated the positions by thin allows (with molecular masses). The protein fractions were prepared from *E. coli*/pSHP11 cells obtained from induced (lane 3-6) or uninduced culture (lane 2) using LM medium. Lane 2 and 3, total membrane fraction; 4, soluble fraction; 5, inner membrane fraction; 6, outer membrane fraction. The proteins in the inner membrane fraction of induced culture were solubilized by DDM, then soluble (lane 7) and insoluble (lane 8) fractions were obtained. $MhpH_6$ was then purified to electrophoretical homogeneity by using Ni-NTA resin (lane 9). Thick arrows indicate the position of $MhpH_6$.

**Conclusions** The *hyu* gene of *M. liquefaciens* was shown to encode a protein for transport **of** hydantoins. This is sufficiently important for Ajinomoto to protect the IPR by patenting the discovery. The purified protein yields diffracting crystals.

**Acknowledgements** This work was funded by Ajinomoto Inc and BBSRC.

# Characterization of the severe acute respiratory coronavirus RNA binding protein

Jae-Hwan You and Julian A. Hiscox.

## Introduction

Coronaviruses are a group of positive strand RNA viruses, which replicate in the cytoplasm of the infected cell and cause a variety of respiratory and gastrointestinal illnesses. In 2003 coronaviruses were identified as the causative agent of severe acute respiratory syndrome (SARS). The death rate following infection approached almost 10%. Work in our laboratory has focused on characterizing the coronavirus nucleocapsid protein (N protein), which is involved in binding viral RNA, the regulation of viral replication and also signaling in the host cell. N protein is composed of three distinct regions, containing RNA binding motif(s), and appropriate signals for modulating cell signalling. We studied the sub-cellular localisation of SARS-CoV N protein. In infected cells SARS-CoV N protein localised exclusively to the cytoplasm. In contrast to the avian coronavirus N protein, when over expressed SARS-CoV N protein remained principally localised to the cytoplasm with very few cells exhibiting nucleolar localisation. Bioinformatic analysis and deletion mutagenesis coupled to confocal microscopy and live cell imaging revealed that SARS-CoV N protein regions I and III contained nuclear localisation signals and region II contained a nucleolar retention signal. However, cytoplasmic localisation was directed by region III and was the dominant localisation signal in the protein.

## Collaborators

Prof. Luis Enjuanes, Centro Nacional de Biotecnología, Spain.
Dr. Maria Zambon, Health Protection Agency, UK.

## Publications

You, J.H., Dove, B.K., Enjuanes L., DeDiego, M.L., Alvarez, E., Howell, G., Heinen, P., Zamon, M. & Hiscox, J. A. (2005). Sub-cellular localisation of the severe acute respiratory syndrome coronavirus nucleocapsid protein. *Journal of General Virology.* 86, 3303-3310.

# Delineation of nucleolar trafficking signals

Mark L. Reed and Julian A. Hiscox.

## Introduction

Coronaviruses are a group of positive strand RNA viruses, which replicate in the cytoplasm of the infected cell and cause a variety of respiratory and gastrointestinal illnesses. Unlike nuclear localisation signals there is no obvious consensus sequence for the targeting of proteins to the nucleolus. The nucleolus is a dynamic sub-nuclear structure which is crucial to the normal operation of the eukaryotic cell. Studying nucleolar trafficking signals is problematic as many nucleolar retention signals (NoRSs) are part of classical nuclear localisation signals. In addition, there is no known consensus signal with which to inform a study. The avian infectious bronchitis virus coronavirus nucleocapsid (N) protein localises to the cytoplasm and the nucleolus. Mutagenesis was used to delineate a novel eight amino acid motif that was necessary and sufficient to for nucleolar retention of N protein, and co-localise with nucleolin and fibrillarin. Additionally, a classical nuclear export signal functioned to direct N protein to the cytoplasm. Comparison of the coronavirus NoRSs with known cellular and other viral NoRSs revealed that these motifs have conserved arginine residues. Molecular modelling, using the solution structure of severe acute respiratory (SARS) coronavirus N protein, revealed that this motif is available for interaction with cellular factors which may mediate nucleolar localisation. We hypothesise that the N protein uses these signals to traffic to and from the nucleolus and the cytoplasm.

## Collaborators

Prof. Gavin Brooks, School of Pharmacy, University of Reading.

## Publications

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*, in press.

# Interaction of coronaviruses with the nucleolus

Brian K. Dove and Julian A. Hiscox.

## Introduction

Coronaviruses are a group of positive strand RNA viruses, which replicate in the cytoplasm of the infected cell and cause a variety of respiratory and gastrointestinal illnesses. The nucleolus is a dynamic sub-nuclear structure involved in ribosome sub-unit biogenesis, cell cycle control and mediating responses to cell stress, amongst other functions. Whilst many different viruses target proteins to the nucleolus and recruit nucleolar proteins to facilitate virus replication, the affect of infection on the nucleolus in terms of morphology and protein content is unknown. Previously we have shown that the coronavirus nucleocapsid protein will localise to the nucleolus. In this study, using the avian infectious bronchitis coronavirus, we have shown that virus infection results in a number of changes to the nucleolus both in terms of gross morphology and protein content. Using confocal microscopy coupled with fluorescent labelled nucleolar marker proteins we observed changes in the morphology of the nucleolus including an enlarged fibrillar centre. We found that the tumour suppressor protein, p53, which localises normally to the nucleus and nucleolus, was redistributed predominately to the cytoplasm.

## Collaborators

Prof. Gavin Brooks, School of Pharmacy, University of Reading.

## Publications

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*, doi: 10.1111/j.1462-5822.2006.00698.x.

# Interaction of coronaviruses with the cell cycle

Brian K. Dove and Julian A. Hiscox.

## Introduction

Coronaviruses are a group of positive strand RNA viruses, which replicate in the cytoplasm of the infected cell and cause a variety of respiratory and gastrointestinal illnesses. In eukaryotic cells, cell growth and division occurs in a stepwise, orderly fashion described by a process known as the cell cycle. The relationship between positive strand RNA viruses and the cell cycle and the concomitant effects on virus replication are not clearly understood. We have shown that infection of asynchronously replicating and synchronised replicating cells with the avian coronavirus infectious bronchitis virus, a positive strand RNA virus, resulted in the accumulation of infected cells in the G2/M phase of the cell cycle. Analysis of various cell cycle regulatory proteins and cellular morphology indicated a down-regulation of cyclins D1 and D2 (G1 regulatory cyclins) and that a proportion of virus-infected cells underwent aberrant cytokinesis, in which the cells had undergone nuclear, but not cytoplasmic division. We assessed the impact of the perturbations on the cell cycle on virus infected cells and found that IBV-infected G2/M phase synchronised cells exhibited increased viral protein production when released from the block when compared to cells synchronised in the G0 phase or in asynchronously replicating cells. Our data suggested that IBV induces a G2/M phase arrest in infected cells to promote favourable conditions for viral replication.

## Collaborators

Drs Paul Britton and Dave Cavanagh, Institute for Animal Health. Compton, UK.
Dr. Ian Tarpy, Intervet, UK.
Prof. Gavin Brooks, School of Pharmacy, University of Reading.

## Publications

Dove, B.K., Brooks, G., Bicknell, K., Wurm, T. & Hiscox, J. A. (2006). Cell cycle perturbations induced by infection with the coronavirus infectious bronchitis virus and their affect on virus replication. *Journal of Virology*, 80, in press.

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

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

## Introduction

The interactions of proteins with small molecule ligands are of central importance to much in biology. The ability to predict and manipulate such interactions will open a number of future avenues in research, biotechnology and therapeutics. In particular, an understanding of the molecular basis of protein-small molecule interactions is crucial to attempts to design novel drug technologies. Extensive research efforts have been aimed at relating the known structure of protein-ligand complexes to the thermodynamics of that interaction. Currently, such attempts have resulted in only limited improvements in our ability to predict or design novel protein-ligand interactions. One reason for the limited success of these approaches is their neglect of the role of protein and ligand conformational dynamics in determining ligand binding thermodynamics.

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

## NMR relaxation studies of ABP

Previously, we have established near-complete spectral assignment for apoABP and for the complex of ABP with the ligand D-galactose using conventional triple-resonance assignment strategies. Full sets of $^{15}N$ relaxation measurements (T1, T2 and NOE at three magnetic fields) have been made for ABP in its apo form and in the complex. These data have been analysed to extract information on the extent and timescales of ps-ns motion of the protein backbone. We make the somewhat surprising observation that on these timescales, apoABP is less dynamic than the protein in complex with its ligand. Although no data is available for residues directly involved in ligand binding, this effect appears quite general, with changes in dynamics dispersed across both domains of ABP.

In order to rationalise this unexpected finding, it is useful to consider the entropic costs of the protein-ligand interaction. Loss of translational, rotational and internal degrees of freedom of the ligand, as well as loss of translational and rotational degrees of freedom of the many bound water molecules involved in the ligand-binding interaction, are expected to make large unfavourable contributions to the entropy of binding. This raises the possibility that the observed changes in protein dynamics on ligand interaction reflect the protein paying some of this entropic cost of binding. Similar measurements of relaxation parameters for sidechain methyl and NH2 groups in ABP are also planned. In addition, we propose to use analogous approaches to probe the conformational dynamics of the bound ligand.

**Funding:** This project is funded by the Wellcome Trust

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

Elizabeth Barratt, Richard Bingham, Natalia Shimokhina, Neil Syme, Simon Phillips and Steve Homans

## Introduction

A better understanding of the factors that determine molecular recognition between a protein and a ligand remains an important goal in biophysical chemistry. Despite enormous advances in the three-dimensional structure determination of proteins and their complexes, our ability to predict binding affinity from structure remains severely limited. The main reason for this problem is that affinities are governed not only by energetic considerations concerning the precise spatial disposition of interacting groups, but also by the dynamics of these groups. Hence, to predict accurately the affinity of a protein for a given ligand, it is essential to have prior knowledge of both the enthalpy and the entropy of binding. However, these components are enormously complex, and involve contributions arising from new ligand-protein interactions, solvent rearrangement and changes in protein and ligand degrees of freedom. Detailed decomposition of the thermodynamics of a model hydrophobic system, the binding of 2-methoxy-3-isobutyl pyrazine (IBMP) to the major urinary protein (MUP), was recently initiated. Using a combination of isothermal titration calorimetry (ITC), X-ray crystallography, NMR relaxation measurements, all-atom molecular dynamic simulations, and site-directed mutagenesis facilitated the acquisition of thermodynamic contributions from restrictions in protein and ligand motions and formation of new protein-ligand interactions. By carrying out complete thermodynamic decomposition, we hope to further understand individual factors contributing to binding parameters.

## Ligand desolvation

Ligand solvation undoubtedly plays an important role in interaction thermodynamics, but still remains notoriously difficult to quantitate. Our approach for quantitation of ligand desolvation terms involves the experimental measurement of standard solvation free energies, enthalpies and entropies for the relevant ligands using air/water partition coefficient measurements, supported by free energy perturbation calculations. The theoretically derived solvation free energy, enthalpy and entropy are, in general, in good agreement with those derived experimentally and fit well in the context of binding data previously reported for this model, but suggest that ligand desolvation is not a driving force for binding or MUP specificity.

## Solute-solute dispersive interactions

The complex formed between MUP and IBMP contains a single hydrogen bond, formed between the ring nitrogen of IBMP and the hydroxyl group of Y120 in MUP. The interaction between these molecules is enthalpically driven and one would expect removal of this hydrogen bond to make ligand binding less favorable. The interaction of IBMP with the MUP Y120F mutant is less favorable than with the wild-type, but ITC data show it is still enthalpically driven. Crystal structures and molecular modeling have shown the binding site of MUP is poorly solvated, which implies that van der Waals interactions are predominant in driving MUP-ligand association.

The discovery that MUP binds primary aliphatic alcohols permitted a verification of this suggestion. ITC measurements of the interaction between MUP and panel of alcohols from pentanol to nonanol demonstrated a linear dependence of the enthalpy and entropy of binding on aliphatic chain length. The enthalpy of binding became more favorable and the entropy of binding less favorable. The absence of the "classical" hydrophobic thermodynamic signature confirmed predictions that dispersive interactions play a major role in ligand association to MUP. The results of this work, combined with the suggestion that solvent reorganisation

around the ligand does not play key role in binding, strengthening the hypothesis that van der Waals interactions are the driving force for hydrophobic association in systems with a poorly solvated protein binding pocket.

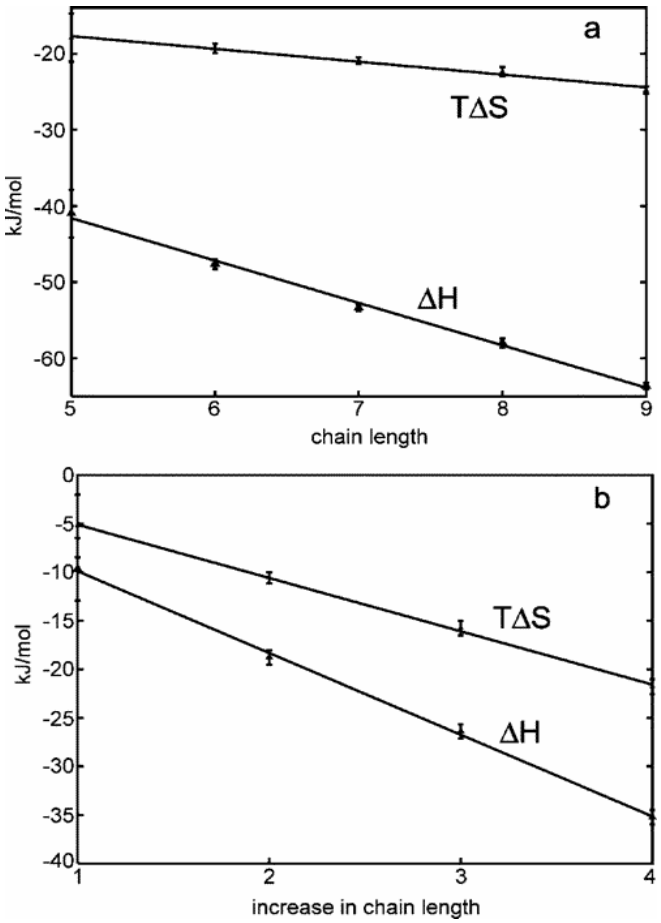



**Fig.1.** Thermodynamics of binding of primary aliphatic alcohols to MUP-I. (a) Global enthalpies of binding ($\Delta H°_b$) and entropies of binding ($T\Delta S°_b$) plotted versus carbon chain length. (b) Differences between "intrinsic" enthalpies of binding ($[\Delta H°_{i2} - \Delta H°_{i1}]$) and entropies of binding ($T[\Delta S°_{i2} - \Delta S°_{i1}]$) plotted versus increase in chain length.

**Collaborators**
Daniel J. Warner and Charles A. Laughton at the University of Nottingham.

**Publications**
Barratt E., Bingham R.J., Warner D.J., Laughton C.A., Phillips S.E.V. & Homans S.W. (2005) Van der Waals Interactions Dominate Ligand-Protein Association in a Protein Binding Site Occluded from Solvent Water. *J. Am. Chem. Soc.* **127**, 11827-11834

Malham R., Johnstone S., Bingham R.J., Barratt E., Phillips S.E.V., Laughton C.A. & Homans S.W. (2005) Strong Solute-Solute Dispersive Interactions in a Protein-Ligand Complex. *J. Am. Chem. Soc.* **127**, 17061-17067

**Funding**
We acknowledge the support of the Wellcome Trust and BBSRC.

# Dynamics of the MS2 coat protein

Gary Thompson, Nicola Stonehouse, Peter Stockley and Steve Homans

## Introduction

The MS2 bacteriophage presents a rich system for understanding the processes involved in assembly of large macromolecular systems and especially viruses with Icosohedral geometry. Understanding this system is especially rewarding as it provides both a paradigmatic model for the assembly mechanism, and insights into the design of anti-viral drugs and nanotechnology systems. Molecular assembly in these systems is not a simple problem as, due to steric and symmetry constraints, identical coat proteins have to take up multiple conformations in the final virus (quasi-equivalence). In the case of the MS2 protein this involves forming protein dimers where the FG loop [residues 66 -83] take-up one of three conformation A, B or C, with each dimer containing either AB or CC pairs.

It has previously been shown that addition of the 19mer MS2 translational repressor RNA (TR) to MS2 coat protein (MS2CP) initializes assembly of the complete virus. It has been postulated that the response to the TR RNA may be due to a templating reaction where the binding of the RNA affects the dynamics and/or conformation of the FG loop.

The aim of this project is to analyse the dynamics and structure of the protein in solution using NMR spectroscopy. To carry out this study we have used a non-assembling mutant of the MS2 coat protein, MS2W82R, which still forms dimers. This system allows us to measure conformational and dynamic (entropic) changes within the protein, with and without RNA present, by NMR. A nearly complete assignment of *apo*-MS2W82R protein has been acheieved using a triple labeled NMR sample [100% $^{15}$N,100% $^{13}$C, 50%D], and spectra of the triple labeled *holo*-MS2W82R–TR-RNA complex have also been acquired and are being assigned. We also have an almost complete analysis of ps-ns dynamics in the *apo* state and are collecting and analyzing data for the *holo* state.

A. B.

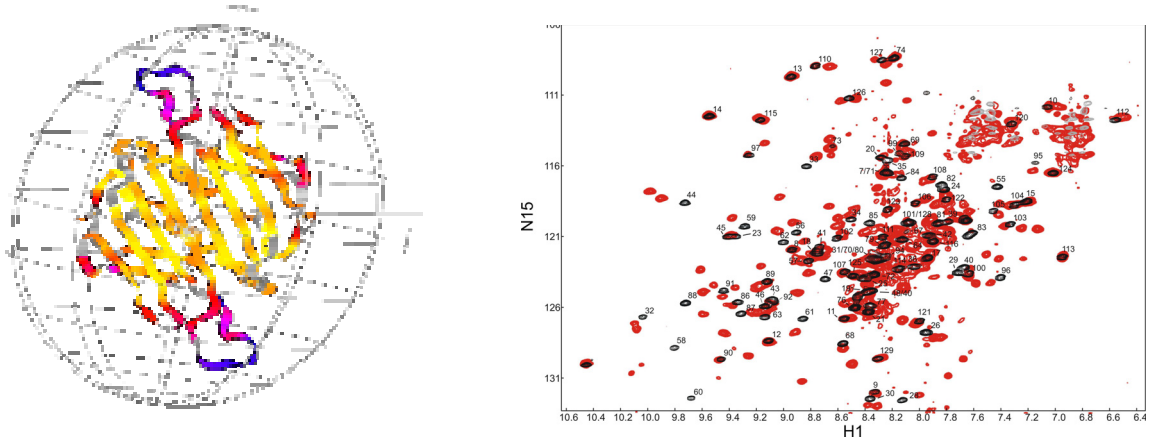



**Fig. 1. Panel A;** Dynamics on ps-ns timescales for the backbone of *apo*-MS2W82R (order parameters ($S^2$) as detected by $^{1}$H-$^{15}$N relaxation experiments). The image shows a ribbon representation of the backbone of MS2W82R [PDB code 1MSC] with $S^2$ shown by colour, yellow - blue; $S^2$ = 0.9-0.3. **Panel** B: $^{1}$H-$^{15}$N HSQC spectra of *apo* MS2W82R (black) and *holo* MS2W82R (red) with TR RNA, assignments are shown for the *apo* state.

## Publications

Kalverda, A.P., Thompson, G.S., & Turnbull W.B. (2005) The importance of protein structural dynamics and the contribution of NMR spectroscopy in 'The Encylopedia of

Genetics, Genomics, Proteomics and Bioinformatics', Wiley Interscience M. J. Dunn, *et al*. editors ISBN 0470849746.

Stockley, P.G., Ashcroft, A.E. Francese, S. Thompson, G.S., Ransom, N., Smith, A., Homans, S.W., & Stonehouse, N.J. (2005) "Dissecting the details of assembly of a T=3 Phage Capsid. *Journal of Theoretical Medicine.* **6,** 119–125.

**Funding**
We wish to thank the BBSRC for funding GST.

# Software tools for docking and structure based drug design

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

## Introduction

Structure Based Drug Design (SBDD) is an increasingly popular method for identifying novel lead compounds. It uses the three dimensional structure of a protein-ligand binding site to construct or dock a drug-like molecule such that it has complementary chemistry and shape. Several aspects of SBDD have been addressed in this project. Identification of druggable binding sites was addressed in a previous research project. Having identified a potential ligand binding site, the task remains to fit drug molecules inside the site. Two alternative methods are being addressed in this project; (1) Flexible ligand docking, (2) Lead optimisation.

## 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.

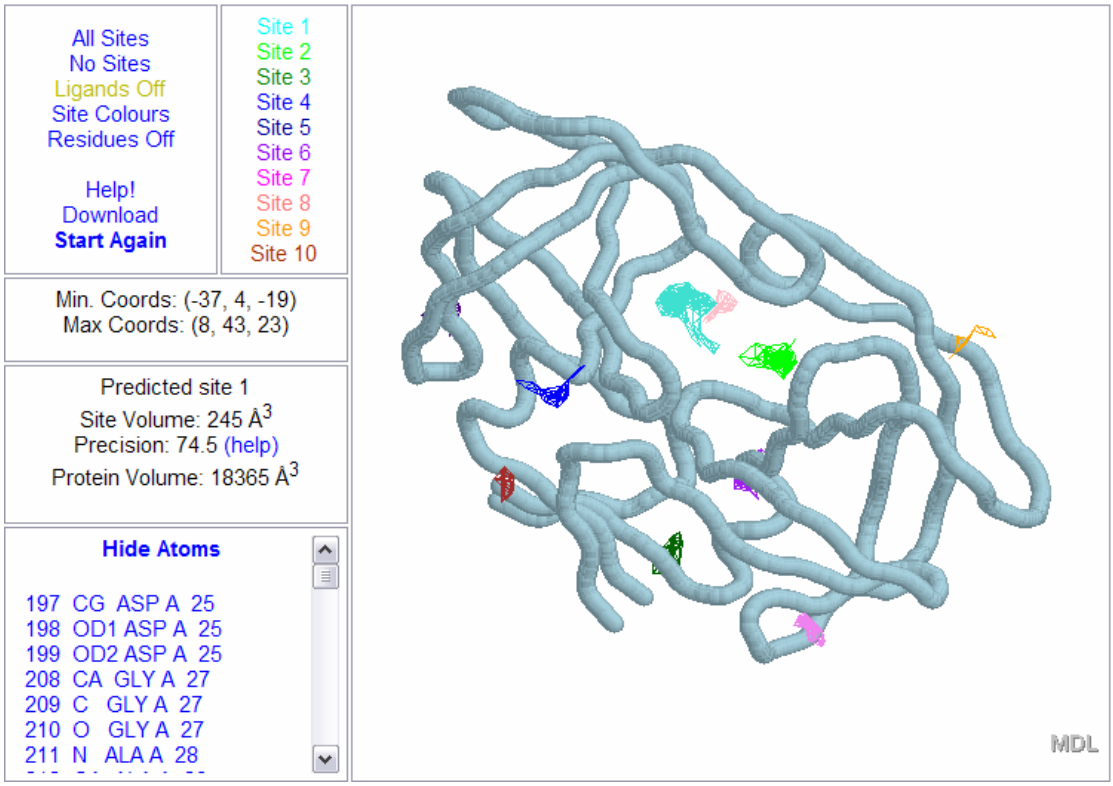



**Fig. 1** Ligand binding site prediction for HIV protease (1aaq). The protein is shown in grey tubing with ten colour-coded predicted binding sites, each represented by a cluster of methyl probes. The location of the ligand binding site is identified by the 1st (turquoise), 2nd (light green) and 8th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

Part of the current project was to develop an online user-friendly interface to the program.

**Lead optimisation**

A lead optimisation tool has been developed and is being tested. It is pre-publication and will not be described in detail here.

**Publications**

Laurie, T. & Jackson, R.M. (2005) Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics,* **21**, 1908-16.

**Funding**

This project is funded by the BBSRC.

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

Nicola Gold and Richard Jackson

Structural genomics projects produce large amounts of data including some 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 (Fig. 1). A multiple alignment of structurally equivalenced atoms is also provided.

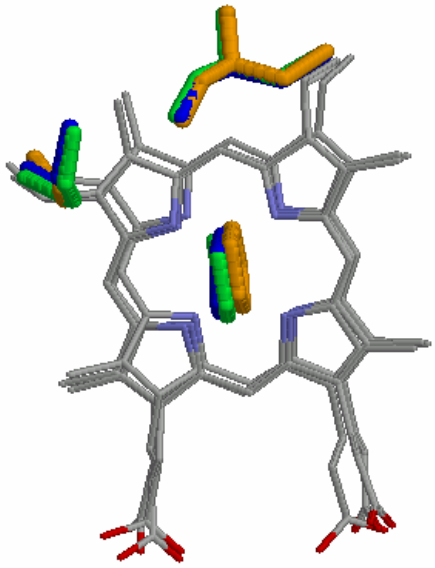


**Figure 1**. Superposition of the haem binding site of cytochrome c4 (1m6z), flavocytochrome c (1dxr) and cytochrome f (1e2z). These proteins have different folds. 1m6z has a cytochrome c fold, 1dxr has a multihaem cytochrome fold and 1e2z has a common fold of diphtheria toxin/transcription factors/cytochrome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

Comparison of each binding site to all other binding sites provides a matrix of similarity scores. Fig. 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.

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.

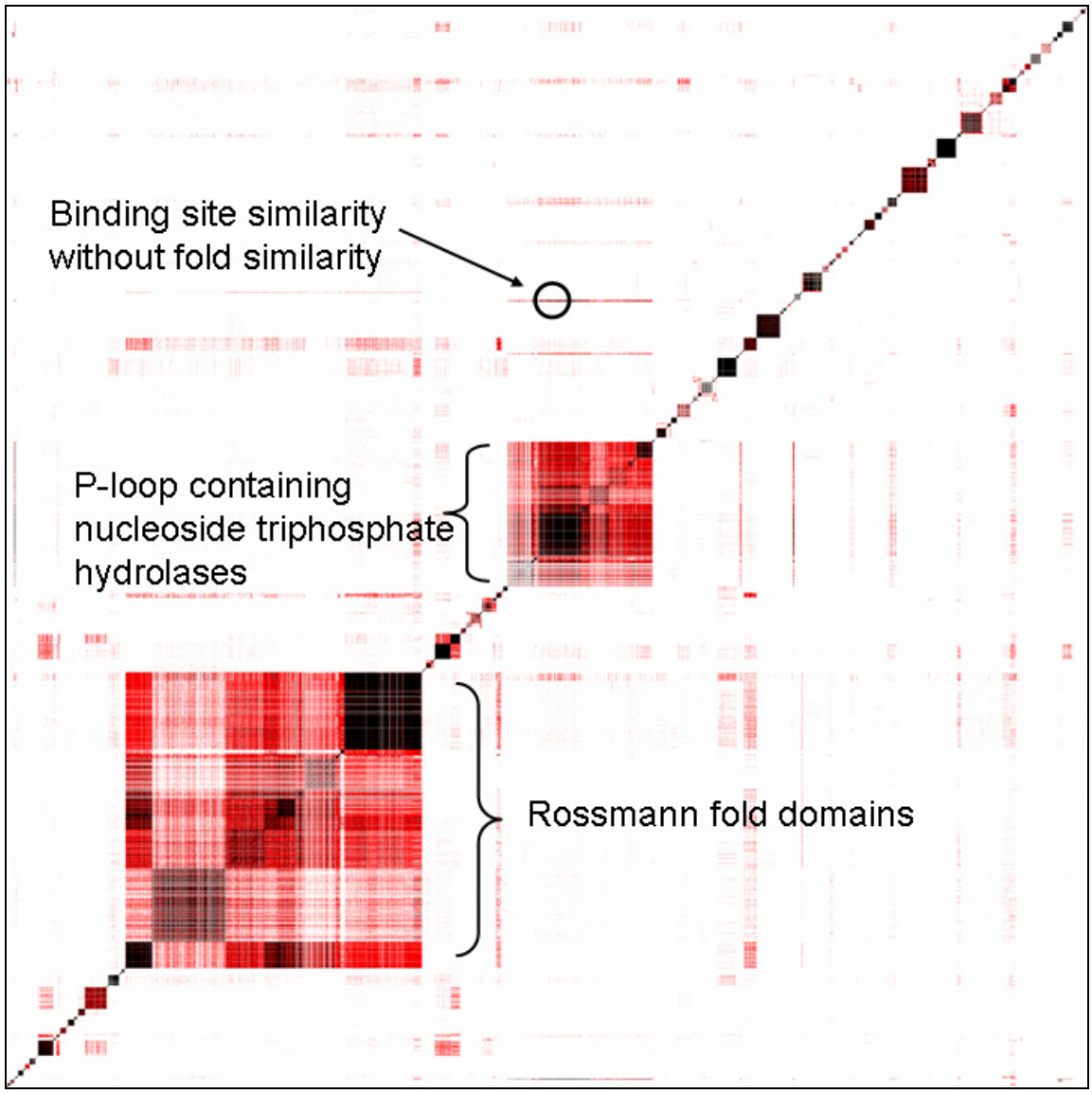



**Fig. 2**. Similarity matrix from the database of nucleotide binding sites. Strong similarity (score>39) is indicated by a black dot and weaker but still significant similarity (score 25-39) is indicated with a red dot. Binding sites have been ordered along the axes by their evolutionary relationships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

**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* (in press).

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.

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

Monika Rella, Christopher Rushworth, Jodie Guy, Tony Turner and Richard Jackson

## Introduction

Angiotensin Converting Enzyme 2 is a new member of the M2 metalloprotease family and a unique human homologue of the Angiotensin Converting Enzyme (ACE) with distinct catalytic activity but a conserved active site. ACE plays a central role in the Renin Angiotensin System (RAS) and represents a critical drug target for the treatment of hypertension and heart disease. The discovery of ACE2 increases the complexity of the RAS and makes it an interesting new cardio-renal disease target. ACE2 has been associated with hypertension, heart and kidney disease and more recently as the functional receptor for the SARS-Coronavirus and its full physiological roles are currently under intense investigation.

In this study, a computational virtual screening approach was applied to identify novel ACE2 inhibitors based on the ACE2 crystal structure in complex with a potent inhibitor. Virtual screening has become an important *in silico* technique next to traditional high throughput screening in the pharmaceutical industry in search for new active compounds including ligand- and structure-based methods. Pharmacophore modelling is one such approach where active molecules are analysed for common steric and electronic features responsible for specific drug-receptor interactions. Alternatively, protein-ligand crystal structure complexes can be used to construct specific receptor-based pharmacophore models. The aim of this study was to generate a set of protein structure-based ACE2 pharmacophore models to apply as filters for 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. We show that all compounds displayed some inhibitory effect on ACE2 activity. 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% and represents a successful virtual screening application.

## Collaborators

Monika Rella thanks Thierry Langer, University of Innsbruck, Austria, for the opportunity to carry out the major part of this work as visiting research scientist in his lab.

## Publications

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 *J. Chem. Inf. Mod*., in press.

## Funding

This work was funded by the University of Leeds, the BBSRC and The Wellcome Trust.

# Predicting protein-protein interactions

Nicholas Burgoyne and Richard Jackson

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

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

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

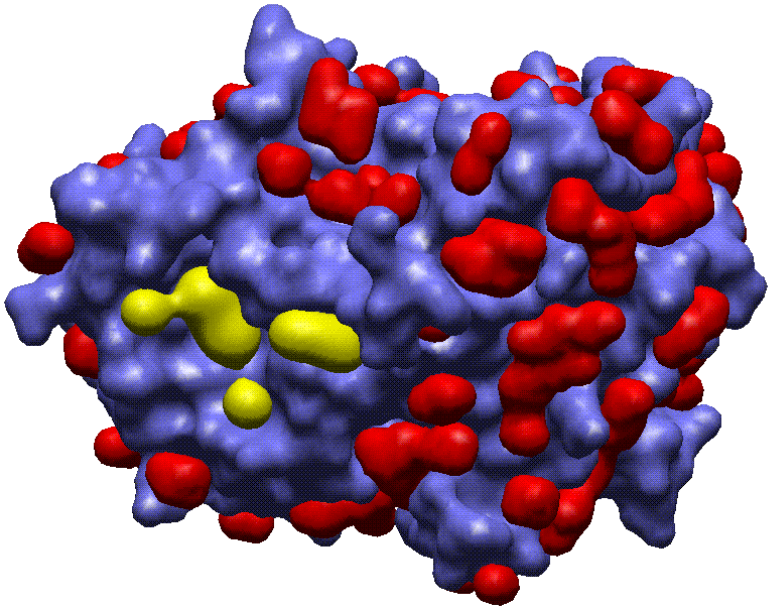


**Fig. 1** Probes (red and yellow) fill the identified clefts on the protein surface of Ran (blue). Those clefts that are occupied during the interaction with nuclear transport factor-2 are coloured yellow; these are more easily desolvated than many of the other clefts on the protein surface.

Despite the different proteins that interact, we have shown that there are consistencies between the interface clefts of different proteins. We have demonstrated that interface clefts are more easily desolvated than the non-interface clefts. Interestingly this result is true for a

wide range of protein interactions including antibodies, antigens, enzymes and their protein inhibitors. We have also shown that protein-ligand interfaces are more readily desolvated than other clefts found on the protein surface. However, there are often better discriminators of the interface. It is known that these sites barely alter on formation of the protein complex. Therefore the prediction of the relevant clefts could act as a basis for the generation of possible protein-protein interactions

**Publications**

Burgoyne N.J. & Jackson R.M. (2006) Predicting protein interaction sites: Binding hot-spots in protein-protein and protein-ligand interfaces, *Bioinformatics*, In Press.

**Funding**

This work is funded by the MRC.

# Electrodes for redox-active membrane proteins

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

## Introduction

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

## Cholesterol tethers to 'wire' membrane proteins

We have prepared electrode surfaces which enables 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 (Fig. 1). For this purpose lipid derivatives have been used, which, via a hydrophilic linker, are connected to a thiol group that forms self-assembled monolayers (SAMs) on gold electrodes.

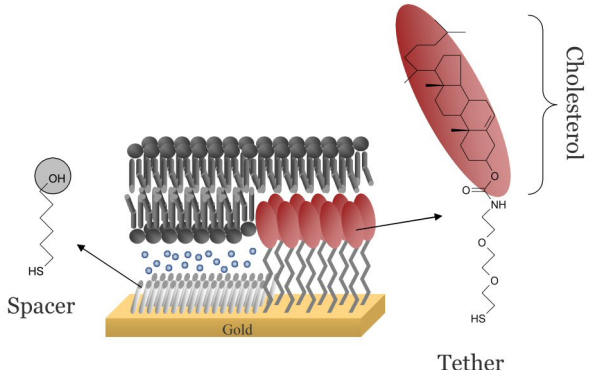
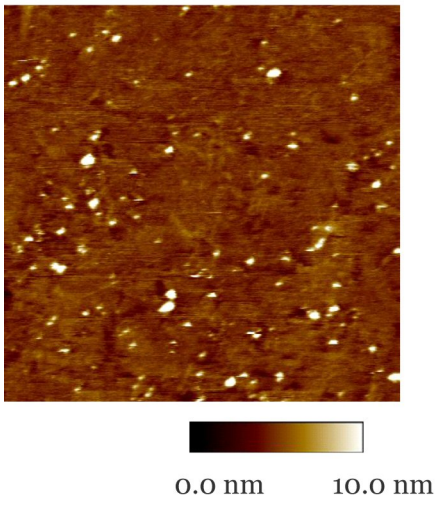



**Fig. 1**: Chemical structures of 6-mercaptohexanol (spacer) and the cholesterol tether molecule used to form tBLMs.

An ubiquinol oxidase from *Escherichia coli*, cytochrome $bo_3$ ($cbo_3$), has been co-immobilised into these tethered bilayer lipid membranes (tBLMs). A planar membrane architecture is formed by self assembly of proteoliposomes on cholesterol modified electrodes. The structure of these electrode surfaces are characterised by surface plasmon resonance (SPR), electrochemical impedance spectroscopy (EIS) and tapping-mode atomic force microscopy



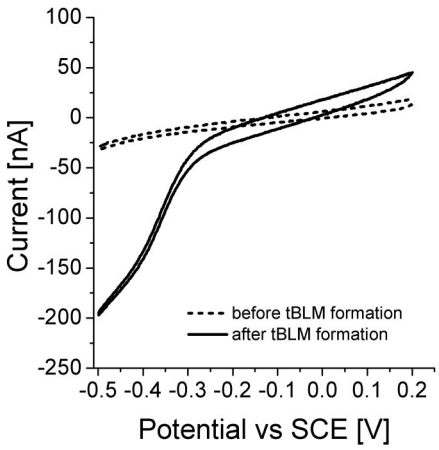




**Fig. 2:** (Left) TM-AFM results of a tBLM with $cbo_3$.The membrane proteins are observed as heightened (5-10 nm ) areas protruding above the planar lipid bilayer. (Right) CV results indicating catalytic oxygen reduction after formation of a tBLM containing $cbo_3$

(TM-AFM) and all confirm that a planar lipid bilayer is formed in which the membrane proteins are incorporated. The functionality of c*bo*$_3$ is investigated by cyclic voltammetry (CV) and is confirmed by the catalytic reduction of oxygen, resulting in the signals shown in Fig. 2, which are typical for electrocatalysis. Interfacial electron transfer to c*bo*$_3$ is found to be mediated by the membrane-localised ubiquinol-8, the physiological electron donor of c*bo*$_3$. Enzyme coverages observed with TM-AFM and CV coincide (2-8.5 fmol·cm$^{-2}$) indicating that most - if not all - c*bo*$_3$ on the surface is catalytically active and thus retains its integrity during immobilisation.

**Collaborators and Funding**
Dr. Robert B. Gennis (University of Illinois, Urbana, USA)
This work was funded by the BBSRC (David Phillips fellowship).

**Publication**
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.

# NMR facility

Arnout Kalverda and Steve Homans

**Overview of facility**

The NMR facility is equipped with 750 MHz, 600 MHz and two 500 MHz Varian Inova NMR spectrometers. All instruments are setup to use $^{1}H$, $^{13}C$, $^{15}N$ and $^{2}H$ during normal operation. In December 2005 a cryoprobe has been installed on the 750 MHz NMR spectrometer, thereby further increasing the sensitivity by a factor of three. This addition is crucial to the success of studies on large complex systems where sensitivity is of prime importance due to the fast relaxation of signals and often lower yields from expression. The Biopack suite of pulse sequences is used for most data acquisitions, supplemented with in house written sequences where needed.

**Dynamics and thermodynamics of ligand binding**

The Mouse Urinary protein is being used as a model system for studying the interplay between dynamics and thermodynamics in protein ligand binding. ITC measurements have shown that the binding of a series of small hydrophobic ligands to the protein is enthalpically driven. Crystallography, molecular dynamics simulations, ITC measurements and NMR experiments on wild type and site directed mutants bound to a variety of closely related ligands are used to build a picture of the thermodynamic features of ligand binding in this system. Hydrogen exchange measurements show no difference between the apo-protein and ligand bound protein. Previously $^{15}N$ and $^{2}H$ relaxation measurements have indicated both increases and decreases in backbone and side chain dynamics that are largely compensatory upon ligand binding. These have now been complemented with measurements of correlated chemical shift modulations that are sensitive to slower timescale motions. The latter hint at changes in slow dynamics in similar areas of the protein for which differences in fast time scale dynamics are also seen.

**NMR of protein folding and misfolding**

The dynamics of the partially folded ensemble of $\beta_2$-microglobulin at low pH have been probed by NMR. Measurement of transverse relaxation (in this case of $^{15}N$) gives an exquisite probe that is sensitive to the structural reorganisations that occurs in such ensembles of partially folded and more highly unfolded molecules. Transitions between different species in such ensembles often occurs on the ms timescale leading to either extreme broadening of the spectrum or to the appearance of an unfolded spectrum with broadened peaks arising from regions of the protein that are exchanging between more folded and more highly unfolded states. $^{15}N$ T2 relaxation and relaxation dispersion were measured on the wild type and two site directed mutants in which solvent exposed aromatic residues had been mutated. The results point to the existence of two or more distinct species that include at least one that has substantial non-native structure in the central region of the protein stabilised by hydrophobic clusters.

**A stable single α-helix domain in Myosin 10**

A variety of Myosins are thought to dimerise through a coiled coil domain based on results from Myosin 2. In Myosin 10 part of the predicted coiled coil domain is highly charged with a repetitive sequence of the form ((R/K)XXXEXX). CD and NMR experiments show that a 36 residue peptide covering this region of Myosin 10 is helical throughout the repeat region. This helix is stabilised by favourably charge interactions between the Arg/Lys and the Glu side chains. The presence of this helix instead of the coiled coil makes it likely that Myosin 10 is monomeric which is supported by electron microscopy data. It forms a novel domain

that elongates the neck of the Myosin. This increase in the length of the lever would result in a longer power stroke for the protein.

**Publications**

Knight, P.J., Thirumurugan, K., Xu, Y.H., Wang, F., Kalverda, A.P., Stafford, W.F. & Peckham, M. (2005) The predicted coiled-coil domain of Myosin 10 forms an elongated domain that lengthens the head. *J. Biol. Chem.* **280**, 34702-34708.

Platt, G.W., McParland, V.J., Kalverda, A.P., Homans, S.W. & Radford, S.E. (2005) Dynamics in the unfolded state of beta(2)-microglobulin studied by NMR. *J. Mol. Biol.* **346**, 279-294.

Homans, S.W. (2005) Probing the binding entropy of ligand-protein interactions by NMR. *Chembiochem* **6**, 1585.

Barratt, E.A., Bingham, R.J., Warner, D.J., Laughton, C.A., Phillips, S.E.V. & Homans, S.W. (2005) Van-der-waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water. *J. Am. Chem. Soc.* **127**, 11827-11834.

Perazzollo, C., Wist, J., Poggi, L., Homans, S.W. & Bodenhausen, G. (2005) Effects of protein-pheromone complexation on correlated chemical shift modulations. *J. Biomol. NMR* **33**, 233-242.

Kalverda, A.P., Thompson, G.S. & Turnbull, B.W. (2005) The importance of protein structural dynamics and the contribution of NMR spectroscopy. In The encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics, Wiley Interscience, Dunn, M.J., et al, editors, ISBN 0470849746.

Stockley, P.G., Ashcroft, A.E., Francese, S., Thompson, G.S., Ranson, N., Smith, A., Homans, S.W. & Stonehouse, N.J. (2005) Dissecting the details of assembly of a T=3 phage capsid. *J. of Theoretical Medicine* **6**, 119-125.

**Funding**

We thank the Wellcome Trust and BBSRC for funding

# Discovery of stable, single α-helices in myosins

Peter Knight, Kavitha Thirumurugan, Arnout Kalverda and Michelle Peckham

## Introduction

Myosins are a diverse family of motor proteins. The human genome alone contains about 40 myosin genes, of which about one third are 'conventional' myosins (i.e. the well-studied myosin 2 that drives muscle contraction) and the rest are unconventional, falling into about 10 different classes. The structure, properties and functions of the majority of the unconventional myosins are poorly characterised, and have largely been inferred from sequence comparisons rather than direct experiments on purified proteins. A fundamental question to ask about any unconventional myosin is whether it can dimerise, because only the dimeric molecule would be able to transport cargoes by a succession of steps along its actin track. We know that myosin 2 dimerises through its α-helical coiled-coil tail. Therefore, it has been assumed that an unconventional myosin will also be dimeric if it contains a region predicted to be coiled coil. This assumption depends on the accuracy of coiled-coil prediction programs, such as Coils or Paircoil, which are also used by protein-fold prediction sites on the web such as SMART. Recently, myosin 6 was shown to be monomeric and not processive despite containing predicted coiled coil. We noticed that part of the predicted coiled coil region of myosin 10 is rich in charged amino acids and lacked the hydrophobic residues that favour dimerisation. We have therefore studied the properties of a synthetic peptide of this part of the amino acid sequence and the dimerisation behaviour of myosin 10 itself.

## Structure of the highly charged predicted coiled coil of myosin 10

Residues 800-950 of the myosin 10 sequence have a high score for predicted coiled coil (Fig. 1a). However, in the first part, the expected hydrophobic 'seam' is absent, and instead there are many positive and negative charged side chains ordered in the sequence such that they can form chains of alternating charge if the sequence is wound into an α-helix (Fig. 1b & c). Such intra-helical interactions stabilise a single α-helix against unfolding by salts and water. Purification of a 36-residue peptide containing this sequence by HPLC was guided by MALDI-MS. UV-CD shows it is strongly α-helical, unfolding reversibly at high temperature. Unfolding does not show the cooperative transition seen in a dimeric coiled-coil segment of myosin 2 tail, and the peptide is very resistant to unfolding by salt. NOESY-NMR shows the N-terminal 6 residues are not α-helical, but the rest is. Finally, equilibrium analytical ultracentrifugation shows the helical peptide is monomeric, not dimeric, even at high concentrations (0.7 mM). Thus this part of the predicted coiled coil, despite its high score, is stable, single α-helix (we call it a SAH domain). SAH domains appear to be rare in Nature, but may have escaped detection by structure prediction software.

## Structure of Myosin 10

Electron microscopy of a recombinant myosin 10 in which the C-terminal cargo-binding domains were deleted but the entire predicted coiled coil was still present shows that most molecules are monomeric. Dimers are joined together at their distal tips, rather than by a coiled-coil tail. The heads are longer than expected on the basis of the current dogma, which holds that the head is the sum of just the motor and light chain domains. Thus the predicted coiled coil sequence adds to the head, and we measured that the additional length it produces is similar to that expected if it was all α-helix, *i.e.* the whole predicted coiled coil was a SAH domain. The SAH domain could increase the working stroke of this motor protein beyond that produced by the conventional lever mechanism, where the length of the light chain domain alone determines the working stroke (Fig. 2). Inspection of amino acid sequences of other myosins reveals that myosins 6, 7a and MyoM (from *Dictyostelium*) also contain highly-charged sequences which could have similar properties.

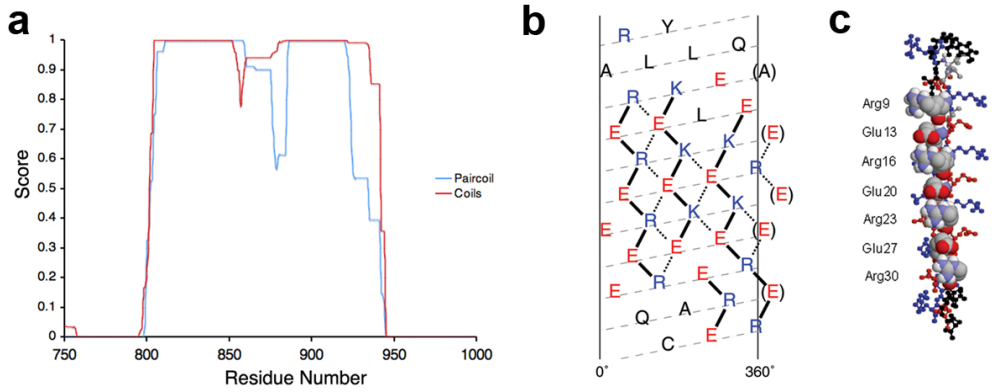



**Fig. 1** *Character of the predicted coiled coil of myosin 10.* **a** shows the high score by prediction software; **b** shows a helical net of the initial part of the high scoring sequence, with solid and dotted lines linking the positive (blue) and negative (red) residues that could form intrahelical ion pairs; **c** shows the sequence modelled as an α-helix, with the left-hand chain of residues from **b** modelled in spacefilling format.

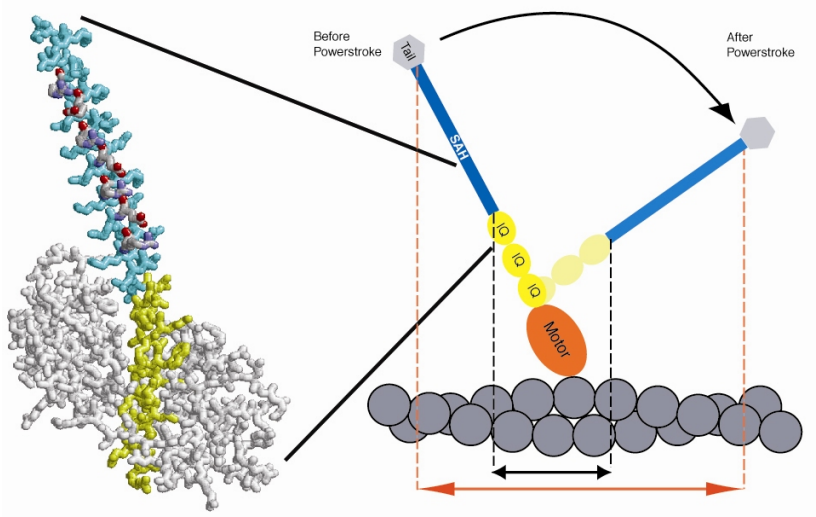



**Fig. 2** *Possible function of the SAH domain in myosin.* Left panel shows the last calmodulin (grey) of the light chain domain attached to the yellow heavy chain, and the SAH domain extending beyond it. The right panel shows how much the conventional axial working stroke (black arrow) could be extended by the predicted coiled coil (red arrow), if present as a SAH domain.

## Publication

Knight, P.J., Thirumurugan, K.M., Xu, Y., Kalverda, A.P., Wang, F., Stafford III, W.F., Sellers, J.R. & Peckham, M. (2005). The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head. *J. Biol. Chem*. **280**, 34702-34708.

## Collaborators

James R. Sellers, Fei Wang and Yuhui Xu, Laboratory of Molecular Physiology, NIH, USA; Walter F. Stafford III, Boston Biomedical Research Institute, MA, USA.

## Acknowledgements

We thank Drs Jeff Keen and Chris Adams for advice on peptide synthesis and purification, Andy Baron for performing the analytical ultracentrifuge runs and Prof. Stephen Baldwin for discussions. Funded in part by The Wellcome Trust and the BBSRC.

# Dissecting the functions of a transcriptional activator of antibiotic production by disrupting its 'reading head' to create a dominant-negative mutant

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

## Introduction

The *Streptomyces* produce a plethora of secondary metabolites including antibiotics and undergo a complex developmental cycle. As a means of establishing the pathways that regulate secondary metabolite production by this important bacterial genus, the model species *Streptomyces coelicolor* and its relatives have been the subject of several genetic screens. However, despite the identification and characterization of numerous genes that affect antibiotic production, there is still no overall understanding of the network that integrates the various environmental and growth signals to bring about changes in the expression of biosynthetic genes. To establish new links, we are taking a biochemical approach to identify transcription factors that regulate antibiotic production in *S. coelicolor*. This year we described the identification and characterisation of a transcription factor, designated AtrA, that regulates transcription of *actII-ORF4*, the pathway- specific activator of the actinorhodin biosynthetic gene cluster in *S. coelicolor*. Disruption of the corresponding *atrA* gene, which is not associated with any antibiotic gene cluster, reduced the production of actinorhodin, but had no detectable effect on the production of undecylprodigiosin or the calcium dependent antibiotic. These results indicated that *atrA* has specificity with regard to the biosynthetic genes it influences. An orthologue of *atrA* is present in the genome of *Streptomyces avermitilis*, the only other streptomycete for which there is a publicly available complete sequence. We also showed that *S. coelicolor* AtrA can bind *in vitro* to the promoter of *strR*, a transcriptional activator unrelated to *actII-ORF4* that is the final regulator of streptomycin production in *Streptomyces griseus*. These findings provided further evidence that the path leading to the expression of pathway-specific activators of antibiotic biosynthesis genes in disparate *Streptomyces* may share evolutionarily conserved components in at least some cases, even though the final activators are not related. They also suggested that the regulation of streptomycin production, which serves an important paradigm, may be more complex than represented by current models.

## Construction and use of a dominant negative mutant

AtrA is a member of the TetR family of transcription factors, which bind DNA as dimers. Recently we constructed integrating plasmids that overproduce either wild-type AtrA or a mutant that binds DNA with greatly reduced affinity. Residues in AtrA that contribute to DNA binding were identified by aligning the known helix-turn-helix, DNA-binding domains of TetR and QacR with the corresponding segment of AtrA (Fig. 1). Subsequently, residues R71 and R72 were disrupted in AtrA.

```
                -------α1-------      ---α2---     --α3---   -α4-
                                     ***          **** ***   *
TetR     RESVIDAALELLN-ETGIDGLTTRKLAQKLGVEQPTLYWHVKNKRALL
AtrA     NLEHVLRAAREVFGELGYGA-PMEDVARRARVGVGTVYRRRFPSKDVLV
QacR     KDKILGVAKELFI-KNGYNATTTGEIVKLSESSKGNLYYHFKTKENLF
                                     *        ***** ***   *
```

**Fig. 1.** Alignment of AtrA, TetR and QacR
Green are residues within turns between helices. Bold residues are conserved in all three sequences. Asterisks mark those residues shown to be involved in DNA-binding by crystallographic studies. For TetR and QacR see Orth *et al.*, 2000. *Nature Structural Biology* **7**: 215-9 and Schumacher *et al.*, 2002. *EMBO J* **21**: 1210-18, respectively. The helix-turn-helix motif is comprised of helices 2 and 3.

When transformed into *S. coelicolor* M145, the integrating plasmid overexpressing wild-type *atrA* caused an increase in actinorhodin production relative to wild-type M145 cells (see strain L646, Fig. 2). Moreover, this phenotype has been correlated with increased transcription of *actII-ORF4* (data not shown). These results indicate that the production of actinorhodin is limited by the availability of AtrA and are consistent with a model in which the regulation of AtrA is at the level of its production. Overproduction of the DNA-binding mutant resulted in a decrease in actinorhodin production relative to M145 indicating that this mutant is dominant (see L644, Fig. 2). The simple explanation for this result is that the overproduction of mutant AtrA results in significant pairing with the chromosomally encoded wild-type polypeptide to produce heterodimers, which because of mutation of one of the reading heads are largely inactive. By introducing this pair of integrating plasmids into different *Streptomyces* species, we are now able to assess the extent to which orthologues of AtrA have a regulatory role in other species as well as using this system to explore the AtrA regulon, and the role of AtrA in known pathways in *Streptomyces coelicolor*. For example, we have preliminary evidence that the overexpression of wild-type AtrA stimulates sporulation, as well as antibiotic production, in *S. lividans*.

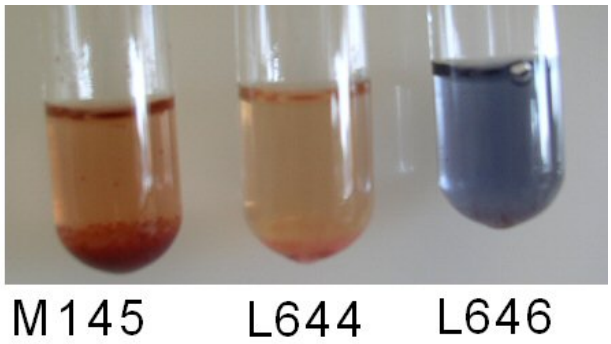



**Fig. 2.** The effects of overexpressing wild-type and a dominant-negative mutant of atrA in *S. coelicolor*. The parental strain is M145. Strains L644 and L646 overexpress the dominant negative and wild-type atrA, respectively. The blue pigment in the culture supernatant is actinorhodin.

**Publication**

Uguru, G.C., Stephens, K.E., Stead, J.A., Towle, J.E., Baumberg, S. & McDowall, K.J. (2005) Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Mol. Microbiol*. **58**, 131-150.

**Funding**

This work is funded by the BBSRC.

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

Stephanie Jourdan, Jonathan Stead and Kenneth McDowall

## Introduction

RNase E is an evolutionarily conserved endoribonuclease that controls the balance and composition of the transcript population in *Escherichia coli* and has the unusual property of being able to modulate its rate of cleavage of substrates depending on the structure of their 5' ends. Our study of this key enzyme culminated this year in the solving of the crystal structures of the catalytic domain with RNA substrates bound and a mutational analysis of residues that contact the 5' end or form the catalytic centre. The crystal structure was solved by the laboratory of our collaborator Dr. Ben Luisi (University of Cambridge). Consistent with previous published findings from our collaboration, the atomic view revealed four subunits associate as an interwoven quaternary structure.

The subdomain encompassing the active site was found to be structurally congruent to a deoxyribonuclease, making an unexpected link in the evolutionary history of RNA and DNA nucleases. In all the complexes, the RNA was found to follow the same stacking arrangement over the surface of a channel. Aside from one hydrogen bond with a base, there appears to be no sequence recognition *per se*. This suggests that the preference of RNase E for A and/or U-rich substrates stems primarily from the recognition of RNA conformation.

## Probing 5'-end recognition

RNaseE cleaves 5'-monophosphorylated substrates more rapidly than the corresponding primary transcripts, which have a terminal 5'-triphosphate group. The crystal structure revealed that the 5' monophosphate is accommodated in a pocket by a semicircular ring of hydrogen-bonding donors (Fig. 1), and that this pocket is physically separated from the catalytic site. This raises the question of how the status of the 5' end of the substrate is relayed to the active site. It seems most likely that the communication is mediated through allosteric change in the protein. In support of this proposal, our collaborators have observed using neutron solution scattering with contrast variation that the binding of RNA induces compaction of the protein. Based on the crystal structure and the solution experiments, a 'mouse-trap'-like mechanism has been proposed. In this model the engagement of the 5' monophosphate of the RNA organises the 5' sensor pocket, which in turn interacts with an S1-like domain causing it to clamp down on the RNA downstream, thus orienting the phosphate backbone to favour an in-line attack on the scissile phosphate. Thus, catalysis proceeds through classical induced-fit. The dual requirement for the simultaneously engaged of the 5' end in the sensing site and an A and/or U-rich segment in the active site may explain why RNase E does not indiscriminately cleave all RNA with a 5' monophosphate, including the structured substrates that it processes, such as the precursors of ribosomal RNA and RNase P.

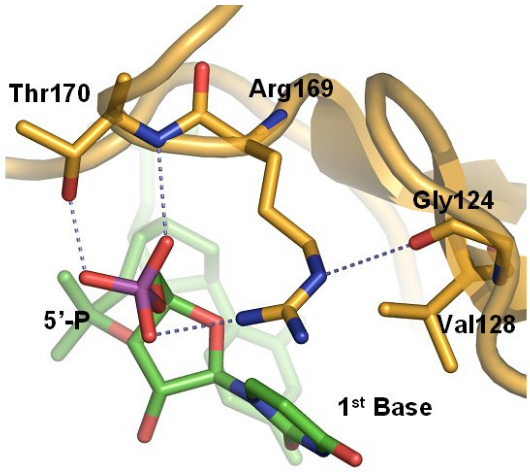



**Fig. 1.** The 5' monophosphate-binding pocket of *E. coli* RNaseE. The phosphate is engaged by a semicircular ring of hydrogen-bonding donors from the side chain and peptide amide of Thr 170 and the guanidino group of Arg 169. The interaction of Arg 169 is consolidated by a hydrogen bond to the peptide backbone of Gly 124 in the neighbouring strand.

The crystal structure of *E. coli* RNase E opens the possibility of developing new anti-infectives and analysing the complex interaction of different components in dynamic cellular systems. Residues that contact the terminal monophosphate group have been disrupted and two mutants identified (V128A and R169K) that appear to be insensitive to 5'-end stimulation. Currently these mutants are being used to test whether 5'-end stimulation does indeed endow on mRNA decay its 'all-or none' nature and 5' to 3' directionality.

**Publications**

Callaghan, A.J., Marcaida, M.J., Stead, J.A., McDowall, K.J., Scott. W.G. & Luisi, B. F. (2005) Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature* **437**, 1187-1191.

**Funding**

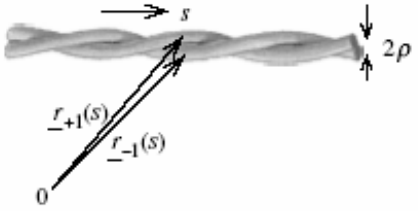
This work was funded by the Wellcome Trust, BBSRC, and the EU.

# Dynamic allostery of the dynein stalk coiled coil

Rhoda Hawkins and Tom McLeish.

## Introduction

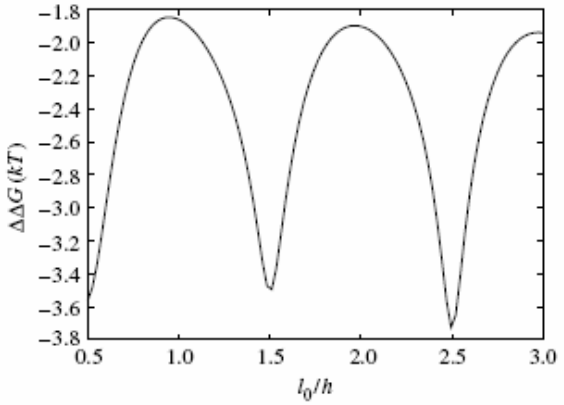
Allosteric signalling is at the heart of protein networks: it is the effect by which the binding of a protein to a substrate molecule may be affected by the simultaneous binding of a second substrate ("co-effector"), often at a site distant from the first. It permits a protein to act as a logical "gate", but poses a puzzle: how do the two binding sites communicate with each other? The canonical explanation invokes binding-induced conformational change that affects the distant site, but increasing evidence suggests that in many cases the information is transmitted not structurally, but dynamically. The dynamic modulation of thermal motion directly affects the entropic component of the free energy of binding of the substrate, so generating allosteric control. One example of such a system is the *lac* repressor dimer



Schematic model for an elastic coiled coil.

## Results for dynein

We have now applied this idea and method to the allosteric control of binding to microtubules of the stalk region in the molecular motor dynein. Studied extensively using electron microscopy by Burgess and Knight at Leeds, there is evidence that modification of the charge state of the dynein head controls the attachment of the distant end of the stalk by moderation of its global dynamics. We model the stalk coiled coil of two α-helices by a coarse-grained elastic system of two rods with specific interactions along their interface. By including the modes of mutual sliding, bending and twisting, we have calculated the effect of altering the local attractive interactions at the stalk head and tip, as well as of its bending and twisting moduli on the free energy of binding to the microtubule.



Predicted allosteric free energy of a coiled coil as a function of length (in units of the pitch). The dynein system lies at the point 1.2 in these units.

Among the predictions of the model is that the allosteric signal is a nearly-periodic function of the length of the coiled-coil. This could be checked by binding assays on dynein mutants, work currently under way in other laboratories.

## Publications

Hawkins, R.J. & McLeish, T.C.B. (2004) Coarse-Grained Model of Entropic Allostery, *Phys. Rev. Letts*, **93**, 098104.

Hawkins, R.J. & McLeish, T.C.B., (2005) Dynamic allostery of protein alpha helical coiled-coils", *J. Roy. Soc. Interface*, **100**, 3143-3148.

## Funding

This work was funded by the EPSRC, we gratefully acknowledge helpful discussions with Stan Burgess and Peter Knight, both of the Astbury Centre.

# High dimensional models of protein folding

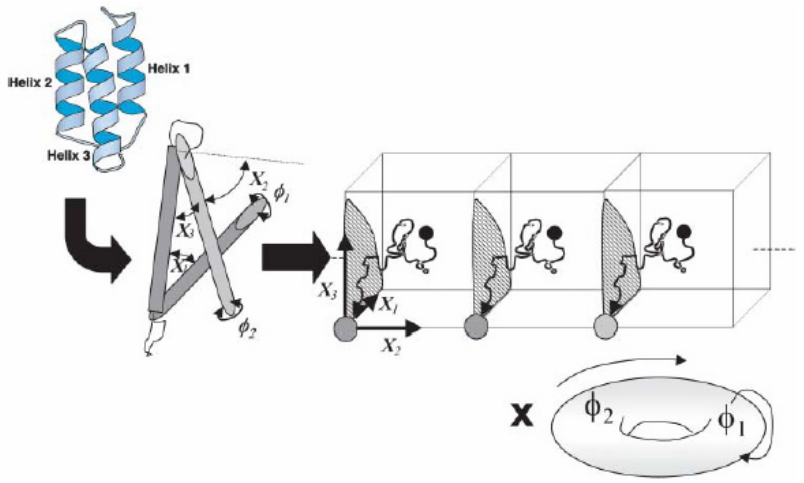
Tom McLeish.

## Introduction

Protein folding is a very active international field, now increasingly concentrating on the way that the folding pathway, or "funnel" is created by the interactions between the residues of a protein. Most of the models currently used "project" the large space of folding interactions down to one or occasionally two degrees of freedom. However, the real situation constitutes a search problem in a space of very high dimension – coming from the hundreds of degrees of freedom in the configurations of a polypeptide. The approach I have taken is to make other simplifications, but to retain the feature of high dimension, to see what predictions this makes for the encoding of folding spaces in proteins of different tertiary structure.

## Three helix bundles

One consequence of the search calculation in very high dimensions is that to find the native state, low-dimensional subspaces need to be stabilised along the way. This may be done with native interactions, but more flexibly with non-native interactions, especially in early stages of folding, and when secondary structured pieces are searching for their correct registry.




Schematic 5-dimensional space for the folding of a three-helix bundle. The rotations of the helix create a 2-torus at each point in the 3-d space of the relative helix orientation.

I have applied the model in detail to the folding of a three-helix bundle (see figure) to find that non-native attractions between the outer helices strongly accelerate folding – and can actually contribute to the measured "Phi-value" of residues even though there is no native association in the effect. The prediction agrees with the pattern found for the bacterial immunity protein Im9 by the Radford group.

A further consequence of the theory is that single exponential kinetics may emerge even when the folding space is by no means "two-state". This is a consequence of all high dimensional diffusive searches.

## Publications

McLeish, T.C.B. (2005) Protein Folding in High-Dimensional Spaces: Hypergutters and the Role of Non-native Interactions, *Biophysical Journal*, **88**, 172–183.

McLeish, T.C.B. (2006) Diffusive searches in high-dimensional spaces and apparent "two-state" behaviour in protein folding, *J. Phys.: Condens. Matter* **18**, 1861-1868.

## Funding

This work was funded by the EPSRC, we gratefully acknowledge helpful discussions with Sheena Radford of the Astbury Centre.

# High yield recombinant production of self-assembling peptides in microbes and plants

Jessica M Riley, Kier James and Michael J McPherson

Organisation of molecules into defined hierarchical structures through non covalent interactions is the foundation of molecular self assembly. In nature, examples of some well studied self-assembling proteins are collagen, keratin, elastin and spidroin (spider silks). Alzeimer's and the prion diseases are also characterised by aggregation of proteins via self-assembly of peptidic sequence motifs.

Exploration of such peptide self-assembly (SA) has opened up an array of applications in nanotechnology, promoting rational design of peptides that self-assemble upon external triggers such as pH, temperature and light.

The biocompatibility and biodegradability of some SA peptides therefore presents an attractive building block for the fabrication of nanoscale devices in both the materials industry and academic research. However, the appliance of these systems will require large quantities and current chemical synthesis of peptide is limited by cost and short sequence length. Large scale chemical production of peptides and proteins is difficult and prohibitively expensive.

The opportunity exists to exploit molecular biology as a production vehicle by using the nucleotide sequences coding for a pre-designed peptide as a building block for cloning, and then expression in a host organism. Exploitation of biological systems as bio-factories for the production of proteins on a large scale is well established. Thus large scale production of a variety of peptides becomes accessible.

Various biological hosts are being tested in the lab such as bacteria, yeast, fungi and plants. Peptide yields from insoluble and soluble expression systems are being compared in *E. coli*. In plants, a designed peptide has been fused with a green fluorescent protein to aid expression screening and also to look at self-assembly *in vivo*.

Purification handles, protease and chemical cleavage mechanisms are being tested, all of which may affect overall peptide yield.

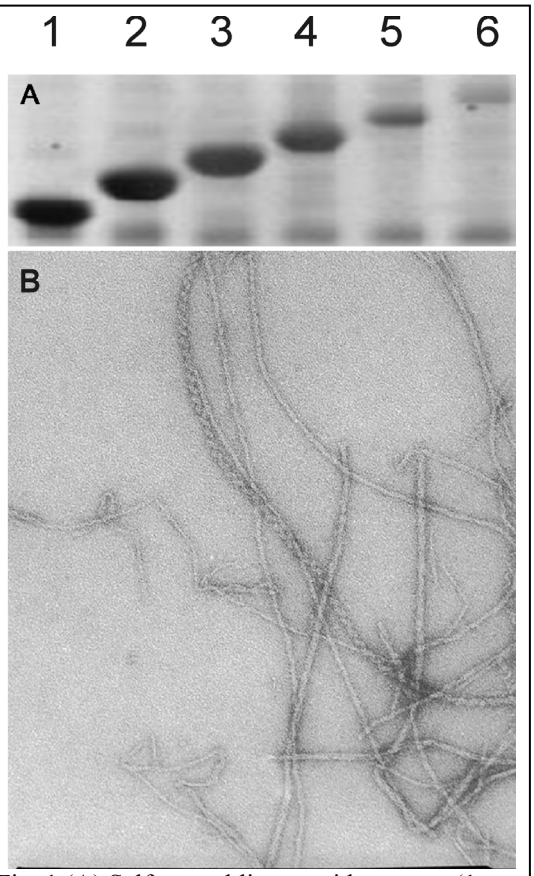


Fig. 1 (A) Self-assembling peptide repeats (1 to 6) expressed in *E. coli* cells. (B) TEM of fibrils formed from a self-assembling peptide.

We acknowledge the help and advice of Dr Amalia Aggeli in the SOMS Centre.

**Funding**

This work was supported by BBSRC studentships to JR and KJ and a CASE studentship from The Dow Chemical Company.

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

Pascale Pirrat, Christian Kurtis, Mark Smith, Simon E.V. Phillips,
Peter J. Knowles and Michael J. 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 homodimeric protein containing one copper ion and a post-translationally modified tyrosine cofactor; 2,4,5-trihydroxyphenylalanine quinone (TPQ). 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 metal in the oxidative half cycle remains elusive, despite numerous proposed mechanisms. Valuable insights into the mechanism of $O_2$ activation and the role of active site and auxiliary metals are key, not only to amine oxidases, but also could provide a paradigm for investigations concerning other more complex metalloenzymes, such as cytochrome *c* oxidase.

To investigate the role of the metal, studies involving removal of copper from ECAO and reconstitution with $Cu^{2+}$, $Ni^{2+}$, $Co^{2+}$ and $Zn^{2+}$ have been recently undertaken. A combination of spectroscopic, crystallographic and steady-state kinetic analyses have been carried out. The results suggest that metal redox chemistry is not required for ECAO activity. Kinetic studies of the metal reconstituted ECAOs implied that calcium plays an important role in ECAO activity. In ECAO two calciums (Ca) have been identified per monomer. The buried calcium and its ligands lie at the ends of 2 antiparallel β-strands, the other ends being occupied by the 3 histidines that coordinate the copper ion in the active site (Fig. 1). The same connection between copper and the non active site metal is observed for the other amine oxidases. Furthermore, along each β-strand, many residues are conserved. Mutagenesis targeting these strictly conserved Ca ligands has been carried out and kinetic and crystallographic studies are under way to elucidate the role of Ca.

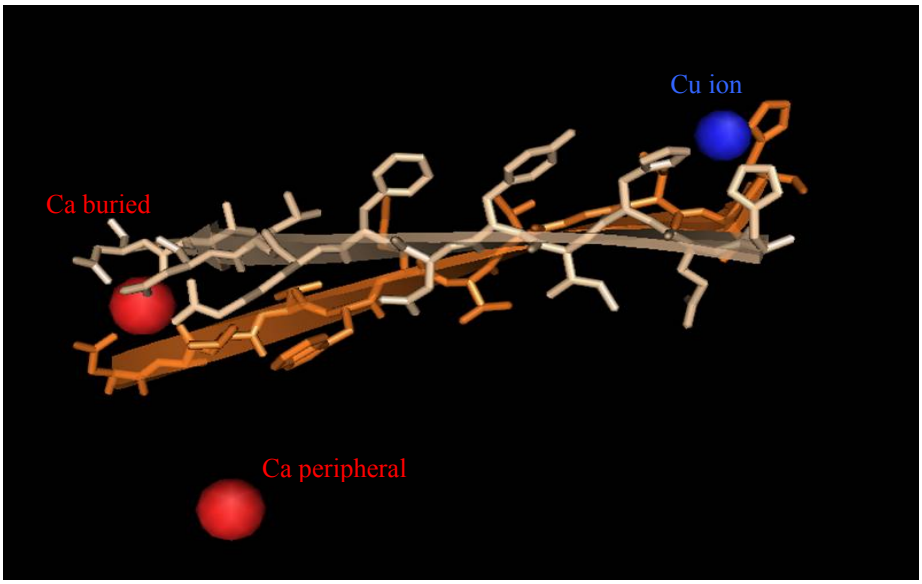



**Fig 1**: Illustration of the structural relationship between the buried Ca and the active site Cu.

**Funding**
This work was funded by the BBSRC and the Wellcome Trust.

# Galactose oxidase, processing, substrate specificity and directed evolution studies

Sarah Deacon, Nana Akumanyi, Melanie Rogers, Peter Knowles, Malcolm Halcrow, 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 (Fig. 1). GO catalyses the stereo- and regio-specific oxidation of a broad-range of primary alcohols, from small aliphatics to large poly-saccharides, generating their corresponding aldehydes with associated reduction of oxygen to hydrogen peroxide, as in the reaction scheme:

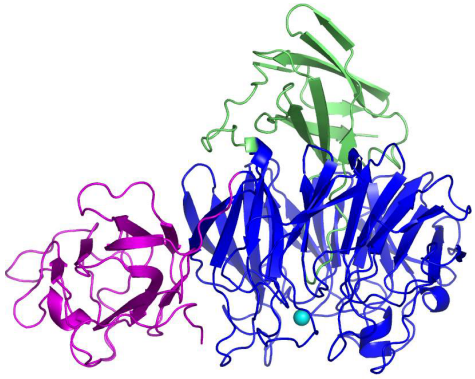


**Fig. 1.** Ribbon diagram showing overall fold of GO, Domain 1 is shown in magenta, Domain 2 in blue and Domain 3 in green, copper is shown in cyan.

$$RCH_2OH + O_2 \longrightarrow RCHO + H_2O_2$$

## Substrate binding studies

Modelling studies have shown a binding site complementary to D-galactose (Fig. 2.). We demonstrated that the mutations F464A, R330A and R330K resulted in increased $K_M$ for D-galactose, indicating the importance of these residues in substrate binding. However, R330K shows a switch in specificity towards D-fructose, potentially useful in the sugar industry. C383S, highlighted from previous directed evolution screens, is currently being examined through structural and saturation mutagenesis studies.

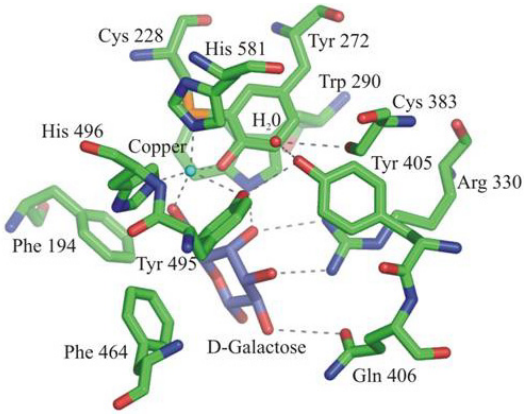



**Fig. 2.** Model of substrate binding in GO

## Directed evolution

The enzyme is used in a number of medical and commercial applications, but has great potential for novel applications in generating synthetic chemistry intermediates, improved medical sensors and as a green industrial catalyst. A major obstacle for directed evolution studies is developing a host expression system suitable to generate extensive mutant libraries. Functional expression of soluble GO was not possible in *E. coli* until recently. We have now developed a system for high level expression of functional GO by applying a combination of approaches, such as the introduction of silent mutations and optimising host strain and culture conditions. We are now using this system to develop mutational libraries to generate enzymes with novel specificity, enhanced activity under standard and altered pH, organic solvent and elevated temperature conditions.

## Collaborators

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

## Funding

This work was sponsored by the BBSRC.

# Applications of synthetic organic chemistry to biological problems

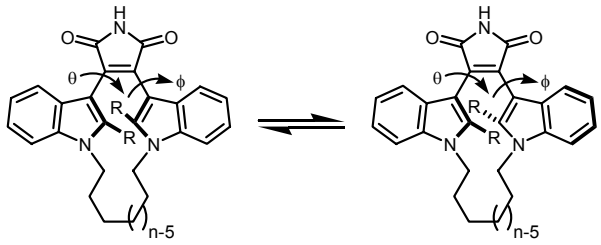
Steve Bartlett, Alan Berry, Mandy Bolt, Blandine Clique, Jacqui Colley, Lorna Farnsworth, Alan Ironmonger, Richard Jackson, Catherine Joce, Adam Nelson, Peter Oledzki, Alexis Perry, Peter Stockley, James Titchmarsh, Stuart Warriner, Ben Whittaker, Gavin Williams and Tom Woodhall

## 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 of our published work in 2005 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 conformationally diverse bisindolylmalemides

We have used a range of experimental and theoretical approaches to determine the conformational properties of constrained bisindolylmaleimides **1**. We have shown that the population of the limiting *syn* and *anti* conformations varies as a function of the tether length n, and that this family of compounds is conformationally diverse. The family of compounds is a useful tool for comparing the active sites of the protein kinases; this chemical approach requires no structural knowledge and yet still concentrates on the comparison on the proteins' active sites.




***syn*-1**
θ and ϕ have opposite signs

***anti*-1**
θ and ϕ have the same sign

**Figure 1**: Limiting conformations of bisindolylmaleimides

In other collaborations, our bisindolylmaleimides are being exploited as chemical tools for dissecting the mechanisms of stem cell differentiation and for structural studies of protein kinases.

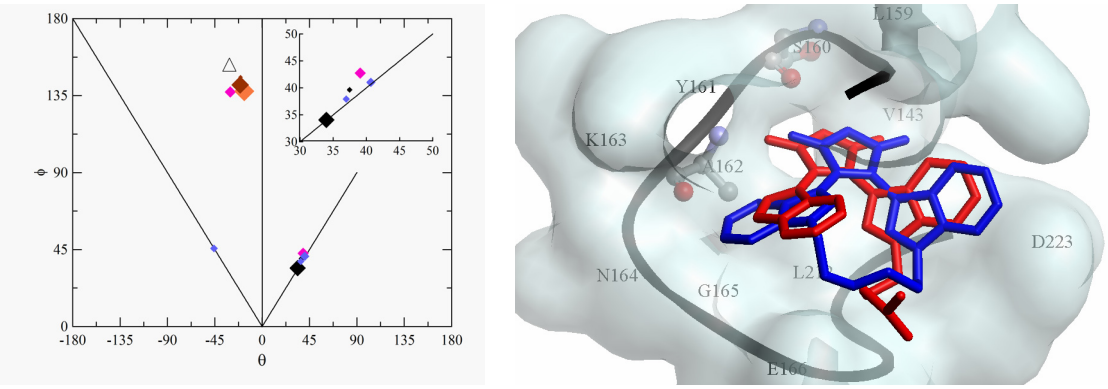



**Fig. 2.** Left panel: Conformational diversity of bisindolylmaleimide cyclophanes. Right panel: Docked bisindolylmaleimide cyclophane within the context of a kinase.

## A two-directional approach to aza-*C*-linked disaccharide mimetics

We have developed a two-directional approach to the synthesis of aza-*C*-linked disaccharide mimetics. Unusually, for a two-directional approach, the target molecules do not possess any hidden symmetry. The approach is highly general, and can be applied to the synthesis of a wide range of disaccharide mimetics; for example, complementary conditions were identified for the two-directional functionalisation of **2** to give the protected mimetics **3Bd'** and **3Ad'**

(Fig. 3). This work was published as the inaugural paper in *Beilstein Journal of Organic Chemistry*.

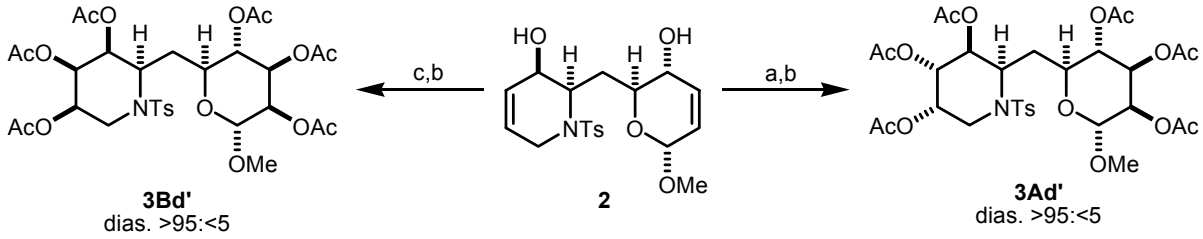



**Figure 3.** Complementary two-directional synthesis of aza-*C*-linked disaccharide mimetics
Conditions: a) NMO, cat. $OsO_4$, acetone-$H_2O$; (b) $Ac_2O$, pyridine; (c) $OsO_4$, TMEDA, $CH_2Cl_2$, -78 **°C**

In a related synthetic study, we have also prepared a library of diverse aminoglycoside derivatives which we have exploited as powerful chemical tools for probing the functions of RNA molecules.

**Parallel synthesis of a library of sialic acid mimetics using a variant aldolase**

In collaboration with Alan Berry, we have directed the evolution of a range of aldolases with catalytic activity of value to the synthetic chemist. For example, we have evolved a variant of sialic acid aldolase with sufficiently broad substrate specificity for application in the parallel synthesis of a range of precursors of influenza A sialidase inhibitors. Ozonolysis of the alkenes **5**, and variant aldolase-catalysed C-C bond formation yielded a small library of mimetics (Fig. 4). In related work, we have evolved a pair of complementary aldolases which catalyse the selective formation of either of the diastereoisomeric mimetics **6** and **7**.

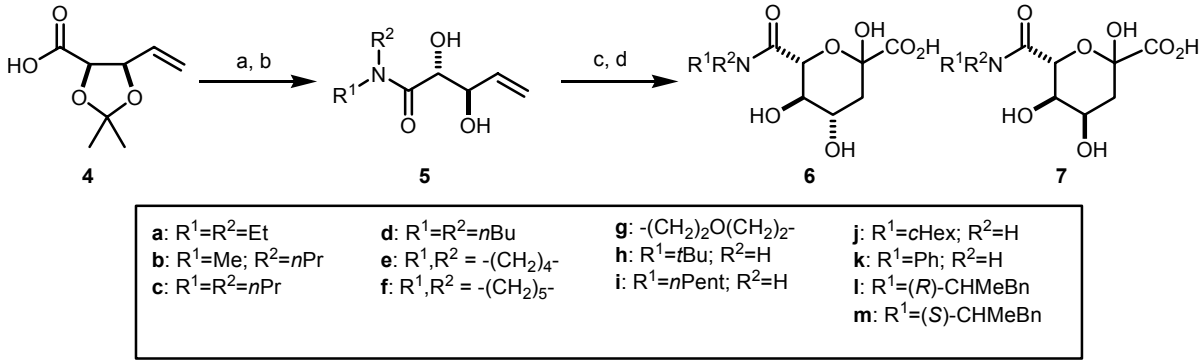



**Figure 4.** Parallel synthesis of the sialic acid mimetics **6**/**7**. Reagents and conditions: a) EDC, HOBt, $R^1R^2NH$, $CH_2Cl_2$; b) 1:1 TFA–$H_2O$; c) $O_3$, MeOH then $Me_2S$; d) 2 x $10^{-2}$ mol% E192N, pyruvate, pH 7.4 buffer.


## Acknowledgements

ASN holds an EPSRC Advanced Research Fellowship (2004-2009), and thanks AstraZeneca for the award of the 2005 AstraZeneca Research Award in Organic Chemistry. 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), Simon Barrett, Professor Godfrey Beddard, Dr Veysel Kayser, Colin Kilner, Dr Gavin Reid (University of Leeds), Dr Andrew Leach (AstraZeneca), Professor Peter Parker (Cancer Research UK), Dr Stefan Knapp and his collaborators at the Structural Genomics Consortium and Andrew Kennedy (GSK).

## Publications

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.* **127**, 11699-11708.

Bullock, A. N., Debreczini, J. É., Fedorov, O. Y., Nelson, A., Marsden, B. & Knapp, S. (2005) "Structural basis of inhibitor specificity of the human proto-oncagene proviral insertion site in Moloney Murine Leukaemia Virus (PIM-1) kinase", *J. Med. Chem.* **48**, 7604-7614.

Barrett, S., Bartlett, S., Bolt, A., Ironmonger, A., Joce, C., Nelson A. & Woodhall T. (2005) "Configurational stability of bisindolylmaleimide cyclophanes: From conformers to the first configurationally stable, atropisomeric bisindolylmaleimides", *Chem. Eur. J.* **11**, 6274-6283.

Kennedy, A., Perry, A. & Nelson, A. (2005) "Methods for the synthesis of polyhydroxylated piperidines by diastereoselective dihydroxylation: Exploitation in the two-directional synthesis of aza-*C*-linked disaccharide derivatives", *Beilstein J. Org. Chem.* **1**, 2.

Kennedy, A., Perry, A. & Nelson, A. (2005) "A general, two-directional synthesis of aza-(1→1')-*C*-disaccharide mimetics", *Chem. Commun.*, 1646-1648.

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

Ironmonger, A., Stockley, P. & Nelson, A. (2005) "Asymmetric double ring-opening of a $C_{2h}$-symmetric bisepoxide: Improved enantiomeric excess of the product through enantioselective desymmetrisation and 'proof-reading' steps", *Org. Biomol. Chem.*, **3**, 2350-2353.

Woodhall, T., Williams, G., Berry, A. & Nelson, A. (2005) "Creation of an aldolase for the parallel synthesis of sialic acid mimetics", *Angew. Chem., Int. Ed.*, **44**, 2109-2112.

Woodhall, T., Williams, G., Berry, A. & Nelson, A. (2005) "Synthesis of screening substrates for the directed evolution of sialic acid aldolase: Towards tailored enzymes for the preparation of influenza A sialidase inhibitor analogues", *Org. Biomol. Chem.* **3**, 1795-1800.

Williams, G. J., Woodhall, T., Nelson, A. & Berry, A. (2005) "Structure-guided saturation mutagenesis of *N*-acetylneuraminic acid lyase for the synthesis sialic acid mimetics", *Prot. Eng. Des. Sel.* **18**, 239-246.

# Theoretical and computational studies of mechanical unfolding of proteins

Daniel West, Emanuele Paci and Peter D Olmsted

Over the past year we have used theoretical tools and molecular dynamics simulations in particular, to help understanding the important molecular features that govern the folding behaviour of proteins. One particularly exciting development concerns the interpretation of single molecule experiments which, thanks to techniques such as dynamic force spectroscopy, are now beginning to shed new light on the 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 (eg. see Report by Anderson *et al* and manuscripts cited below).

**Mechanical resistance of proteins explained using simple molecular models**

Recent experiments have demonstrated that proteins unfold when two atoms are mechanically pulled apart, and that they unfold differently under force compared with thermal or chemical denaturation. Experiments have also shown that the response of proteins to external forces is very diverse, some of them being "hard" and some others "soft". Mechanical resistance originates from the presence of barriers on the energy landscape; together, experiment and simulation have demonstrated that unfolding occurs through alternative pathways when different pairs of atoms undergo mechanical extension. In this work we used simulations to probe the mechanical resistance of six structurally diverse proteins when pulled in different directions. For this, we used two very different models: a detailed, transferable one, and a coarse-grained, structure-based one. The coarse-grained model gives results that are surprisingly similar to the detailed one and qualitatively agree with experiment; *i.e.*, the mechanical resistance of different proteins or of a single protein pulled in different directions can be predicted by simulation. The results demonstrate the importance of pulling direction relative to the local topology in determining mechanical stability, and rationalize the effect of the location of importation/degradation tags on the rates of mitochondrial import or protein degradation *in vivo*.

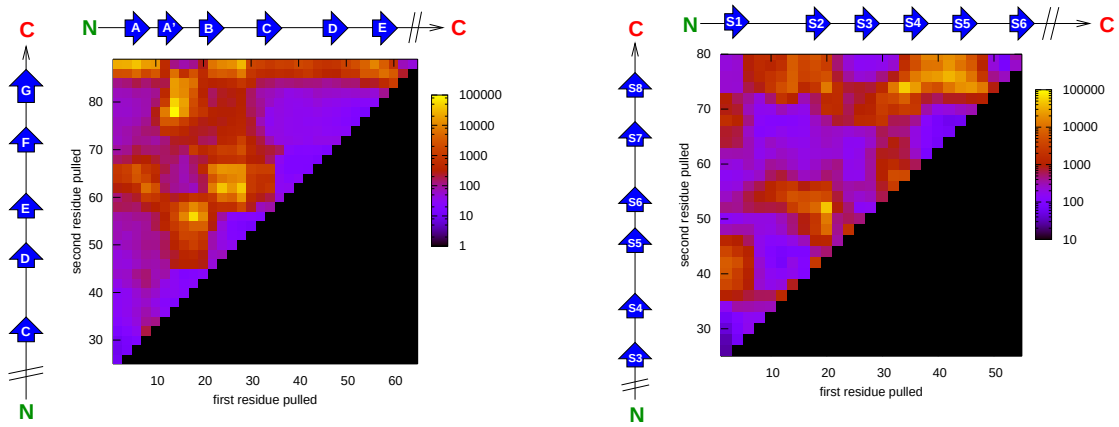



**Fig. 1** These mechanical landscapes are among the most intriguing results that can be obtained with such coarse-grained models. They reveal the anisotropy of the response to a force of proteins. The landscapes show the average unfolding time for a constant force applied to any two residues within a single protein. It is thus clear that classifying a protein of hard or soft is not meaningful. Each protein can be soft or hard depending on how it is handled.

**Exploring non-equilibrium, equilibrium and minimum energy pathways in unfolding**

Single molecule experiments and their application to probe the mechanical resistance and related properties of proteins provide a new dimension in our knowledge of complex biological molecules. Single molecule techniques may not have yet overridden solution

ensemble experiments as a method of choice to characterize biophysical and biological properties of proteins, but they have stimulated a debate and contributed considerably to bridging theory and experiment. Here we demonstrate this latter contribution by illustrating the reach of some theoretical findings using a solvable but non-trivial coarse-grained molecular model whose properties are analogous to those of the corresponding experimental systems. We use molecular dynamics simulations to unfold protein domains both thermally and mechanically. We examine the pathways for both the thermal and mechanical unfolding processes in terms of several different candidate reaction coordinates, such as native contacts, root mean square displacement, N-C distance, etc. We then correlate these pathways with quantities such as the mechanical resistance for different applied forces (*e.g.* the time taken to unfold).

We have thus shown some of the relationships between the thermodynamic and the mechanical properties of a protein. In addition, we have demonstrated how forced and spontaneous unfolding occur through different pathways, by examining candidate reaction coordinates; and we have demonstrated that folding (or unfolding) rates at equilibrium cannot in general be obtained from forced unfolding experiments or simulations. We have also studied the relationship between the energy surface and the mechanical resistance of a protein and shown how a simple analysis of the native state can predict much of the mechanical properties of a protein.

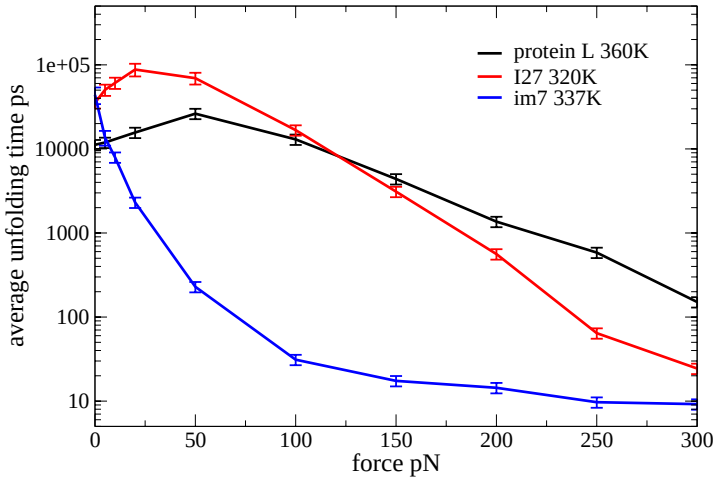



**Fig. 2** Here we show the average unfolding time (over 100 independent simulations) when a force is applied to the C- and N-termini of the protein. The constant applied force accelerates the rate of thermally activated unfolding of the protein. For standard AFM experiments only the large forces regime (>100 pN), where the logarithm of unfolding time is linear in the force, is accessible. Smaller forces are accessible by, e.g., optical tweezer experiments. It was initially believed that, by extrapolating the experimental unfolding time measured using the AFM to zero force, the equilibrium unfolding time in solution could be obtained. This would have also meant that the mechanism of unfolding does not depend on the presence or absence of a pulling force. By simulation, we have discovered a turnover in the unfolding time at very low forces for I27 and protein L. This means that these proteins unfold faster than in the absence of a force when a sizable pulling force is applied, but it might unfold **slower** when a very weak force is applied. This result can be easily explained if one considers that the folding mechanism is different in the presence or in the absence of a force. In the very low force regime both mechanisms are possible, but the very weak force frustrates the thermal mechanism and hence decreases the unfolding rate. At higher forces the mechanical mechanism takes over, and thereafter the unfolding rate increases with increasing force.

## Collaborators

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

## Publications

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

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

# Computational modelling of amyloidogenic peptides

Geraint Thomas, Oliver Clarke, Malcolm McLean and Martin Parker

## Introduction

Amyloid fibril formation is associated with a number of severe human dementias, including Alzheimer's and the so-called prion disease. A definitive explanation of the root cause and molecular mechanism of amyloid fibril formation remains elusive. The general consensus is that the respective proteins or peptide fragments transform into a cross-β-sheet amyloid fibril through an associative mechanism involving intermediates with a predominantly β-sheet conformation. An attractive therapeutic strategy therefore is to create β-sheet 'breaker' peptides, based on the wild-type amyloidogenic peptide sequence. An effective approach is the incorporation of *N*-methyl amino acids: one side of the peptide presents backbone NH groups able to hydrogen bond to the amyloidogenic peptide, while on the other side, methylation prevents hydrogen bonding. Such derivatives have the advantages of high proteolytic resistance, solubility and blood-brain barrier permeability. For improving potency, delivery and metabolic resilience, further peptidomimetic strategies will need to be pursued, e.g. employing cyclic peptides and unnatural amino acids. Computational methods will be crucial here for testing and leading rational approaches for increasing binding affinity and specificity of therapeutic peptides.

## *In silico* simulations - towards improved therapeutic peptides

The various approaches adopted towards simulating amyloidosis fall into two camps. Some run simulations of continuous, full atom models. Whilst physically realistic, with current computational power, only an infinitesimal fraction of the conformation space can be sampled. Indeed, as these simulations do not reach equilibrium, they have to be started from putative models of amyloid oligomers or fibres. Others use greatly simplified models, where atoms are unified into single beads. These methods reduce the size of conformation space, thus facilitating more extensive simulations. However, whilst useful for exploring general principles, the lack of atomic detail precludes their use in drug design.

Accurate *ab initio* thermodynamic calculations necessary for elucidating the efficacy of therapeutic peptides to inhibit amyloid formation require a reduced, yet realistic physico-chemical force field, and a method for directly sampling the Boltzmann distribution of conformation space. Our lab is making real progress on both fronts.

## A highly effective force field

The force field we have developed ('PETAL') pertains to a full heavy atom description of the protein with the solvent treated implicitly. To facilitate extensive sampling of conformation space, the model uses ideal covalent bond angles and lengths, along with discrete backbone and side chain torsion angle sets. PETAL contains van der Waals, solvation, hydrogen bond, salt bridge and torsional potentials. These are semi-empirical; experimental data are used to parameterise them, while the functional forms have a sound physico-chemical basis. To be practicable for *ab initio* simulations, we have strived to make the calculations required simple to reduce the overall CPU time as much as possible. Our approach for calculating local dielectrics, for example, faithfully reproduces the energies of 222 non-equivalent salt bridges calculated using the finite difference Poisson Boltzmann method (the gold standard for environment-dependent electrostatics) whilst being ca. $10^6$-fold faster to calculate.

Protein decoy sets are the current acid test for assessing the performance of force fields. A protein decoy set contains the native, experimental structure together with several thousand non-native decoys. A challenging set has recently been proposed comprising 47,301 non-

native structures for 35 proteins. An effective force field must clearly discriminate the native structure from the decoys for all proteins. In this respect, PETAL yields impressive results (Fig. 1). It has superior discriminatory power relative to standard molecular mechanics force fields and the Rosetta force field (Rosetta is currently considered the best *de novo* protein fold prediction method). Moreover, PETAL is very much faster.

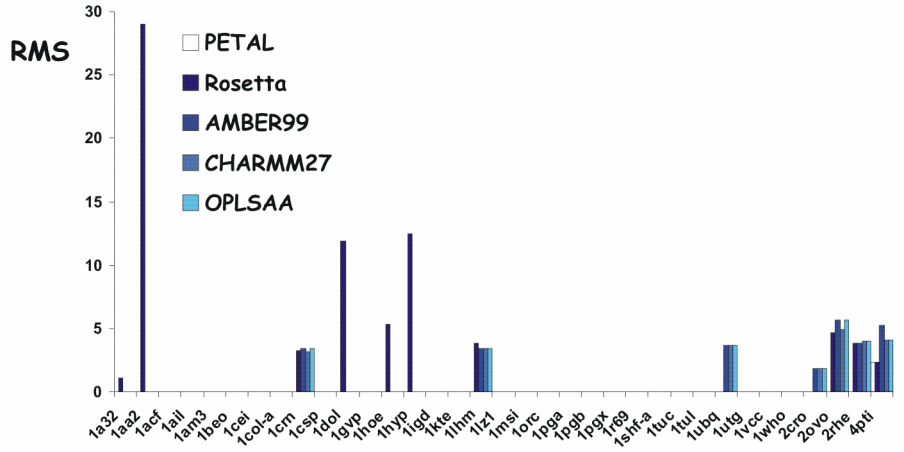



**Fig. 1.** The column chart shows the RMS deviation in heavy atom coordinates, relative to the experimental structure, of the lowest energy conformation among the set of decoys for each protein (identified by PDB ID), calculated using the molecular mechanics force fields AMBER99, CHARMM27 and OPLSAA, the Rosetta force field, and our PETAL force field. An RMS value of zero means that the force field correctly identifies the experimental structure as the lowest energy conformation.

**A highly efficient sampling method**

The above demonstrates that the use of idealised bond geometries and discrete torsion angles in PETAL does not compromise accuracy, while massively reducing the search problem. Still, calculating the Boltzmann distribution of the conformation space necessitates the use of a sampling method. The standard Metropolis Monte Carlo method is inefficient at sampling conformation space. Under physiological conditions the simulations encounter tunnelling problems; trajectories between statistically important energy minima have low transition probabilities. This leads to inaccuracies, as equilibrium is not reached. We have developed a novel Density Guided Importance Sampling (DGIS) method. This method actively 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 is therefore very efficient, and yields results of high accuracy, outperforming all other current sampling algorithms.

**Accurate predictions**

We have applied PETAL and the DGIS method to a set of α-helical and β-hairpin peptides for which high-resolution experimental structures are available. As can be seen in Fig. 2, the calculated and experimental structures agree well in each case. We have also applied PETAL and the DGIS method to the self-association of amyloidogenic peptide fragments of the Alzheimer's peptide (Aβ16-22), running our simulations in virtual space. The time-averaged structure calculated is a β-sheet (Fig. 2), which is entirely consistent with this peptide's β-amyloid forming properties, as studied by solid state NMR. A more impressive illustration of prediction accuracy at the atomic level is provided by simulations of the self-association of the Sup35 yeast prion peptide, where both cross-strand hydrogen bonding and cross-sheet side chain interactions closely mirror those seen in the crystal structure (Fig. 2).

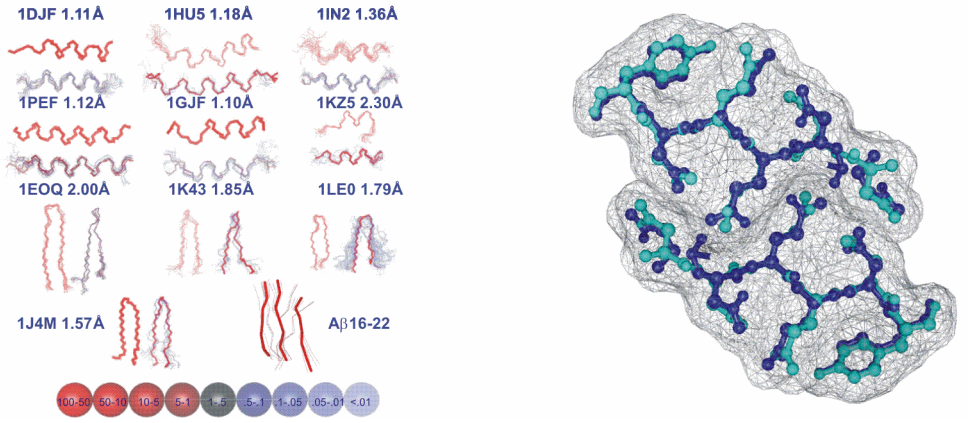



**Fig. 2.** LHS: experimentally-determined peptide backbone structures (top and LHS) along with the ensemble structures calculated using PETAL and DGIS (bottom and RHS). RMS values between the experimental and predicted structures are shown. The ensemble structure calculated for the self-association of three Aβ16-22 peptides is also shown. Individual structures within the ensembles are superimposed, and coloured and rendered according to their probabilities. RHS: atomic structure of the 'steric zipper' unit from the Sup35 yeast prion peptide amyloid, viewed down the sheet (cyan). The corresponding unit of structure obtained from a simulation using PETAL and DGIS has been superimposed (blue). The RMS deviation in heavy atom coordinates between these corresponding units is 0.8Å

## Future directions

Evidently, our approach to computational modelling of amyloidosis holds real potential, and in this regard it is important to stress that, unlike all other current methods, our simulations are both fully reversible (i.e. reach equilibrium) and atomically detailed. We are now using experimental data collected for a large number of peptides and their sequence variants, which pertain to both folding and self-association, to parameterise PETAL to enable accurate thermodynamic calculations under various conditions. This is vital for drug development. We are also refining our simulation methods to increase computation speed as much as possible. Our ultimate aim is to evolve potent peptide-based inhibitors *in silico*.

## Publications

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

Thomas, G., 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 MRC. M.J.P. is a BBSRC David Phillips and University Research Fellow.

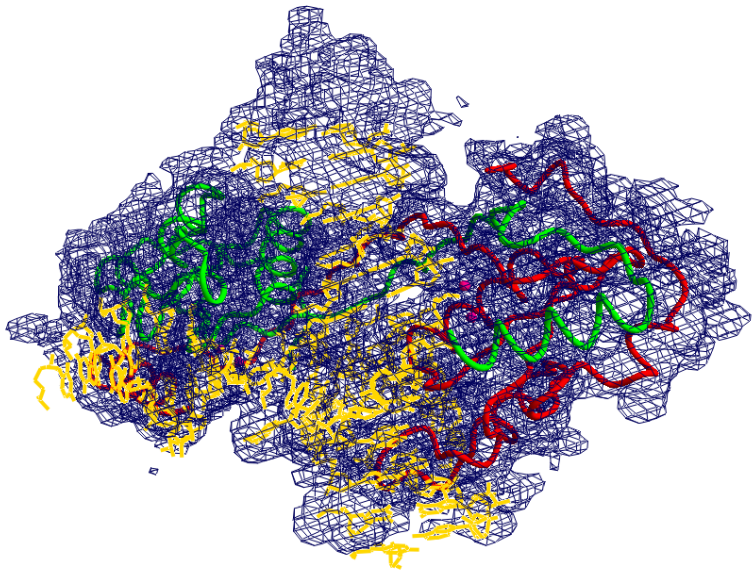
# Towards the structure of a bacteriophage T7 endonuclease I / Holliday Junction complex

Jonathan M. Hadden and Simon E.V. Phillips

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

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

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



We have recently grown crystals of a stable complex of endonuclease I and a synthetic Holliday junction and have collected X-ray diffraction data to 4.1Å resolution using station 14.2 Daresbury SRS. The structure of this complex has been solved using molecular replacement using the structure of free endonuclease I as the search model.

**Fig. 1:** Crystal structure of the complex between T7 endonuclease I and a synthetic Holliday junction. The electron density map is displayed in blue and Endonuclease I subunits are shown in red and green, DNA is shown in gold.

We intend to increase the resolution of our current structure by collecting data from optimised crystals at the European Synchrotron Radiation Facility (ESRF).

**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.

# Crystallographic studies of RNA binding discrimination between bacteriophages MS2 and Qβ.

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

## Introduction

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

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

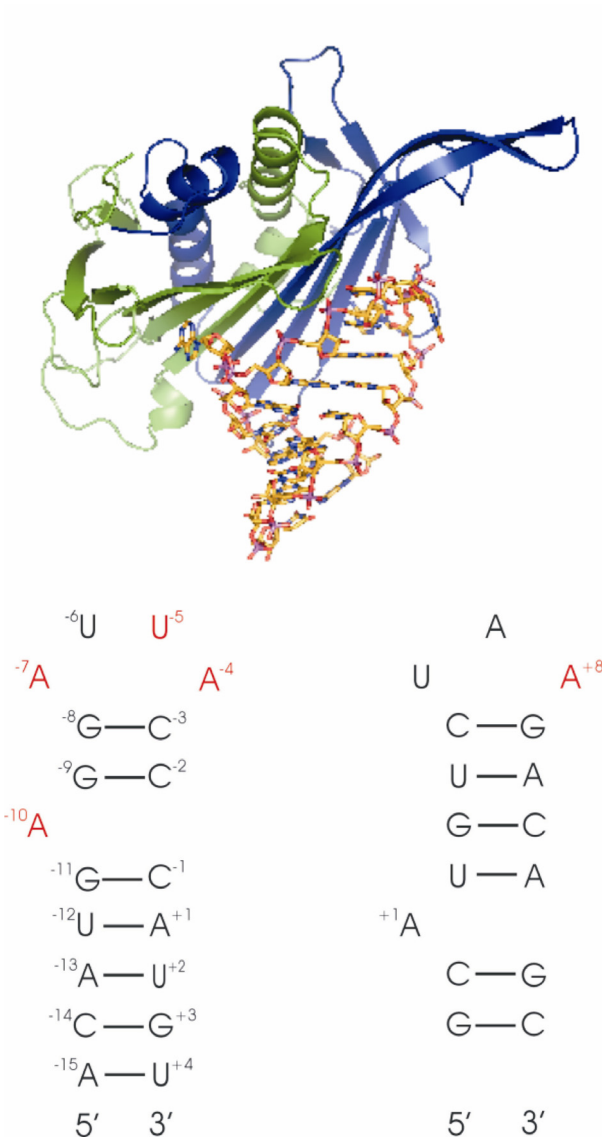



**Fig. 1;** (Top) Structure of the WT MS2 RNA stemloop-coat protein dimer complex, protein subunit A (blue) and subunit B (green).
(Bottom) Diagram of the secondary structure of the MS2 (left) and Qβ (right) stemloop operators. Residues that have been shown to be important for high affinity binding to capsid protein are highlighted in red.

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

*In vivo*, each bacteriophage preferentially discriminates against binding the stemloop operator of the other. Affinity binding studies have, however, identified specific coat protein mutants

of MS2 at residues 87 and 89 that overcome this discrimination mechanism, some of the mutations allowing the binding of the Qβ RNA operator to MS2 mutant capsids with an affinity comparable to that of the wild type MS2 operator. In order to gain new insight into this discrimination mechanism, Qβ RNA stemloop operators were soaked into pre-crystallised MS2 mutant capsids and the structure of the capsid/RNA complexes determined via X-ray crystallography.

## Results

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

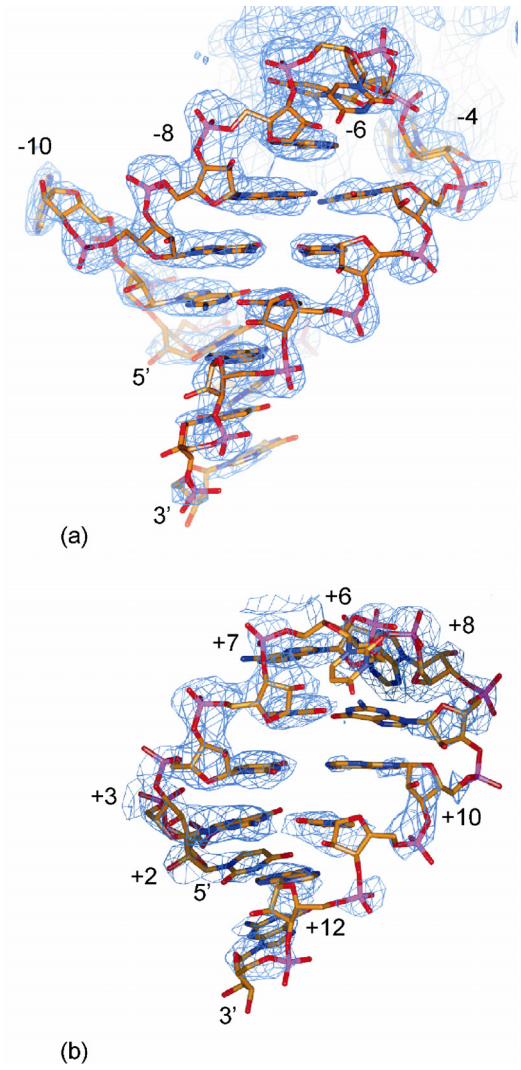



Fig. 2.; Images of (a) the MS2 and (b) the Qβ RNA stemloops modelled into the *2Fo-Fc* electron density, shown in blue and contoured at 1 R.M.S., in the N87S and N87S,E89K mutant capsids respectively.

## Publications

Horn, W.T., Tars, K., Grahn, E., Helgstrand, 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 Qβ and MS2. *Structure.* **14,** 1-9.

## Collaborators

Lars Liljas and Kaspars Tars, Uppsala University, Sweden

Funding from BBSRC.

# Crystal structure of *N*-acetylneuraminic acid lyase mutants

Chi H. Trinh, Alan Berry, Adam Nelson, Gerard Huysmans,
Malavika Ramaswamy and Simon E.V. Phillips

## Introduction

Sialic acid (*N*-acetyl-*D*-neuraminate or *N*-acetylneuraminic acid) is a nine carbon sugar molecule that is an essential component of cellular oligosaccharides, often occurring as the terminal carbohydrate on glycoconjugates that coat cell surfaces. Since this molecule is directly involved in modulating host-pathogen interactions, sialic acid analogues have been identified as potential pharmaceutical agents that could be used against bacteria and viruses. In attempts to produce these analogues, *de novo* synthesis has proved complex and costly. Sialic acid lyase can be used to catalyze the aldol condensation of pyruvate and *N*-acetyl-*D*-mannosamine to produce sialic acid, but attempts to harness the wild-type enzyme for synthesis have yielded poor results. We have, therefore, engineered the wild type enzyme, using a combinatorial approach of rational design and directed evolution. This has created several stereoselective sialic acid lyase variants with varying substrate specificities. Crystallographic analysis has been employed to elucidate the basis of substrate specificity and modified stereoselectivity of these variants.

## Crystallographic studies

This study focuses on the sialic acid lyase mutants E192N/T167G (the most stereoselective enzyme for the 'S' configuration of the product) and E192N/T167V (which shows a preference for the 'R' configuration). Two crystal structures of the E192N/T167G variant have been solved at 2.2Å; as a free enzyme and in complex with pyruvate. In addition the E192N/T167V mutant has been solved in complex with pyruvate.

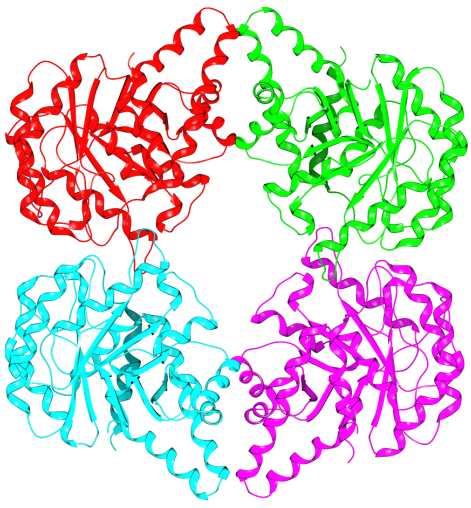
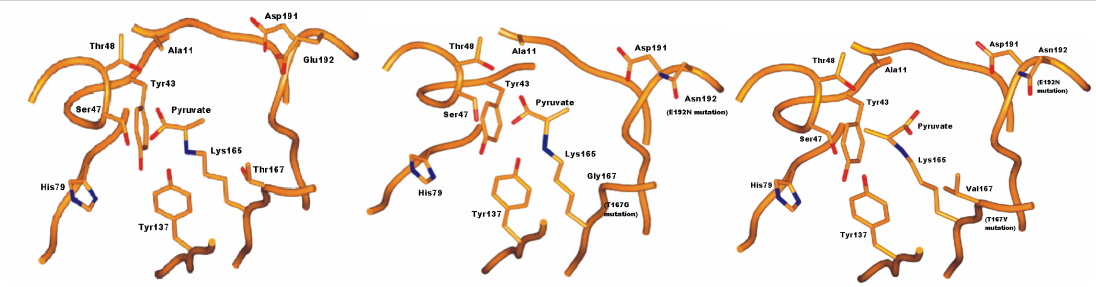


**Fig. 1.** A ribbon representation of sialic acid lyase. Each monomer subunit is shown in a different colour

All X-ray diffraction data were collected using synchrotron radiation at the Daresbury Laboratory at a temperature of 100 K. The structures were determined by the method of molecular replacement using the structure of wild type sialic acid lyase structure (Fig. 1) (Izard *et al*, 1994) as the starting model. Model building and refinement was accomplished using the programs O and programs from the CCP4 suite, respectively.

Sialic acid lyase is a tetrameric protein. Each monomer subunit consists of a $(\alpha/\beta)_8$ barrel fold with the eight inner β-strands surrounded by eleven α-helices; the standard eight helices plus a three helix extension. The active site is located in a deep pocket at the carboxy-terminal end of the $(\alpha/\beta)_8$ barrel fold.

Comparison of the structures of the wild type, and the E192N/T167G and E192N/T167V variants, each complexed with pyruvate, have shown that while pyruvate binding is apparently identical in the E192N/T167G and the wild-type enzymes, there is a possible change in orientation of pyruvate in the E192N/T167V mutant. Further work will be needed to assess the significance of the orientation of the pyruvate within the active site and the stereo-selectivity of this enzyme.




A B C

**Fig. 2.** Representations of the active sites of (A) wild-type NANA lyase, (b) E192N/T167G and (C) E192N/T167V each in complex with pyruvate. In each case, a pyruvate molecule is seen covalently linked to the $N^{\varepsilon}$ of the active site lysine. Several other residues lining the enzyme active site are shown as labelled. The pyruvate in T167V is observed to be oriented differently as compared to pyruvate in the wild type and E192N/T167G active sites.

**Funding**
We thank the BBSRC, EPSRC and Wellcome Trust for funding.

# Structure and redox-dependent regulation of a fasciclin I domain protein (Fdp) involved in *in vitro* adherence of *Rhodobacter sphaeroides*

Eun-Lee Jeong, Robert Moody, Samantha Broad, Stewart Goodwin, Jeff Keen, Alison Ashcroft and Mary Phillips-Jones

## Introduction

Members of the fasciclin I family of proteins occur in a wide range of vertebrates, invertebrates and microorganisms. They are generally cell-surface and membrane-anchored proteins involved in homophilic cell adhesion or symbiotic processes. One of the best studied examples is *Drosophila* FAS1, which is expressed during embryonic development, and guides axons from axon-generating neural cells to other target neurons or muscle cells Examples in mammals include transforming growth factor-β-induced gene product (βig-h3), periostin, osteoblast-specific factor 2, and the scavenger receptor FEEL-1 proteins. Amongst plants, fasciclin I-like domains occur in *Arabidopsis thaliana* SOS5 protein required for normal cell expansion, and cell surface arabinogalactan proteins required for plant growth and development. Microbial fasciclin I proteins include the MPB70 proteins of mycobacteria (for which a structure is available) and those important for symbiotic relationships by cyanobacteria and in cnidarian-algal associations. In symbiotic rhizobia such as *Sinorhizobium meliloti*, the fasciclin I protein Nex18 is required for normal nodule formation with leguminous plant partners.

Though implicated in general adhesion properties, the roles of fasciclin I domains have often been difficult to identify, partly because knock-out mutations do not exhibit discernible phenotypes. No specific role has previously been identified for any microbial fasciclin I protein. The function and regulation of fasciclin I proteins remain incompletely understood.

## Project

Here we report on the identification of a new member of the fasciclin I superfamily, Fdp (Fasciclin I Domain Protein), a simple single-domain protein found in the photosynthetic bacterium *Rhodobacter sphaeroides*, that is confirmed here as a member of this protein family by determination of its structure (Fig. 1).

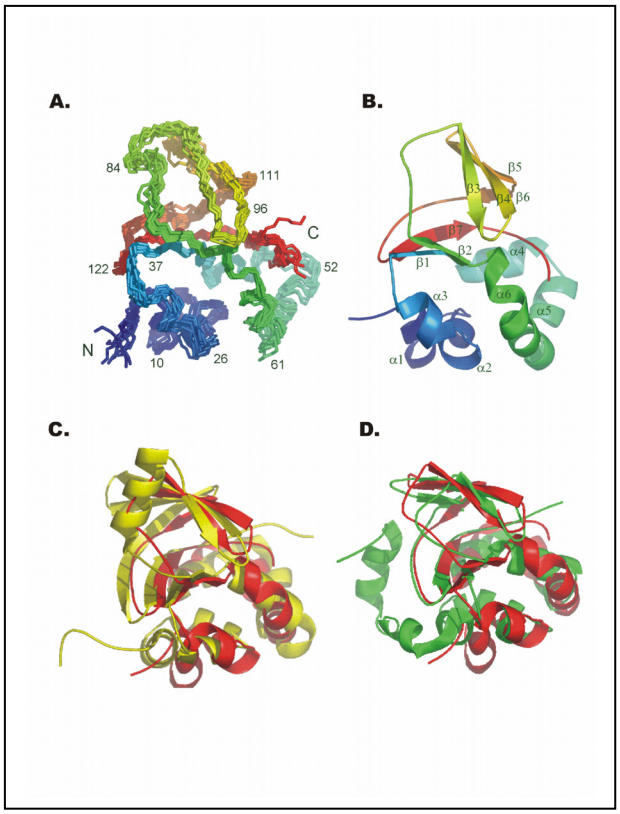



Fig. 1. **The solution structure of Fdp.** 10 overlaid structures, shown as a rainbow view, from blue at the N-terminus to red at the C-terminus. Only backbone atoms (Cα, C', N) are shown. (B) Best structure as a cartoon, same colour scheme and orientation. The α-helix and β-sheet numbering is indicated. (C) Best fit superposition to Drosophila melanogaster FAS1 domain 4 (Fdp red, FAS1 yellow). (D) Best fit superposition to Mycobacterium tuberculosis MBP70 (Fdp red, MBP70 green).

Our study defines a possible role for Fdp in adhesion properties of whole cells, which may be of significance for the bacterium in its natural environment. Inactivation of the *fdp* gene in *R. sphaeroides* resulted in significantly reduced adherence properties of cells, the first clear phenotype identified for any microbial fasciclin I protein (Fig. 2).

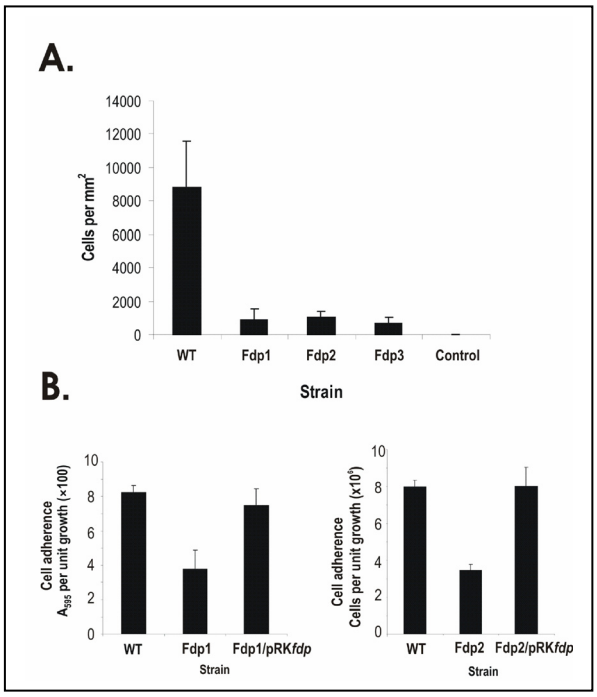



Fig. 2. **Adherence of R. sphaeroides NCIB 8253 strains.** A. Viable cells growing in biofilm attached to 68.1 $mm^2$ pegs following 5 days aerobic growth at 34 $^oC$. WT, wild type; Fdp1, Fdp2 and Fdp3, mutants possessing an insertionally-inactivated fdp gene; control, uninoculated medium (incorporating extremely low cell numbers that arise upon sonication). Standard deviation derived from 9 replicate samples. B. Biofilm formation during complementation studies of Fdp1 (left) and Fdp2 (right) using pRKfdp determined by crystal violet staining (left) or viable counting of biofilm cells (right), and expressed as a percentage of total culture growth in wells. Culturing was after 5 days at 34$^oC$. Fdp1/pRKfdp and Fdp2/pRKfdp are Fdp1 and Fdp2 harboring fdp inserted on pRK415 respectively. Standard deviation derived from 9 (left) and 3 (right) replicate samples.

We also demonstrated that Fdp is negatively regulated by the global redox-sensing two-component regulatory pathway Prr in *R. sphaeroides*, the first regulatory pathway to be associated with any fasciclin I protein. Reporter studies reveal that *fdp* transcription is negatively regulated by Prr under all conditions of aerobiosis tested, and probably by other systems in addition (data not shown). This is the first demonstration of regulation of any fasciclin I protein in microbial cells. We therefore present fasciclin I domains as domains important for bacterial physiology, and Fdp as an exemplar.

**Publications**

Saidijam, M., Bettaney, K.E., Szakonyi, G., Psakis, G., Shibayama, K., Suzuki, S., Clough, J.L., Blessie, V., Abu-bakr, A., Baumberg, S., Meuller, J., Hoyle, C.K., Palmer, S.L., Butaye, P., Walravens, K., Patching, S.G., O'Reilly, J., Rutherford, N.G., Bill, R.M., Roper, D.I., Phillips-Jones M.K. & Henderson P.J.F. (2005) Active membrane transport and receptor proteins from bacteria. *Trans. Biochem. Soc.* **33**, 867-872.

Jeong, E.L., Moody, R.G., Broad, S.J., Goodwin, S, Keen, J.N., Ashcroft, A.E., Williamson, M.P. & Phillips-Jones, M.K. (2006) Structure and redox-dependent regulation of a fasciclin I domain protein (Fdp) involved in *in vitro* adherence of *Rhodobacter sphaeroides. (submitted)*

**Collaborators**

Mike P. Williamson, Krebs Institute Structural Studies Group, University of Sheffield.

**Funding**

This work was funded by the BBSRC and a White Rose Studentship.

# Protein folding mechanisms

Victoria Morton, Claire Friel, Graham Spence, Stuart Knowling, Eva Sanchez Cobos, Daniel Lund, Inigo Rodriguez-Mendieta, Alastair Smith and Sheena Radford

## Introduction

One of the greatest challenges in modern structural biology is to understand how a newly formed polypeptide sequence finds its correct and unique fold. Real progress has been made towards establishing a fundamental and universal mechanism by which protein folding takes place, and such advances have come from the combination of experimental and theoretical techniques. In our laboratory we are studying the folding of three all α-helical proteins. The four-helix bacterial immunity proteins Im7 and Im9 have been the focus of our studies for 8 years. These proteins are structural homologues with high sequence identity, however they fold with mechanisms of different complexity. Im7 has been shown to fold *via* a compact, helical, on-pathway intermediate, while under the same conditions Im9 folds directly from the unfolded to the native state. We have shown the folding landscape of these closely homologous proteins to be finely balanced, such that small changes in sequence, or minor alterations in the folding conditions, can switch the kinetic mechanism of folding from two- to three-state. More recently we have extended our studies to include the ultra-fast folding three-helix bundle B domain of staphylococcal protein A (BdpA). We are using an array of biophysical methods to explore the folding landscapes of these three proteins, including laser temperature jump, ultra-rapid mixing, stopped flow and NMR.

## Ultraviolet resonance raman studies of Im7

Understanding the nature of partially folded proteins is challenging and best accomplished by combining several techniques. Ultraviolet resonance raman (UVRR) spectroscopy studies have been used to study the environments of the tryptophan and tyrosine residues in Im7, Im9 and variants of Im7 specifically designed to trap partially folded states at equilibrium. Results show that the environments of the tryptophan and tyrosine residues in the native state of wild-type Im7 and Im9 are indistinguishable, whereas these aromatic residues experience a more hydrophobic environment in the partially folded intermediate of Im7. These data suggest, therefore, that non-native interactions involving aromatic and aliphatic side chains play a role in the folding of Im7.

## Sulphate-induced effects in the on-pathway intermediate of Im7

Im7 folds through an on-pathway intermediate which has been shown to contain three of the four native helices. The structural properties of this intermediate were previously determined in the presence of 0.4M $Na_2SO_4$. To determine the effect of $Na_2SO_4$ on the properties of the Im7 intermediate, the folding of Im7 has been studied as a function of the concentration of $Na_2SO_4$ (0-0.4M) (Fig. 1). Detailed kinetic analysis, including Φ-value analysis at 0.2M $Na_2SO_4$, has shown that the structural properties of the intermediate are not significantly altered by the addition of $Na_2SO_4$. Therefore, whilst sulphate stabilises compact species on the Im7 folding landscape it does not alter their structural properties, confirming that the non-native interactions observed during Im7 folding are a generic feature of the folding of this polypeptide sequence.

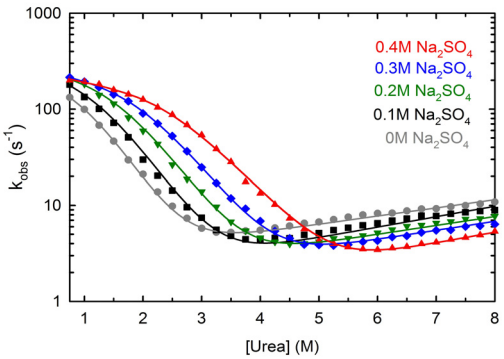



**Fig. 1**. Folding and unfolding kinetics of Im7 at different concentrations of $Na_2SO_4$. Continuous lines show the best fit to the three-state on-pathway model.

**Semisynthesis of a glycosylated analogue of Im7**

To establish a system to address questions concerning the influence of glycosylation on protein folding pathways, we have set up a new project to create *N*-linked glycosylated variants of Im7 using a semisynthetic approach. The strategy involves the native ligation of a glycosylated synthetic peptide with a bacterially expressed fragment (Fig. 2). An initial study of one such variant has shown that semi-synthetic Im7 analogues are well suited for protein folding studies, and that methods are available which will allow the complete characterisation of the role of individual sugar moieties on the folding free energy landscape of a protein. Further studies are currently underway using other glycoprotein analogues of Im7.

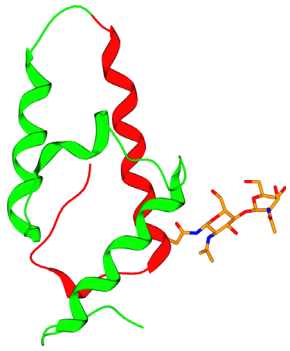


**Fig. 2.** Semisynthesis of Im7 involves the native ligation of a glycosylated synthetic peptide (red) with a bacterially expressed fragment (green), to create a glycoprotein with a disaccharide placed uniquely in the polypeptide chain.

**Helix stability and hydrophobicity in the folding mechanism of Im9**

The mechanism of protein folding has been proposed to be determined by the balance between the stability of secondary structural elements and the hydrophobicity of the sequence. To decipher the role of these factors in the folding of Im9, the secondary structural propensity or hydrophobicity of helices I, II and IV in Im9 was altered by substitution of residues at solvent exposed sites. The results of this study support a diffusion-collision model for immunity protein folding, in which intermediates are predicted to be stabilised both by increased helical propensity and by increasing stabilising contacts between helices by optimisation of either native or non-native contacts.

**Outlook**

In the next stages of this work, we are focusing on the fastest, earliest events in folding, using a combination of site-directed mutagenesis, Φ-value analysis, laser-induced temperature jump, ultra-rapid mixing techniques, and also a newly developed version of diffusion collision theory to analyse the folding of the four helix immunity proteins and the three-helix bundle B domain of staphylococcal protein A.

**Collaborators** Barbara Imperiali, Massachusetts Institute of Technology.

**Publications**

Cranz-Mileva, S., Friel, C.T., & Radford, S.E. (2005) Helix stability and hydrophobicity in the folding mechanism of the bacterial immunity protein Im9. *PEDS*, **18**, 41-50

Rodriguez-Mendieta, I.R., Spence, G.R., Gell, C., Radford, S.E. & Smith, D.A. (2005) Ultraviolet resonance Raman studies reveal the environment of tryptophan and tyrosine residues in the native and partially folded states of the E colicin-binding immunity protein Im7. *Biochemistry*, **44**, 3306-3315

Hackenberger, C.P.R., Friel, C.T., Radford, S.E. & Imperiali, B. (2005) Semisynthesis of a glycosylated Im7 analogue for protein folding studies. *JACS,* **127**, 12882-12889

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

**Funding and Acknowledgements**

We thank the BBSRC, FEBS and The Wellcome Trust for funding. We also thank Keith Ainley for technical support. SER is a BBSRC Professorial fellow.

# Single molecule protein folding using FRET

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

## Introduction

Due the stochastic nature of the folding process, some important aspects of protein folding and dynamics cannot be studied using ensemble methods. Examples of this complexity include the ensemble of unfolded protein molecules which contains a large number of rapidly interconverting protein conformations. In addition, there are rare folding intermediates, including energetically trapped or high energy species, the latter of which cannot be detected in an ensemble experiment. Single pair fluorescence resonance energy transfer (spFRET) has emerged as a powerful tool for characterising the properties and interactions of biomolecules and recently has begun to be used to monitor protein folding.

## Folding studies of Im7 and Im9

The folding mechanisms of Im7 and Im9 have been extensively studied using ensemble techniques. Interestingly, despite possessing 50% 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 at neutral pH. The Im9 variant Im9 S81C was specifically labelled with Alexa Fluor 594 (acceptor) on the naturally occurring cysteine (Cys23), and with Alexa Fluor 488 (donor) on the newly introduced cysteine 81. Another double cysteine Im9 variant (Im9 E2C C23A S81C) and the Im7 double cysteine variant (Im7 N5C K81C) were also labelled with the same dyes, but in this case a donor-acceptor/acceptor-donor mixture of double labelled proteins was produced.

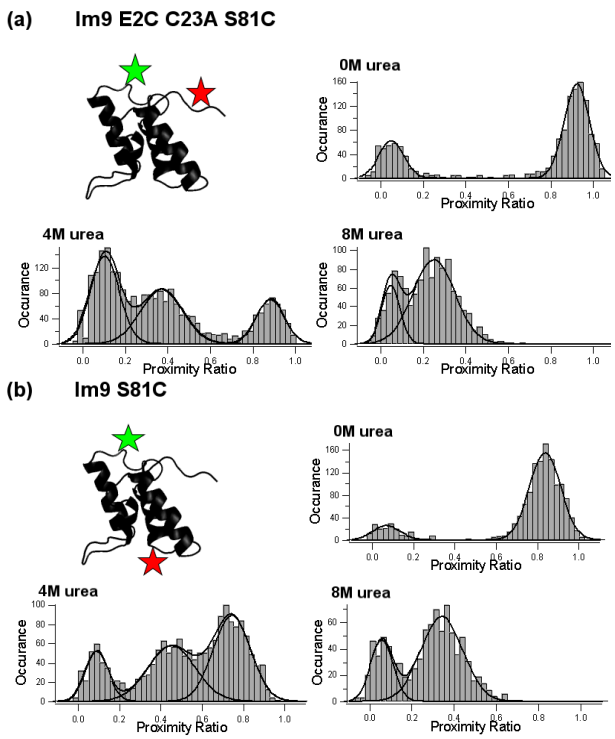



**Fig. 1.** spFRET histograms of (a) double labelled Im9 E2C C23A S81C at pH 7.0, 0.2M sodium sulphate and (b) Im9 S81C at pH 7.0 and 0.3M sodium sulphate. Example histograms calculated from the photon count trajectories of the two Im9 variants in 0M, 4M and 8M urea. The solid lines show Gaussian fits to the peaks. The peaks at high proximity ratio, ~0.9 (a) and ~0.8 (b), are the native protein, while the peaks with a lower proximity ratio ~0.2 (a) and ~0.3 (b), are the unfolded protein. The peak that occurs with a proximity ratio at around 0.1 is known as the "zero peak" and results primarily from photobleaching of the acceptor.

Ensemble folding studies were performed on all three labelled proteins. A low donor fluorescence and high acceptor fluorescence was detected in 0M urea buffer, reflecting the close proximity of the dyes in the native state. By contrast, high donor fluorescence and low acceptor fluorescence were detected in 8 M urea buffer, consistent with reduced energy transfer and therefore an increase in separation between the dyes as the proteins unfold.

The folding of Im7 and Im9 was next probed as a function of denaturant concentration (urea) and for Im9, chaotrope concentration using single molecule FRET, with particular attention

to the changes in compactness of both the native and denatured ensembles; measured by the relative FRET efficiency (the proximity ratio)(Fig. 1). In addition, the widths of the resolved species in the proximity ratio histogram can also contain information about the dynamics and heterogeneity of the proteins within each state; information which is not readily available in ensemble studies. The results revealed that Im9 folds with a two state transition, as measured by single molecule FRET. 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 in a complex manner on the concentration of denaturant. These studies are being continued, by examining other proteins and other variants, to determine whether this property is unique to Im9 or a general feature of the denatured state of many proteins.

**Other proteins for future study**

The Im7 variant L53AI54A (Im7 $I^{eqm}$), is trapped in the intermediate state and at equilibrium this is the most highly populated species. Labeling double cysteine variants of this protein will allow a kinetic intermediate to be studied at the single molecule level. The B domain of protein A (Bdp A), a three-helix bundle protein, is another interesting candidate for single molecule studies (Fig. 2) as it folds and unfolds very rapidly, with a folding rate constant of around 249000$s^{-1}$ at 45°C. Double cysteine variants of BdpA have been created and labelled with Alexa Fluor 488 and 594. Folding studies using single molecule diffusion FRET and fluorescence correlation spectroscopy are currently underway.

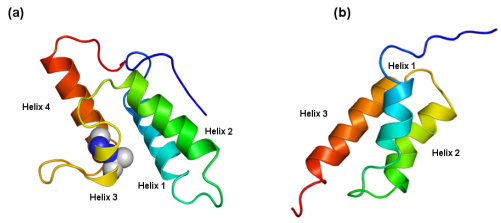



**Figure 2.** Schematic representations of (a) Im7 L53AI54A and (b) B-domain of protein A. The images were prepared using PyMOL (DeLano Scientific).

**Single molecule kinetics**

Diffusion studies do not allow the folding trajectory of a single domain to be studied as a function of time. These types of experiments require a spatially immobilised protein. We are currently developing the methodology to encapsulate labelled proteins in vesicles as a prelude to detailed analysis of folding and unfolding of proteins in real time using single molecule methods.

**Publications**

Gell, C., Tezuka-Kawakami, T., Brockwell, D.J., Radford, S.E. & Smith, D.A.M. (2006) "Denaturation of the Immunity Protein Im9 monitored by spFRET", *Biophysical Journal - Biophysical Letters*, submitted.

**Funding and Acknowledgements**

We would like to thank Graham Spence for helpful discussion and Alison Ashcroft and Victoria Homer for Mass Spectrometry. We acknowledge the Wellcome Trust, BBSRC and the University of Leeds for funding.

# 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 produces 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 (Fig. 1a). Fibre assembly is known to occur by a β-strand exchange mechanism, in which an N-terminal extension 'Nte' strand from a subunit is donated to an adjacent subunit molecule, causing subunit polymerisation (Fig. 1b). 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 in systems such as Saf are largely unknown, but it is 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.

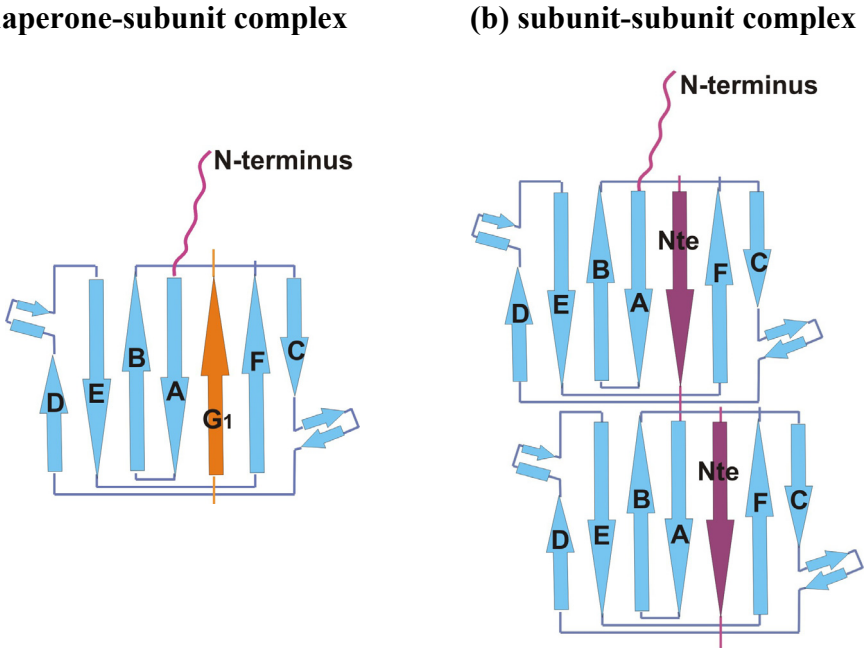



**Fig. 1**. Topology diagrams showing the SafA subunit (strands A-F, blue; N-terminal extension, purple) in complex with (a) the $G_1$ -strand of the chaperone SafB (orange) and (b) the N-terminal extension of adjacent subunits (purple).

## 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, observing the decrease in concentration of the SafA-SafB non-covalent complex (Mw = 37,123Da), and the subsequent appearance and increase in concentration of the SafA-Nte peptide non-covalent complex (Mw = 15,193 Da). The presence of a SafA-SafB-Nte peptide ternary complex (Mw = 39,191 Da) has been detected (Fig. 2), confirming directly and for the first time that the reaction proceeds by a

concerted mechanism involving a stable ternary intermediate. Peptides of the Nte strand with single amino acid substitutions 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 the sequence of the peptide, and this can be attributed to hydrophobic residues on the N-terminal extension sequentially replacing residues of the chaperone, causing dissociation of the chaperone-subunit complex in a zip-in zip-out mechanism.

Thus, by using the power of mass spectrometry to resolve mixtures of protein complexes that are co-populated in solution, mechanistic details of the donor strand exchange mechanism, possibly of generic importance for pilus assembly in many bacteria, have been revealed.

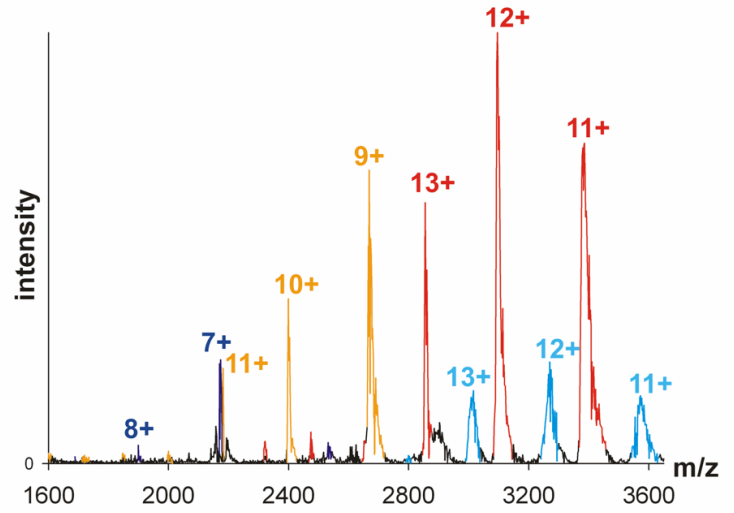



**Fig. 2**. Mass spectrum of a sample taken during a donor strand exchange reaction between SafA-SafB (red peaks) and wild-type Nte peptide (not shown), 4 minutes after reaction initiation. The spectrum provides the first evidence of the existence of a ternary intermediate (light blue peaks) during donor strand exchange. The products of the reaction, free SafB and SafA-Nte, are shown by the yellow and dark blue peaks, respectively.

## Collaborators

This work has been in collaboration with Prof. Gabriel Waksman and Dr. Han Remaut from the School of Crystallography, UCL/ Birkbeck College, London.

## 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 pilus assembly proceeds through a zip-in-zip-out β-strand displacement mechanism. Submitted.

## Funding

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

# Folding and aggregation of $\beta_2$-microglobulin

Thomas Jahn, Geoffrey Platt, Katy Routledge, Martin Parker, Arnout Kalverda,
Steve Homans and Sheena Radford

## Introduction

There are approximately twenty proteins with unrelated amino acid sequences and native structures that aggregate to form highly ordered amyloid fibrils *in vivo* and result in amyloid diseases. More recently, several of these proteins, and some that are not disease-related, have been shown to form amyloid *in vitro* by manipulation of solution conditions. In all cases the normally soluble proteins deposit as insoluble amyloid-like fibres with characteristic cross-β structure. Our research aims to elucidate the mechanism of fibril formation for human $\beta_2$-microglobulin ($\beta_2$m) which causes haemodialysis related amyloidosis in all patients with renal failure.

## An amyloidogenic folding intermediate found at physiological conditions

Studies of the transition between soluble precursors and insoluble amyloid fibrils have been carried out *in vitro* for many proteins, usually in conditions that destabilise the native state (i.e. by the addition of denaturant, low pH, high temperature or amino acid substitutions). Under these conditions the population of partially unfolded conformations (including amyloidogenic precursor states) is increased. However, the most fundamental questions about the identity of such species under *in vivo* conditions as well as their structural characteristics remain unresolved.

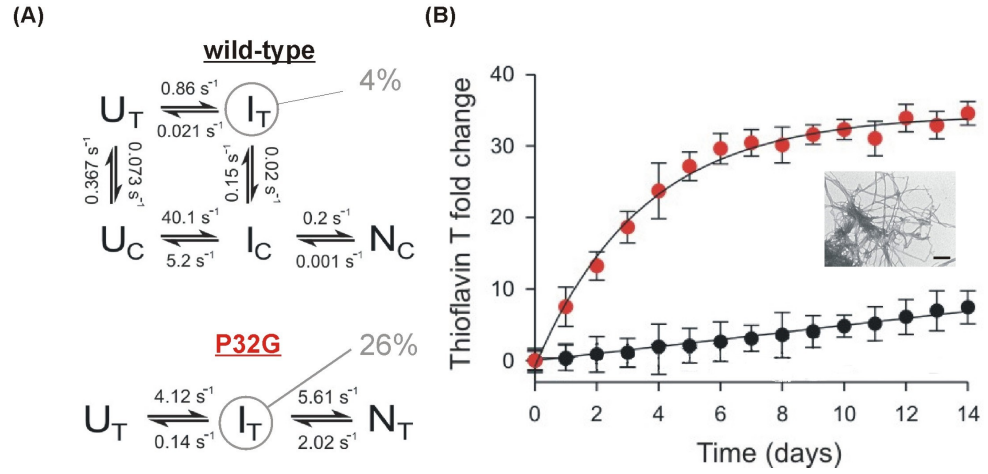



**Fig. 1.** (A) The folding mechanisms of wild-type and P32G $\beta_2$m. (B) Fibril elongation of wild-type (black) and P32G (red) $\beta_2$m monitored using Thioflavin-T fluorescence at pH 7.0, 37 °C. Inset: electron micrograph of P32G fibrils formed after 7 days under these conditions.

To study the link between folding and aggregation of $\beta_2$m, the folding mechanism was determined using a series of single and double-jump stopped flow experiments monitored by intrinsic tryptophan fluorescence (pH 7.0, 37 °C). Creating the protein variant P32G showed that the slow folding phase is limited by *trans-cis* proline-isomerisation. The kinetic data for P32G $\beta_2$m could be well described by a 3-state model, whereas the simplest model able to satisfy all the kinetic data for wild-type $\beta_2$m is a 5-state scheme (Fig. 1A). The results demonstrate that wild-type $\beta_2$m folds to the native state *via* two parallel routes, one of which involves folding in a reaction limited by *trans-cis* isomerisation of Pro32. Analysis of the rate constants for this reaction scheme revealed that the intermediate, $I_T$, is populated to a significant extent (4%) under these physiologically relevant conditions. As this intermediate is populated to 26% in the P32G variant, a comparison of these two proteins allows the

amyloidogenic properties of $I_T$ to be defined. Fibril elongation of wild-type and P32G $\beta_2$m was monitored using Thioflavin-T fluorescence and electron microscopy (Fig. 1B). The results showed a dramatic enhancement in the rate of fibril elongation for P32G, indicating a clear and direct correlation of the rate of fibril formation with the concentration of $I_T$.

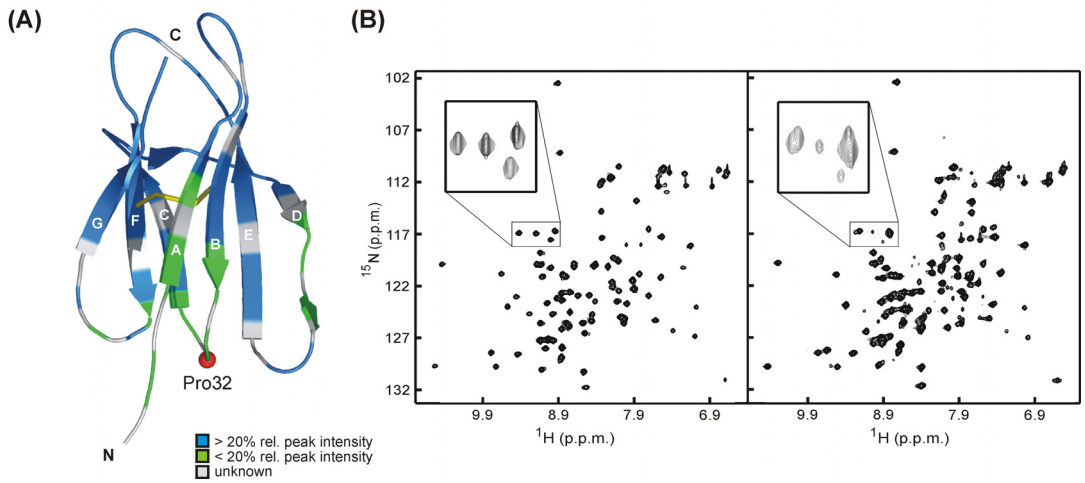



**Fig. 2.** (A) Ribbon diagram of wild-type $\beta_2$m. Regions affected by line-broadening are shown in green. (B) $^{1}$H-$^{15}$N NMR spectra of wild-type (left) and P32G $\beta_2$m (right) at pH 7.0, 37 °C. The spectrum of P32G shows line-broadening for several resonances indicative of conformational exchange between $I_T$ and N.

$^{1}$H-$^{15}$N HSQC NMR spectra show that P32G $\beta_2$m adopts a similar native structure to wild-type $\beta_2$m. However, substantial line-broadening is observed for residues that lie in the A and D-strands, the loop between β-strands B and C (which contains *trans*-Gly32) and the adjacent FG loop, suggesting that conformational changes in these regions are responsible for the enhanced amyloidogenic properties of the variant protein (Fig. 2).

**Fibril formation by variant $\beta_2$m proteins**

Correlating structure and dynamics information with the amyloidogenicity of $\beta_2$m is an important step in understanding the mechanism of fibril formation. We are currently analysing the role of different regions of $\beta_2$m by mutagenesis methods to explore the importance of sequence in promoting amyloidosis and conferring stability to the fibrillar state of $\beta_2$m. We are using a variety of spectroscopic methods such as Thioflavin-T fluorescence, circular dichroism and electron microscopy in conjunction with NMR techniques to study the rate of fibril formation and morphological features of fibrils formed from $\beta_2$m mutants under a range of pH values.

**Publications**

Platt, G.W., McParland, V.J., Kalverda, A.P., Homans, S.W. & Radford, S.E. (2005) Dynamics in unfolded states of $\beta_2$ microglobulin provide insights into the mechanism of fibril formation. *J. Mol. Biol.* **346,** 279-294

Jahn, T.J., 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*. (in press).

**Funding**

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

# Mechanical unfolding of proteins using dynamic force spectroscopy

Kirstine Anderson, Eleanore Hann, Gerard Huysmans, David Sadler, David Brockwell, Alastair Smith and Sheena Radford

**Determining the factors that confer mechanical resistance on globular proteins**

There are many cellular functions that require proteins to exhibit a resistance to applied force. For example, scaffold proteins such as the giant muscle protein titin and extracellular matrix proteins such as fibronectin are constantly subject to changing forces. Proteins are also thought to be subject to force during membrane translocation and protein degradation, hence their mechanical resistance may affect the efficiency of these processes. Mechanical unfolding of proteins can be readily studied using techniques such as force mode AFM and optical tweezer instruments. The role that the type of secondary structure plays in force resistance is understood (helical proteins, in general, are force labile relative to beta-sheet containing proteins), but current knowledge is not sufficient to accurately predict the mechanical properties of a protein.

Protein L is a 62 amino-acid protein with no mechanical function. It has a simple β-grasp topology and demonstrates a surprisingly high resistance to force. These observations, together with experiments on other small proteins have led to the dogma that parallel and directly hydrogen bonded terminal β-strands confer mechanical resistance to a protein. In accord with this, other small proteins with this topology (eg. protein G and ubiquitin) also show mechanical strength, irrespective of their biological function. Interestingly, however, despite their shared topology, ubiquitin is more highly resistant to force than protein L, suggesting that differences in the details of their sidechain packing also contributes to mechanical stability. To test this hypothesis, mechanical phi-value analysis is being carried out on a number of protein L variants that have been designed to alter different sidechain contacts in the native protein. Mechanical unfolding of one of these mutants, L10A, reveals that it shows a similar force resistance to wild-type, despite being significantly less thermodynamically stable.

Another factor which has been shown to affect the mechanical resistance of a protein is the geometry of the applied force. Most experimental data are collected by extending proteins between their N- and C-termini, although a limited number of studies have mechanically unfolded proteins by applying force in alternative directions. I27 is a model protein that has been characterised extensively by application of force from its N- and C-termini. However, a method is being developed by which I27, or indeed any model protein, can be extended between its N-terminus and any other position in the domain. This method can be used to characterise the whole mechanical unfolding landscape of a protein and importantly, will unveil further clues about the factors that determine mechanical resistance and serve as a powerful benchmark for simulations of the mechanical unfolding of proteins using different approaches (see report by West *et al*).

As well as the intrinsic properties of a protein, external factors can affect the extent to which a protein can resist force. These include interactions with a ligand (see below) and the external environment. Membrane proteins are often inaccessible to protein folding studies because of their large size and insolubility. One way in which membrane proteins can be studied is AFM, which can be used to directly unfold proteins out of membranes. To date these studies have been carried out on α-helical inner membrane proteins. In a new project we have initiated studies of the mechanical properties of outer membrane proteins with all β-sheet structures.

**Mechanical consequences of the interaction between Im9 and E9**

The interaction between the colicin E9 and its associated immunity protein Im9 is amongst the strongest protein-protein interactions characterised to date, with a $K_d$ of the order of $10^{-16}$ M. An interesting and readily used application of force-mode AFM is the investigation of mechanical strength of interactions between proteins and their natural ligands. For example, both antibody-antigen and avidin-biotin complexes have been dissociated by force spectroscopy. Since the Im9-E9 interaction is stronger than either of these interactions, we are using this system to determine both the mechanical strength of the interaction and the effect that tight and specific binding has on the mechanical strength of one of the protein in the complex, Im9 (see Fig. 1).

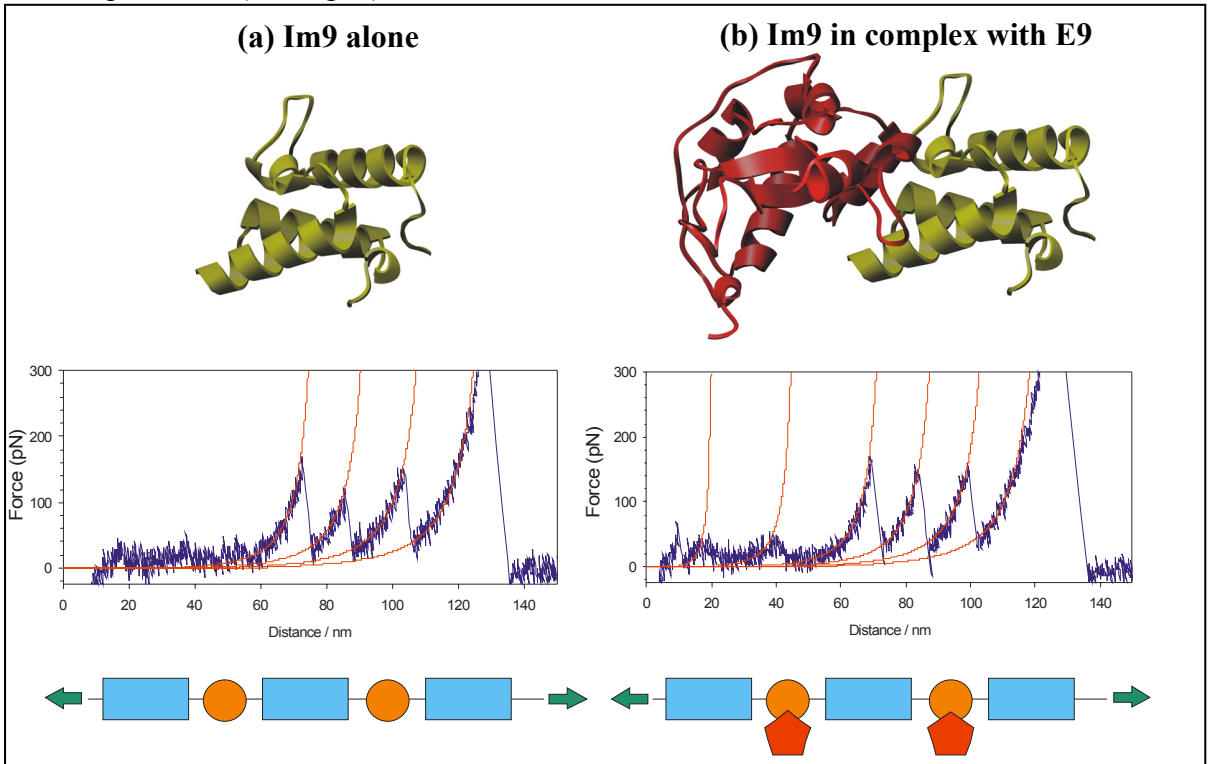



**Fig. 1.** The difference in mechanical stability of (a) Im9 alone and (b) Im9 in complex with E9. Top: Ribbon representations of Im9 (yellow) with and without E9 (red). Middle: Force-extension curves of Im9-protein L concatamers. Peaks at around 150 pN correspond to protein L unfolding whereas peaks at around 40 pN (right hand plot only) correspond to Im9 unfolding. A worm-like chain model has been fitted to each unfolding event and is shown in red. The final peak in each trace corresponds to the protein detaching from the AFM tip. Bottom: Representation of the concatamer in which Im9 is being pulled. Rectangles represent protein L, circles Im9 and pentagons E9. Green arrows: direction of the applied force.

**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**

Brockwell, D.J., Beddard, G.S., Paci, E., West, D.K., Olmsted, P.D., Smith, D.A. & Radford, S.E. (2005) Mechanically unfolding the small, topologically simple protein L. *Biophys. J.* **89**, 506-519.

**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 and SER is a BBSRC Professorial Fellow.

# Investigation into the initiation of fibril formation in dialysis related amyloidosis

Isobel Morten, Antoni Borysik, Sheena Radford and Eric Hewitt

## Introduction

$\beta_2$-microglobulin ($\beta_2$m) is one of approximately twenty proteins that aggregate to form highly ordered amyloid fibrils *in vivo*. $\beta_2$m is a small 99 residue soluble protein, which is non-covalently bound to a membrane-integrated heavy chain, forming the major histocompatibility complex (MHC) class I molecule which is expressed on the surface of all nucleated cells. *In vivo*, $\beta_2$m is continuously shed from nucleated cells into the serum. $\beta_2$m is then transported to the proximal tubule of the kidney where it is degraded and excreted. As a consequence of renal failure the $\beta_2$m serum concentration increases by up to 60-fold. By a mechanism that is currently unresolved free $\beta_2$m then self-associates, forming insoluble amyloid fibrils which typically accumulate in synovial joints. As a consequence, uremic patients who have been dialysed for 10-15 years develop dialysis-related amyloidosis (DRA), a debilitating arthritic-like condition.

## Hypothesis

How $\beta_2$m forms amyloid fibrils *in vivo* is unknown. In addition, like many amyloid disorders fibril deposition occurs site-specifically, in the case of DRA the fibrils develop specifically in the synovial joints. Therefore, we hypothesised that one or more components found within the synovium may promote amyloid formation in this specialised environment. Therefore, using a combination of biochemical, cell biological and biophysical experiments we are testing if either the cells or macromolecules associated with the synovial joints of DRA patients promote fibrillogenesis.

## Do macrophages promote $\beta_2$m amyloid fibrillogenesis?

*In vitro* studies in our laboratory have shown that incubation of $\beta_2$m at acidic pH or the removal of the N-terminal six residues ($\Delta$N6$\beta_2$m) induces rapid fibril formation. We propose that cells within the joint capsule may internalise $\beta_2$m, which enters the endocytic pathway and accumulates in the lysosomes. The acidic microenvironment of lysosomes (pH 4.5), coupled with its high concentration of proteases, may then stimulate fibril formation.

Macrophages are the only known cell type to be associated with $\beta_2$m amyloid fibrils *in vivo*. To determine if $\beta_2$m is endocytosed into cells we N-terminally labelled the protein with fluorescein isothiocyanate (FITC) and incubated the protein at 50μg/mL (the concentration of $\beta_2$m typically found in the serum of uremic patients) with RAW 264.7 cells, a mouse macrophage cell line. Internalisation of $\beta_2$m was analysed using live cell confocal microscopy, and localisation of internalised $\beta_2$m was determined using LysoTracker, a lysosomal specific dye. Internalised $\beta_2$mFITC accumulates in perinuclear regions and co-localises with LysoTracker. These data are consistent with sorting of the internalised $\beta_2$m to lysosomes.

Lysosomal conditions were mimicked using an ammonium acetate buffer at pH 4.5. Rod-like fibrils were formed under these conditions by incubation for 24 hours at 37$^{o}$C whilst longer, worm-like fibrils are generated under the same conditions after incubation for 1 week. Both fibrillar forms have been shown to be amyloid-like by their ability to bind amyloid specific antibodies, the dyes thioflavin T, and Congo red as well as their appearance visualised using atomic force microscopy (AFM) (Figure 1).

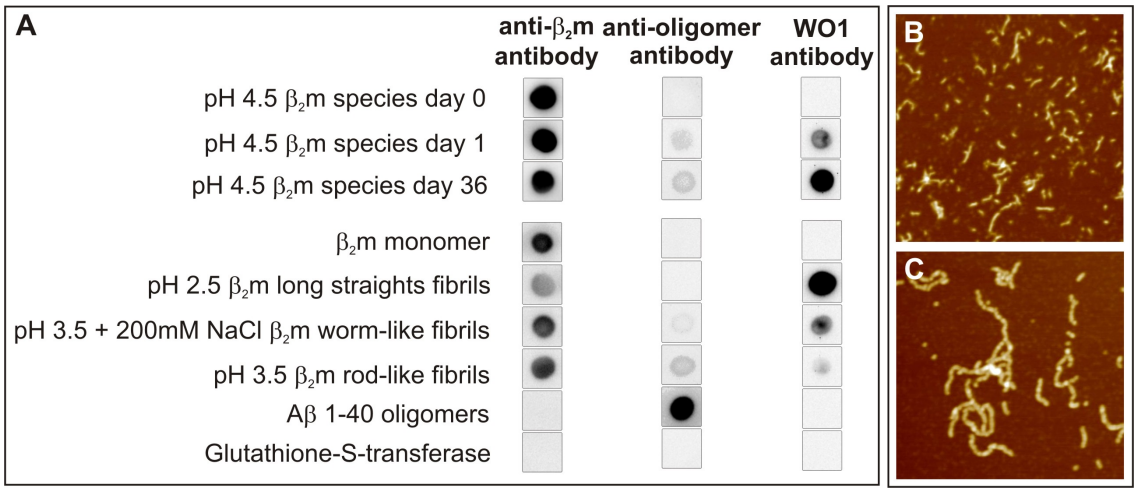



**Fig. 1** – Characterisation of $\beta_2$m species formed at pH 4.5. *A*, Dot blot of $\beta_2$m species incubated at pH 4.5 at time points 0, 1 and 36 days, using anti-$\beta_2$m antibody (DAKO), an oligomer specific antibody (anti-oligomer) and a fibrillar specific antibody (WO1). Aβ1-40 oligomers acted as a positive control for the anti-oligomer antibody and glutathione-S-transferase (GST) acted as a negative control for all antibodies. Representative AFM images (1 $\mu m^2$) of rod-like (*B*) and worm-like (*C*) $\beta_2$m fibrillar species formed at pH 4.5.

To determine the effect that the lysosomal enzymes have on $\beta_2$m, lysosomes were isolated from RAW 264.7 cells by subcellular fractionation and the enzymes extracted. The enzymes were incubated with monomeric $\beta_2$m or preformed rod-like $\beta_2$m fibrils, and the digestion products identified using N-terminal sequencing and mass spectrometry. Digestion of the monomer produces two products, an N-terminal truncation and a single cut within the region corresponding to the native E strand with the subsequent N- and C-terminal regions of the protein remaining attached *via* the native disulphide bond. By contrast, digestion of the fibrillar species yields only one digestion product, the same N-terminal truncation as seen when the monomer was digested. These results demonstrate that the core of the protein is inaccessible to the lysosomal enzymes when it is packed into the fibrillar structure.

**Do macromolecules specific to the synovium promote fibrillogenesis from $\beta_2$m?**

In parallel we are testing whether macromolecules associated with the synovium can nucleate fibrillogenesis of $\beta_2$m, since the monomeric protein does not nucleate fibril formation in isolation at neutral pH. We are monitoring the effect of a range of different glycosaminoglycans on fibril formation from the destabilised truncated form of $\beta_2$m (ΔN6). To date we have shown that some, but not all, forms of these long-chain high molecular weight sugars can significantly enhance the nucleation of ΔN6 fibrils *in vitro* at physiological pH. We are now beginning to unpick the mechanisms involved in how these exciting candidate glycosaminoglycans can possibly contribute to $\beta_2$m amyloidosis under physiological conditions.

**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 $\beta_2$-microglobulin amyloid formation at neutral pH. *Biochemistry*, *in press*

**Funding**

We gratefully acknowledge Kidney Research UK, The Wellcome Trust and the BBSRC for financial support. SER is a BBSRC Professorial Fellow. We thank Keith Ainley for technical support and other members of our research groups, particularly Wooley Gosal for the AFM images in Fig. 1 and Thomas Jahn for advice in the use of GAGS in fibril assays.

# Exploring the assembly pathways of $\beta_2$-microglobulin amyloidosis.

Walraj Gosal, Thomas Jahn, Sarah Myers, Susan Jones, Alastair Smith, Neil Thomson and Sheena Radford.

## The heterogeneous nature of $\beta_2$m assembly pathways

Amyloidoses are a subset of protein misfolding diseases characterised by the pathological deposition of amyloid fibrils. In addition, amyloid fibrils are also the underlying physical feature in other areas of biology. Mechanistic details of how proteins assemble into amyloid fibrils are unclear, and furthermore, it is beginning to emerge that the assembly mechanism may be heterogeneous for a number of proteins. One example is the assembly of $\beta_2$-microglobulin ($\beta_2$m), a 99 residue all-β-sheet protein, which is associated with the condition 'haemodialysis-related amyloidosis', which affects all patients with renal failure on long-term dialysis.

Using atomic-force microscopy (AFM), we have shown that *in vitro* at low pH, $\beta_2$m forms distinct classes of amyloid fibrils, as defined by differences in their morphology and persistence-length. The appearance of these fibrils types (worm-like (WL); rod-like (RL) and long, straight (LS)) is highly dependent on the solution conditions (Fig. 1A), and we have mapped their occurrence into 'state-diagrams'. Using these data, we have performed a number of experiments to show that fibrils of different morphology form on distinct and competitive pathways of assembly, defining an energy landscape that rationalises the sensitivity of fibril morphology on the solution conditions (Fig. 1B). We are currently investigating how, why and what governs the bifurcation of these pathways. The heterogeneous nature of assembly mechanisms has a number of implications in amyloid-related biology, one of which is determining their differing pathological role. For example, assembly pathways that give rise to small oligomeric assemblies (*i.e.* either pore-like, small globular oligomers, and/or worm-like (flexible) fibrils), have been shown to be harmful to cells in cell-viability assays, and therefore, determining their role in the mechanism of amyloid formation is imperative.

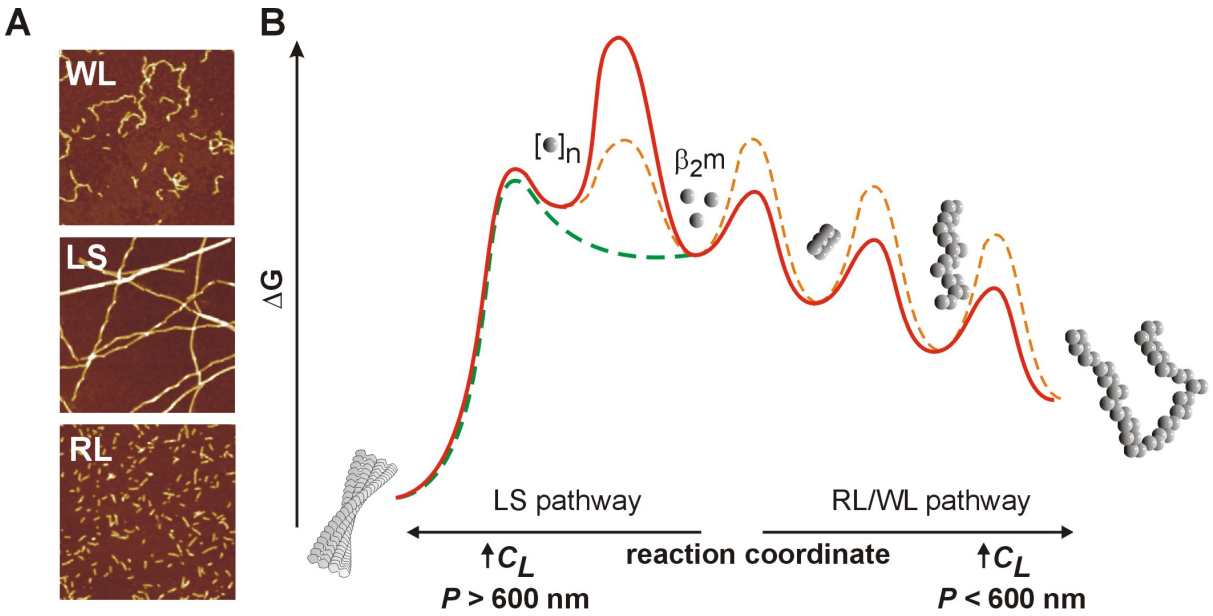



**Fig. 1. A.** AFM images (1μm$^2$) of the various types of fibrillar morphologies formed when $\beta_2$m assembles into amyloid-like fibrils under different conditions *in vitro*. **B.** A schematic energy-landscape diagram of the bifurcated assembly mechanism of $\beta_2$m, where LS fibrils are formed through a nucleated process, which competes with a non-nucleated pathway, leading to the formation of RL and WL fibrils.

**Physiological factors at neutral pH**

Although $\beta_2$m forms fibrils readily at low pH *in vitro* when partially unfolded, how $\beta_2$m forms fibrils *in vivo* is not at all well understood. Indeed, under most conditions, fibrils formed *in vitro* at low pH are not stable at neutral pH, which is not the case in the disease state. One possibility may be that other physiological factors may stabilise the fibril deposits in patients, and we have systematically investigated this hypothesis, using factors such as heparin, serum amyloid P protein (SAP), apolipoprotein E (apoE), collagen and uremic fluid. We were able to show that some of these factors, to varying extents, are to stabilise fibrils formed at low pH (Fig. 2A), and that these stabilised seeds could be used to seed further growth at neutral pH (Fig. 2B). A possible therapeutic avenue may thus be to inhibit specific stabilising factors in haemodialysis-related amyloidosis.

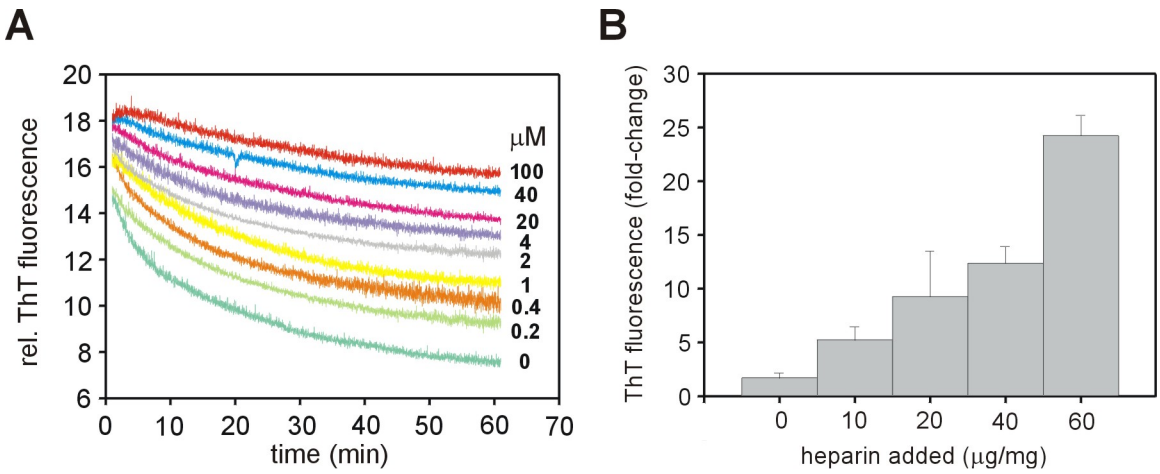



**Fig. 2. A.** Depolymerisation at neutral pH of $\beta_2$m fibrils formed at pH 2.5 is significantly reduced by the addition of heparin, as monitored by the fluorescent dye Thioflavin-T. **B.** Heparin stabilised seeds (50$\mu$g) can be used to enhance fibril-formation at neutral pH in a solution of 500 $\mu$g of $\beta_2$m.

**Publications**

Gosal, W.S., Morten, I.J., Hewitt, E.W., Smith, D.A., Thomson, N.H. & Radford, S.E. (2005) Competing pathways determine fibril morphology in the self-assembly of $\beta_2$-microglobulin into amyloid. *J. Mol. Biol.* **351**, 850-864.

Jahn, T.R., & Radford, S.E. (2005) The Yin and Yang of protein folding. *FEBS J.* **272**, 5962-5970.

Radford, S.E., Gosal, W.S. & Platt, G.W. (2005) Towards an understanding of the structural molecular mechanism of $\beta_2$-microglobulin amyloid formation *in vitro*. *Biochim. Biophys. Acta. Proteins & Proteomics.* **1753**, 51-63

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.

**Funding**

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

# RNA aptamers that inhibit the replicase enzyme of foot and mouth disease virus

Mark Ellingham, David Rowlands and Nicola J. Stonehouse

## Introduction

Foot-and-mouth disease virus (FMDV) is a highly important animal pathogen. As a member of the picornavirus family, it consists of a single strand of RNA surrounded by a protein capsid. 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. There is an urgent need to develop antiviral drugs and efficient diagnostic tests to facilitate more effective management of future outbreaks of the disease.

## Results

SELEX is a powerful technique that works on the principle that nucleic acid sequences isolated on the basis of their binding affinity to a specific target can be preferentially amplified over competing sequences. Repetition of this process results in a population of sequences (aptamers) that specifically bind to the target with a very high affinity. In addition to therapeutic applications, aptamers can also be used as diagnostic tools and as probes to define a binding site for the design of small-molecule inhibitory molecules.

15 rounds of selection have been performed against the 3Dpol protein (an RNA-dependent-RNA-polymerase) of FMDV-C using a random sequence library of RNA as a starting pool and with the intention of generating high affinity RNA ligands against the protein. A number of cDNA clones derived from the selected pool were sequenced and found to have significant areas of homology with the 5'-untranslated region of the viral genome.

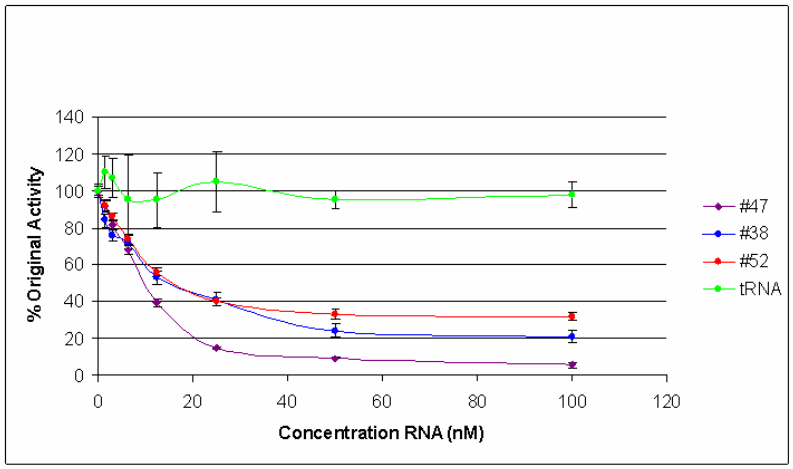



**Fig. 1**
Reduction of activity of FMDV 3D in the presence of RNA aptamers 38, 47 and 52

Certain aptamer species (80mers) were chosen from the pool to assess their ability to inhibit the *in vitro* catalytic activity of the 3D protein. 3 aptamers were found to inhibit the ability of the enzyme to transcribe RNA with an $IC_{50}$ of 15-20 nM (Fig. 1). One of these aptamers has been truncated to a 30mer, with no loss of affinity or inhibitory activity and studies are underway to define a minimal binding region for use in structural studies of complexes of the RNA with the enzyme (in collaboration with Nuria Verdaguer (Barcelona). In addition, in collaboration with Esteban Domingo (Madrid), preliminary experiments suggest that two of the aptamers inhibit viral replication when delivered to cells infected with FMDV.

The sequences of the aptamers have been protected with a patent since their inhibitory properties could potentially lead to applications such as diagnostics or molecular probes in the future.

**Collaborators**
Esteban Domingo (Madrid)
Nuria Verdaguer (Barcelona)

**Funding**
With thanks to the BBSRC for funding.

# Investigating the affinity of molecular interactions of the ϕ29 packaging motor

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

## Introduction

ϕ29 is a bacteriophage responsible for the infection of *Bacillus* species and its proposed mode of DNA packaging, via a molecular motor, has stimulated much recent interest. Commonly molecular motors display either a rotary motion e.g. bacterial flagella, or a linear movement e.g. muscle contraction.

The ϕ29 molecular motor is thought to consist of a multimeric RNA-protein complex which packages a 19kb dsDNA genome into a preformed prolate icosahedral prohead. This is thought to occur via the central channel in the gp10 connector complex (Fig. 1). ϕ29 DNA packaging occurs against a large internal force and as such is an energetically unfavourable process. The current model to explain ϕ29 DNA packaging involves ATPase activity, in combination with rotation of the components, allowing concomitant passage of genomic DNA through the connector complex.

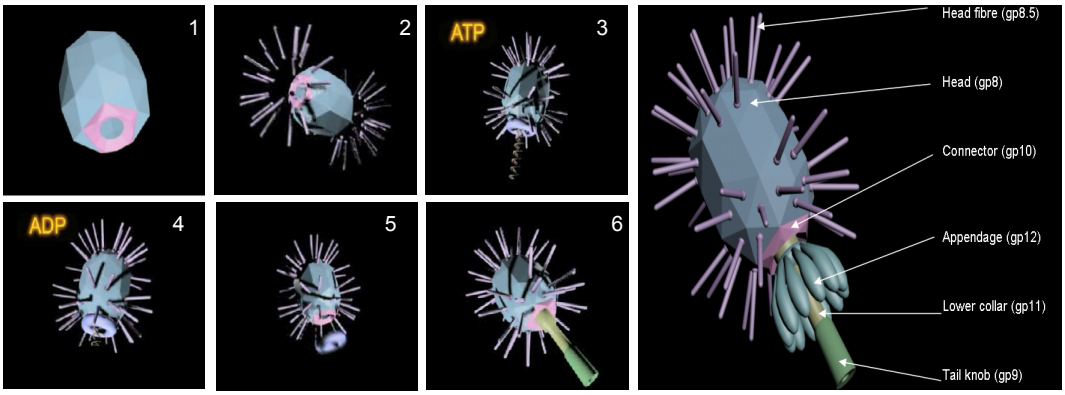



**Figure 1 Cartoon representation of ϕ29 DNA packaging**
1 and 2) The preformed prohead / connector assembly; 3 and 4) Association of pRNA (in purple) and ATP dependant dsDNA packaging; 5) dissociation of the pRNA and 6) final mature virion.
Images from Katie Radcliffe

## Results

The details of the protein:RNA interactions involved in DNA packaging are largely unknown and the mechanistic details of packaging are yet to be elucidated. The aim of this work is to develop a detailed understanding of this system and the interactions therein. Both RNA (pRNA) and connector protein form multimeric complexes (Fig. 1). The multimerisation of the pRNA in solution has been investigated in this laboratory by analytical ultracentrifugation (AUC) and light scanning experiments. Results from these experiments indicated that pRNA forms monomeric, dimeric and trimeric species driven by interactions with divalent metal ions such as magnesium. Current work has involved the chemical and enzymatic synthesis of novel pRNA species in order to study the importance of this RNA:RNA interaction. We have shown that the affinity of the pRNA:pRNA interaction is μM, whereas surface plasmon resonance and binding affinity studies of the RNA:protein interaction resulted in a $K_D$ in the nM range. Interestingly, connector binding affinity is increased in the presence of magnesium; however, this affinity does not seem to be solely dependent on pRNA multimerisation, even though pRNA multimerization is essential for DNA packaging.

Intrinsic tryptophan fluorescence quenching and circular dichroism spectroscopy studies suggested a magnesium-induced conformational change in the connector. This, combined with the differences in the magnitude of affinity between components of the motor, lead us to the hypothesis that it is unlikely that there is a rotation between the connector and pRNA during DNA packaging, and we propose that magnesium-induced conformational changes in both connector and pRNA could drive the packaging event.

**Collaborators**
Peixuan Guo, Purdue University
Alistair Smith, University of Leeds

**Funding**
Funding from the BBSRC is gratefully acknowledged.

**Publications**
Wood, J.P.A., Capaldi, S.A., Robinson, M.A., Baron, A.J. & Stonehouse, N.J. (2005) RNA multimerisation in the DNA packaging motor of bacteriophage phi29. *Journal of Theoretical Medicine* **6**, 127-134.

# Determination of the structure of a type IV topoisomerase from *Staphylococcus aureus*

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

## Background

DNA topoisomerases are ubiquitous enzymes responsible for resolving topological problems arising during DNA transcription, recombination, replication and chromosome partitioning. Topoisomerase IV (topoIV) of *Staphylococcus aureus* is a type II topoisomerase and is composed of two homodimeric subunits: GrlA, which is responsible for DNA binding, strand cleavage and religation, and GrlB, which hydrolyses ATP enabling enzyme turnover. The action of topo IV *in vitro* include a reduction in the superhelical density within the target DNA; it has also been shown to be one of the primary targets of fluoroquinolone antibiotics in *S. aureus*

A 56 kDa proteolytic fragment of GrlA (GrlA56) has been produced and this displays atypical DNA cleavage activity *in vitro,* as it is active in the absence of the ATP hydrolysing subunit. GrlB. GrlA56 has been crystallised and diffraction data collected at the ESRF (Grenoble, France) to a resolution of 2.8 Å.

## Recent findings

Molecular replacement using fragments of GyrA (a homologous protein from *E. coli*) resulted in interpretable electron density maps for approximately 50% of the molecule. Iterative rounds of model building and refinement generated electron density for the missing regions, allowing the entire polypeptide chain to be traced (Fig. 1).

The overall fold of topoIV 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. The relative orientation of the domains at the larger interface (boxed in Fig. 1) is significantly different to that seen elsewhere. The structure could therefore represent another crystallographic "snapshot" of an intermediate conformation involved in the catalytic cycle of topoIV, providing yet more information about the order of events within the protein during catalysis.

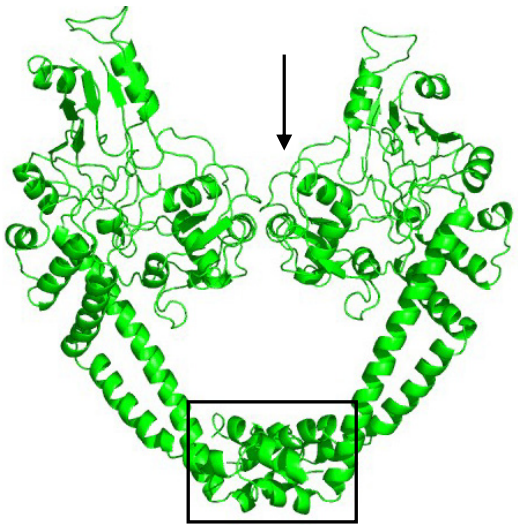


**Fig. 1.** TopoIV structure with domains in a novel orientation (boxed) and active site (arrow) indicated

The DNA-binding/cleavage site (indicated by an arrow in Fig. 1) contains two tyrosine residues (Tyr 119) positioned on either side of a positively-charged DNA-binding groove. These residues undergo a transesterification reaction, attacking the "target" phosphoryl groups forming a covalent link to the DNA. This results in a covalent link between protein and DNA and cleavage of the DNA strand. There are many conserved, basic residues in the vicinity of Tyr119 thought to be involved in stabilisation of the transition state and of the 3' end of the DNA in the complex containing cleaved DNA (Figs. 2a and b).

This region is also involved in the binding of quinolone antibiotics and naturally occurring resistant phenotypes have altered residues at these positions (Fig. 2a). The various quinolones also display differing propensities for inactivating topoIV from various organisms and it is likely that changes in these residues contribute to this preference. The quinolone interaction region of GrlA contains of a number of different residues to those seen in the homologous proteins from *E. coli*, and these differences could contribute to variation in quinolone activity against topoisomerases from these different sources.

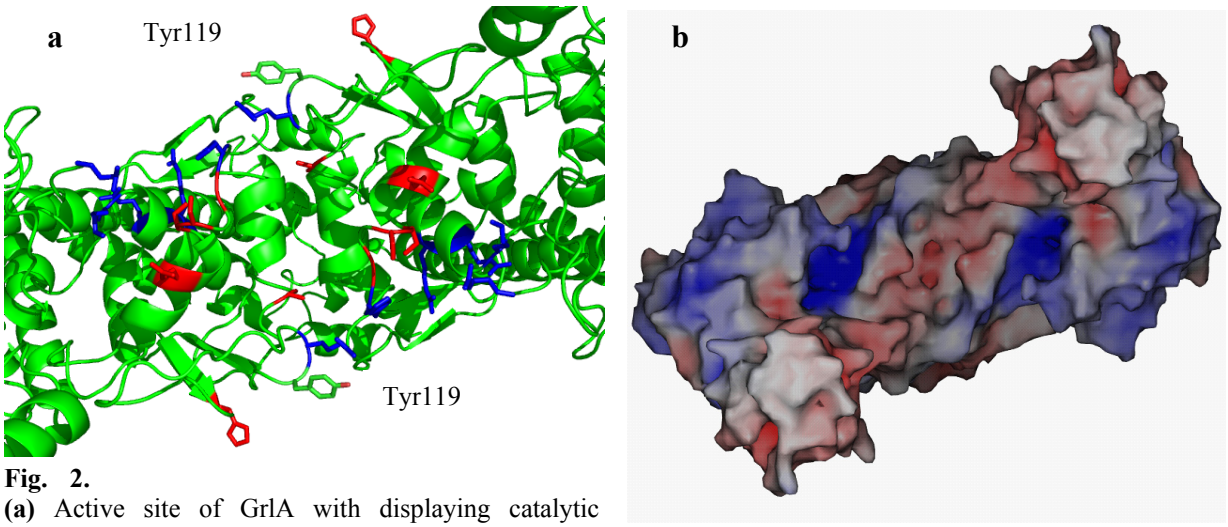



**Fig. 2.**
**(a)** Active site of GrlA with displaying catalytic tyrosine residues and conserved basic residues in blue and residues involved in quinolone binding in red. **(b)** Surface representation of GrlA displaying electrostatic charges highlighting the basic nature of the active site, positive charged regions are coloured blue and negative charges coloured red.

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

# Molecular mechanism of Staphylococcal plasmid transfer

Jamie A. Caryl and Christopher D. Thomas

**Background**

Horizontal gene transfer in bacteria results in genetic diversity with important medical consequences. Small, non-self transmissible, mobilisable staphylococcal plasmids such as pC221 offer a simple system that embodies the initial events in plasmid mobilisation. pC221 is a 4.6 kb chloramphenicol resistance plasmid of *Staphylococcus aureus*. Although not self-transmissible it can be mobilised by a co-resident self-transmissible plasmid such as pGO1. Typically for a small plasmid, pC221 contains only those genes required for its own DNA processing and contains four such loci: an origin of transfer (*oriT*); a DNA relaxase, MobA; and the putative accessory proteins MobB and MobC.

**Recent findings**

MobC binds to *oriT* at two sites (*mcb*) containing a 9 bp consensus sequence, and additionally within the *mobC* gene at a 7 bp conserved sequence. The *oriT* of pC221 has been functionally characterised with respect to the sites of Mob protein-DNA interaction. A functional region, sufficient for nicking, is defined by 77 bp sequence encoding a 9 bp MobC binding site (*mcb2*), a site of MobA interaction (*sra*) and the nick site (*nic*). Mutagenesis of *mcb2* has demonstrated that binding of MobC is a pre-requisite for subsequent nicking by MobA.

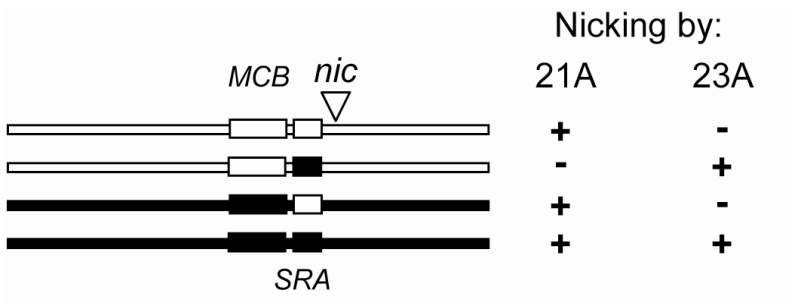



**Fig. 1** Schematic detailing SRA-swap between pC221 *ori*T (white) and pC223 *ori*T (black), and the nickability of recombinant substrates by pC221 MobA (21A).or pC223 MobA (23A).

The MobA relaxase recognises a plasmid-specific sequence within *sra*, permitting cleavage at *nic*. The *sra* of pC221 and pC223 differ by 4 bp. Swapping of these base-pairs between the two substrates, which otherwise share identical nick sites, effectively exchanges the substrate specificity and therefore recognition by their respective cognate relaxases (Fig. 1).

Recently, each of the MobC binding sites has been mutated, in the context of pC221, to investigate the requirement of each in isolation. Four mutant plasmids that differ from the wild-type only at *mcb1*, *mcb2*, *mcb3* or *mcb1+3* were assayed for *in vivo* nicking and mobilisation activity. In each case, mutation was found to abolish mobilisation. Agarose gel electrophoresis of whole cell lysates showed that with the exception of the *mcb2*, all the mutant plasmids were susceptible to nicking. Thus MobC may function at several levels: primarily to initiate nicking by MobA, presumably via interaction at *mcb2*; but furthermore by mediating a potential high-order complex formation, additionally involving *mcb1* and *mcb3*, to yield a mobilisable substrate.

**Publications**

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

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

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

# Atomic force microscopy of DNA structure and interactions

Neal Crampton, Fabien Bourbigot and Neil Thomson

The atomic force microscope (AFM) is being used increasingly to studied DNA molecules and their interactions at the single molecule level. While the resolution is usually only sufficient to map the overall morphology of DNA, i.e. tertiary and quaternary structure, the contrast is high enough to easily distinguish protein binding, the positions of which can be accurately mapped to a few nanometres. AFM has been used previously to quantify the wrapping of dsDNA around *E. coli* RNA polymerase and DNA gyrase on linear templates, while the exit to entry angle of the DNA contour through a protein has enabled analysis of DNA bend angles induced by RNA polymerases and DNA methylases.

## Competitive binding of DNA to mica

To image DNA by AFM, the molecules need to be immobilised to a support surface, usually mica crystals. These are layered and cleave easily with sticky tape to reveal a fresh hydrophilic surface for each experiment. Binding of DNA to mica can be controlled to some extent by varying counter-ion composition and concentrations in the deposition buffer. Binding of DNA to mica is limited by the diffusive motion of the DNA in solution to the surface in sufficiently dilute solutions. Recently, we have shown that for a two component system this process can either be inhibitive or competitive for the larger fragment relative to the smaller one. Fig. 1 shows a plot of the $t^{1/2}$ dependence of the surface molecular densities for a 1074bp DNA fragment.

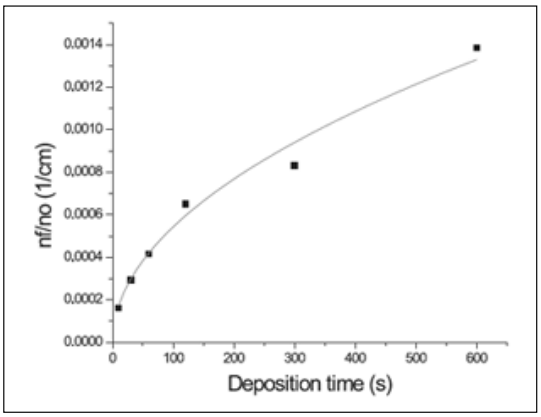



**Fig. 1**: The deposition kinetics of a 1074bp dsDNA fragment onto mica in a Mg(II) containing buffer. The experimental data are fitted to $t^{1/2}$ dependence.

## Tailoring of DNA binding to surfaces

Mica surfaces can be modified using self-assembled monolayers to control the functionality of the surface and the DNA binding characteristics. Recently, we have studied in detail the formation of aminosilane layers on mica to produce positively charged surfaces on which to bind DNA for AFM studies. One can see from Fig. 2 that (linear) DNA adopts different conformations on bare mica and mica modified with the aminosilane (APTES). This reflects the two different binding mechanisms under these conditions, termed surface equilibrated (mica) and kinetically trapped (APTES-mica).

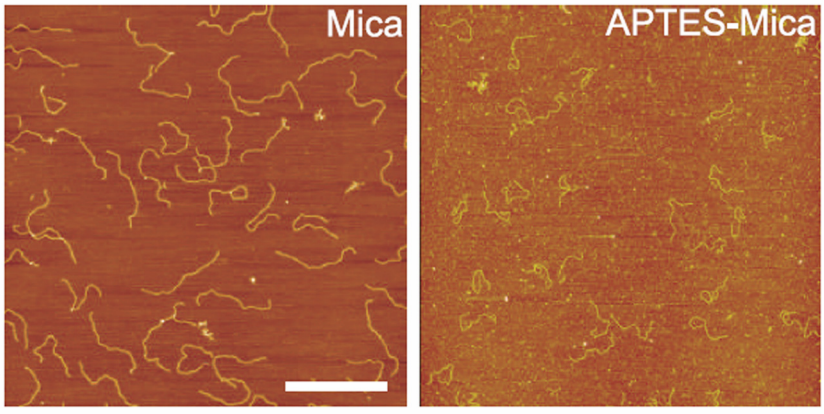



**Fig. 2**: Tapping-mode AFM images of 1074bp deposited on mica and APTES-mica and dried. Scale bar = 500nm.

**Studying convergent transcription by AFM**

In a collaboration, we have been using AFM to study what happens at the molecular level when two genes are simultaneously expressed. "Nested gene" is a term coined for a gene that lies completely within the sequence of another gene and frequently in the opposite orientation. Several such genes are known to exist in the human genome and eukaryotes such as *Saccharamyces cerevisiae.* Whether these genes come about as an unavoidable consequence of their compressed genetic arrangement, or whether such genes have an intrinsic effect on transcriptional regulation is unclear. Preliminary sets of experiments demonstrate that in a significant proportion of the active complexes, the RNAPs, stall against each other and are unable to pass.

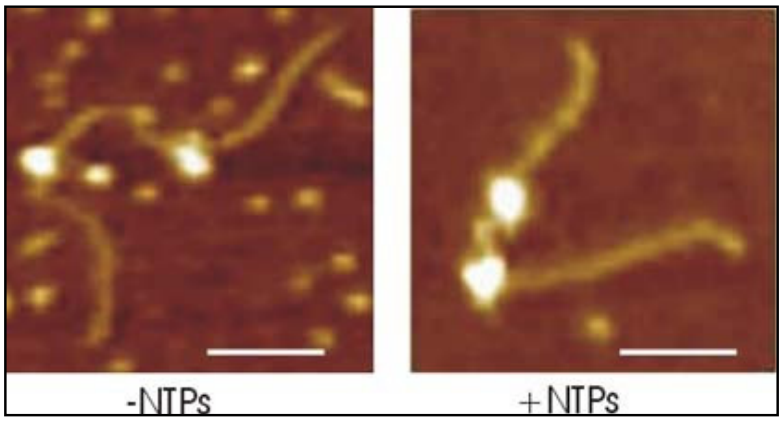



**Fig. 3**: TM AFM images in air of an 1150bp DNA template with two *E. coli* RNA polymerases bound and deposited on mica before and after addition of ribonucleoside triphosphates (NTPs). In the absence of all 4 NTPs the polymerases remain at their promoters (left panel). Addition of NTPs cause the RNAP to initiate from their promoters, transcribe the DNA convergently and, in this particularly case, collide and stall while still physically separated (right panel). Scale bars = 75 nm.

**Collaborators**

- Jennifer Kirkham and William A. Bonass (Oral Biology, Leeds Dental Institute, UK)
- Carolyn W. Gibson (University of Pennsylvania, USA)
- Claudio Rivetti (University of Parma, Italy)

**Publications**

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

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,* in press.

Crampton, N., Bonass, W.A., Kirkham, J. & Thomson, N.H. (2005) Formation of aminosilane functionalised mica for atomic force microscopy of DNA. *Langmuir* **21**, 7884 - 7891.

Gibson, C.W., Thomson, N.H., Abrams, W.R. & Kirkham, J. (2005) Nested genes: biological implications and use of AFM for analysis. *Gene* **350**, 15-23.

**Funding**

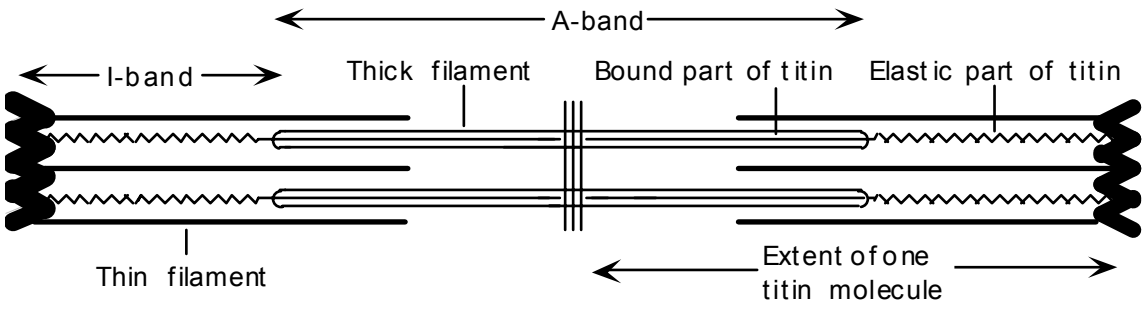
We gratefully acknowledge the EPSRC and University of Leeds for financial support. Neil Thomson is a Leeds University Advanced Fellow.

# Titin

Larissa Tskhovrebova, Tanniemola Liverpool, Tom Waigh, Emmanuela Di Cola,
Ahmed Houmeida, Nasir Khan and John Trinick

Titin is the largest known protein (chain weight 3-3.7 MDa) and the third most abundant protein of vertebrates skeletal and heart muscle. More than half the titin molecule is bound to myosin filaments in the sarcomere, where we have suggested it regulates assembly of the exactly 294 myosin molecules known to comprise the filament. The remainder of the titin molecule forms an elastic connection between the end of the thick filament and the Z-line, which is the main source of elasticity of relaxed muscle. This elastic element also ensures that the myosin filaments stay in the middle of the sarcomere, which ensures that even forces are developed by myosin in opposing halves of the filament during active contraction.

Studies of titin elasticity are of considerable interest and have notably been pursued in many single molecule studies (including some by ourselves) using optical tweezers and AFM. It has been temping to assume that the single molecule data can be scaled directly to explain muscle passive elasticity. However, several factors suggest that such direct extrapolation may not be possible, including interactions of the elastic region of titin *in situ*, both with itself and with other proteins. Effects of molecular crowding and confinement within the lattice of actin (thin) filaments are also under consideration, which would also be expected to modulate the the behaviour of the molecule.




## Publications

Tskhovrebova, L. & Trinick, J. (2005) Muscle disease: a giant feels the strain. *Nature Medicine* **11**, 478-479.

Di Cola, E., Waigh, T. A., Trinick, J., Tskhovrebova, L., Houmeida, A., Pyckhout-Hintzen, W., Dewhurst, C. (2005) Persistence length of titin from rabbit skeletal muscles measured with scattering and microrheology techniques. *Biophysical Journal* **88**, 4095-4106.

Tskhovrebova, L., Houmeida, A. & Trinick, J. (2006) Can the passive elasticity of muscle be interpreted directly from the mechanics of individual titin molecules? *J. Musc. Res. Cell Motil.* (published online).

## Funding

British Heart Foundation

# Myosin XVIII

Kasim Sader and John Trinick

Myosins are identified on the basis of sequence similarities in their motor domain regions, which bind actin and have ATPase activity. Extending from the motor domain is a single α-helix, which binds light chains of the calmodulin class. This acts as a lever arm and produces motion by changing its angle relative to the motor domain. At the end of the lever arm the molecule tail binds cargo. The vast majority of work on myosin has focussed on the type found in muscle, which is called class II. Muscle myosin is a dimeric molecule, with two heads, each with a motor domain and a lever arm. The heads are dimerised through the tail, which is a long (150 nm) coiled-coil a-helix and polymerises to form muscle thick filaments.

Recent years have seen the identification of many other, non-muscle myosin classes. Based on their sequences, these have similar motor domains, but vary widely in other respects. They may or may not dimerise and are also likely to vary in kinetic properties, regulatory mechanisms and cargo binding. Mouse myosin 18A expression levels have been correlated with the hematopoietic ability of stromal cells in bone marrow. Two isoforms have been identified: the alpha isoform contains an N-terminal extension with a lysine-glutamic acid (KE) rich sequence followed by a PDZ domain, which is thought to be involved in protein-protein interaction. The beta isoform is identical but without the N-terminal extension. Myosin 18A has the longest sequence of any unconventional myosin, with a predicted long coiled-coil alpha helical tail. The conserved sequences for actin binding in the motor domain are not present, but actin binding has been demonstrated via an ATP-insensitive site in the N-terminal extension by in vivo expression in HeLa cells. Using baculovirus, we have expressed HMM constructs of the alpha and beta isoforms. We have also expressed S1-like constructs of both isoforms, truncated immediately after the IQ motif. We have shown the HMM construct to be a dimer by negative stain electron microscopy. We are also carrying out electron microscopy and single particle image processing of the S1 constructs, with and without ATP and bound to actin.

**Collaborators**
James Sellers and Yi Yang, National Institutes of Health, Bethesda, Maryland, USA

**Funding**: We acknowledge funding from the Wellcome Trust.

# Class VI myosins

Matthew Walker and John Trinick

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

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

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

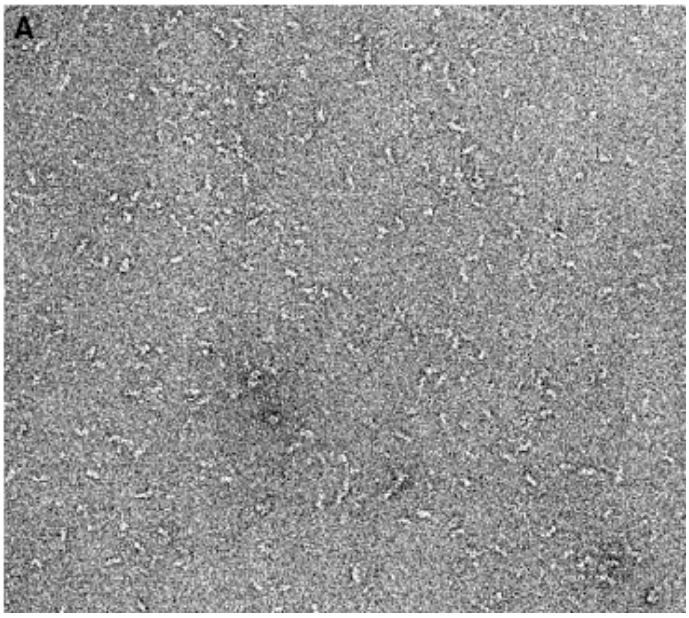



**Fig. 1**: Field of negatively stained monomeric myosin VI molecules (each about 16 nm long.

**Publications**

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

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

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

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

**Funding**
We thank the BBSRC and NIH (USA) for financial support.

# Structural and enzymatic studies on a novel substitutient of *Mycobacterium tuberculosis* lipoarabinomannan

Martin Fascione, Susanne Hesketh and Bruce Turnbull

## Introduction

Tuberculosis (TB) remains a major threat to world health, with approximately eight million cases of TB resulting in two million deaths per annum. The WHO has highlighted the urgent need for more effective anti-TB drugs and for rapid, specific and sensitive diagnostic tests for the causative agent, *Mycobacterium tuberculosis* (Mtb). The lipid-anchored polysaccharide, lipoarabinomannan (LAM, Fig 1), is a major component of the Mtb cell wall, and an important virulence factor for the bacterium. Mannose residues that cap the dendritic arabinan chains of LAM, facilitate entry of Mtb into alveolar macrophages, following interaction with the macrophage mannose receptor. LAM then promotes the intracellular survival of Mtb by down-regulating the immune response and providing anti-oxidative protection for the bacterium.

Recently, we discovered an unusual methylthiopentosyl residue attached to the mannosyl caps of LAM, which we subsequently identified as having an α-xylo configuration. This unusual discovery constitutes the first report of a methylthio-sugar residue incorporated into a polysaccharide, and one of very few examples of a xylo-configured sugar outside the plant kingdom. Mtb invests significant biosynthetic effort into incorporating MTX into its cell wall, which implies that this sugar may provide some advantage to the bacterium. The concurrent discovery of an oxidised form of the sugar (methylsulfinylxylose, MSX, Fig 1), implies that MTX may play a role in oxidative protection for Mtb. Current studies are directed towards refining the structure of the oligosaccharide. We are also investigating enzymatic processes for the interconversion of MSX and MTX which could provide a mechanism for protecting the bacterium from reactive oxygen species *in vivo*.

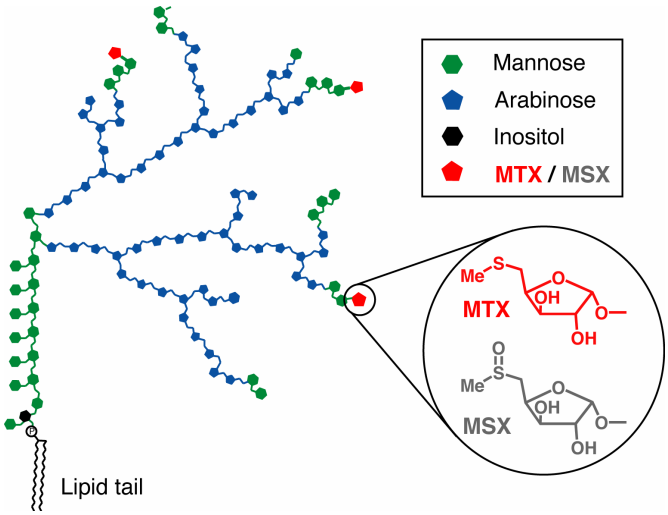



**Fig. 1.** Schematic representation of LAM showing the structures of the MTX and MSX substituents

## Collaborators

Delphi Chatterjee, Department of Microbiology, Immunology and Pathology, Colorado State University, USA.
Subramanian Dhandayuthapani, University of Texas Health Science Center, San Antonio, TX, USA.
Achim Treumann, Royal College of Surgeons in Ireland, Dublin, Ireland.

## Funding

This work is funded by the Royal Society and the University of Leeds.

# Machine learning to predict gene and protein function

James Bradford, Matthew Care, Andrew Garrow, and David Westhead

## Introduction

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

## Protein-protein binding site 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. This last year we have focused on predicting both protein-protein binding site location and interaction type using Bayesian networks in combination with surface patch analysis. In doing so, insights have been gained into the properties that characterise a binding site and drive complex formation. Our method predicts protein-protein binding sites with a high success rate of 82% on a benchmark dataset of 180 proteins, improving on previous work by 6% (see Bradford & Westhead 2005). The method was also able to handle incomplete datasets automatically. With this in mind, we also carried out a study on the Mog1p family for which evolutionary information was sparse and were able to suggest binding sites for Ran and other signalling proteins on Mog1p itself. Our results on other members of the family suggest that proteins can still bind to different proteins and probably have different functions even though they share the same overall fold. We also 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. In a separate study of obligate and non-obligate interfaces, we found that such was the similarity between the two types, we were able to use obligate binding site properties to predict the location of non-obligate binding sites and *vice versa*.

## Modelling the effect of missense mutations on protein function

Prediction of the effects of non-synonymous single nucleotide polymorphisms (nsSNPs) has been studied by various research groups using a variety of probabilistic and machine learning tools. Most methods use a range of structural and sequence attributes to try and predict deleterious or missense mutations that affect protein function.

Bayesian networks have successfully been applied to two protein mutagenesis datasets (*lac* repressor and T4 lysozyme) yielding results that are comparable with those produced by other machine learning techniques. In addition, the results showed that Bayesian networks generalise well to new data, are robust to training from incomplete data, and handle missing data such as structural or evolutionary information. Having discovered the most important contributors to prediction, we reduced our Bayesian network from 15 to only four nodes. This simpler model, even though no evolutionary information was used, maintained similar classification performance to the full network.

Current work has involved producing a larger dataset of SNPs to more accurately predict their effects. The Swiss-Prot Variant database of Human protein variants was parsed to generate ~12,000 disease SNPs (from ~1000 proteins) and ~8,000 polymorphic SNPs (from ~3000 proteins). It is hoped that this diverse "real world" dataset can be used to train machine learning algorithms to analyse existing un-annotated SNP databases.

**Searching genomes for trans-membrane barrel proteins**

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

In this project high accuracy composition based discrimination algorithms have been developed using a number of machine learning techniques (e.g. support vector machines (SVMs) and genetic algorithms; see Garrow *et al.* 2005). Another related project has focused on development of Hidden Markov Models for detection of trans-membrane strands.

**Collaborations**

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

**Publications**

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

Garrow, A.G., Agnew, A. & Westhead, D.R. (2005) TMB-Hunt: An amino acid composition based method to screen proteomes for beta-barrel transmembrane proteins. *BMC Bioinformatics* **6**, 56.

Bradford, J.R. & Westhead, D.R. (2005) Improved prediction of protein-protein binding sites using support vector machines. *Bioinformatics* **21,** 1487-1494.

**Funding**

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

# Storage and analysis of microarray data

Archana Sharma-Oates, Phil Quirke and David Westhead

## Introduction

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

## Human cancer pathology

Tissue samples are obtained from individuals in a clinical trial (both cancerous and normal tissues from the same person) for biochemical and histopathological analyses. Depending on the nature of the trial, these tissue samples undergo extensive characterisation using a number of high-throughput molecular biology techniques. The high-throughput techniques most commonly used in cancer research are cDNA microarrays, comparative genomic hybridisation arrays and tissue microarrays. The purpose of the cDNA microarray approach is to gain an insight into the expression levels of all the predicted genes in the human genome with the aim of identifying a set of genes related to a clinical outcome that may be either up or down regulated in tumour verses normal tissue. Comparative genomic hybridisation (CGH-arrays) arrays are used to study chromosomal instability at a genome level within tumour verses normal tissues. TMA is a technique that enables the analysis of a large cohort of clinical specimens in a single experiment thereby studying the molecular alterations (at the DNA, RNA, or protein level) in thousands of tissue specimens in parallel. The aim of cDNA microarray and CGH-array techniques are to either identify biomarkers that can be verified by TMA. Our involvement in this research involves analysis of the cDNA microarray and CGH-array data using statistical approaches, and the development of storage and analysis software for tissue microarray experiments, an area where we are contributing to the development of international standards, and the integration of this data with MIAME compliant microarray databases.

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

**Collaborators:** Professor P. Quirke

## Publications

Sharma-Oates, A., Quirke, P. & Westhead, D.R. (2005) TmaDB: A repository for tissue microarray data. *BMC Bioinformatics* **6**, 218

Sharma-Oates, A., Quirke, P., and Westhead, D.R. (2005) TmaDB: A tissue microarray database. *J Pathol* **204**, 51A-51A Suppl.

# Molecular modelling of protein structural behaviour

Binbin Liu, Sally Mardikian, Richard Jackson and David Westhead

## Introduction

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

## Modelling the electrostatics of conduction in the voltage-gated $K^+$ channel, Kv1.4, and the effects of charged residues on the $K^+$ occupancy

$K^+$ channels are integral membrane proteins that are universal regulators of cellular function both in excitable and non-excitable cells. The voltage-gated $K^+$ (Kv) channel, Kv1.4, is expressed in the mammalian heart and is partly responsible for the transient outward current. Electrophysiological experiments on key amino acids in Kv1.4 channels show that the charged residues (H508 and K532) in extra-cellular vestibules affect C-type inactivation by changing the pH of the cellular environment. The aim of this current study is to investigate how these charged residues bring about the energy changes that affect the C-type inactivation conduction mechanism.

3D models of Kv1.4 (TM5, p-loop and TM6 domains) were constructed using comparative modelling. Crystal structures of KcsA and Kv1.2 were used as templates in the construction. The electrostatic potential profiles along the pore were calculated by solving the Finite Difference Poisson-Boltzmann equation (FDPB) using the continuum electrostatic model. The differences in potential energy were calculated with H508 and K532 in different protonation states. Electrostatic potential profiles of Kv1.4 with different numbers of $K^+$ ions occupying the selectivity filter binding/coordination sites (S0, S1-S5) were also calculated.

The continuum electrostatics calculations showed that protonation of H508 decreases the interaction energy at the S1 $K^+$ binding site in the selectivity filter by ~5 kJ/mol. Consequently, protonation of H508 is predicted to lead to about a 7-fold reduction in the $K^+$ occupancy at S1. The neutralisation of the positively charged K532 increases the potential energy of a $K^+$ ion binding at S1 by ~10 kJ/mol. Thus, the loss of charge at K532 is predicted to lead to about a 55-fold increase in $K^+$ occupancy at S1. In conclusion, continuum electrostatics calculations predict that the protonation states of H508 and K532 have substantial effects on the $K^+$ occupancy at the S1 binding site.

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

The process of predicting a protein-ligand complex from its constituents' atomic coordinates is known as protein-ligand docking. The docking procedure incorporates two components: the search component explores the ligand's conformational degrees of freedom, whereas the scoring component employs a scoring function to assess the quality of the generated ligand conformations. Most scoring functions use some form of energy evaluation to achieve this. For example, the GRID scoring function calculates non-bonded interaction energy terms (electrostatics, van der Waals, hydrogen bond donor and acceptor) between the protein and ligand atoms, and adds these together to give a total energy. This is regarded as the score for the interaction between the protein and the ligand at that conformation. This process does not take into account the strengths of the different contributions of the energy components and so the influences of all the contributions within the scoring function are assumed to be equal.

We have developed a docking method based on a multiobjective genetic algorithm (MOGA) that is capable of simultaneously optimising individual interaction energy types of the GRID scoring function. The MOGA aims to lead the conformational search to what is known as the Pareto front. Solutions that fall on the front (the Pareto solutions) show different compromises in the different objectives, and they are all considered to be equivalent. By examining where the true solution falls on the Pareto surface, we can gain insights into the relative importance of different contributions to binding for a given protein-ligand complex.

Twenty protein-ligand complexes have been tested using the MOGA optimisation method. For seventeen of these, the MOGA was able to obtain good solutions within the Pareto solution set (RMSDs of less than 2.0Å from the crystal structure, see Fig. 1). An interesting feature of these solutions is that the different objectives do not show an equal influence. In fact, for many of the complexes, the van der Waals interactions alone appear to be guiding the search. These results indicate that the strengths of different interaction energy types vary greatly from one complex to another, and that a scoring function should, in fact, be tailored to a given target. Our method could be used for scoring function development, by assessing the balance of the different interaction types for a single, or a group of related targets. It could also be used as a stand-alone docking tool that allows the user to select the most "realistic" of the Pareto solution set.

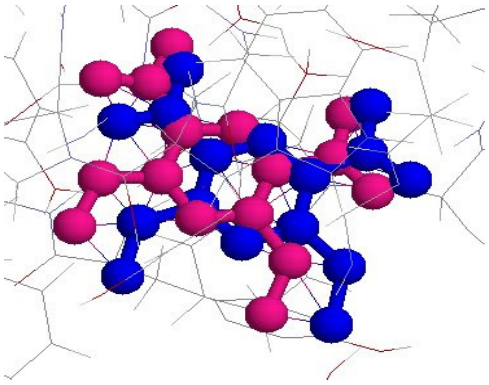


**Fig. 1:** The orientation of Fab fragment of murine monoclonal antibody crystal (blue) relative to a docked solution output from the program (RMSD=1.4Å). The ligand is complexed with tetramethyl dinitrophenyl (1baf).

**Collaborators**
Dr.Val Gillet, University of Sheffield
Professor Mark R Boyett, University of Manchester.

**Funding**
Binbin Liu is funded by the University of Manchester.
Sally Mardikian is funded by the Medical Research Council.

# Reconstruction and analysis of biochemical networks

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

## Introduction

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

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.

## The metaSHARK project

We have developed a comprehensive suite of programs for the representation and analysis of metabolic networks. The **meta**bolic **S**earc**H** **A**nd **R**econstruction **K**it (*metaSHARK*) includes an object-oriented database to store knowledge about networks of chemicals and reactions, as well as an automated system to search an unannotated genome for genes with significant similarity to known enzymes from other organisms. These genes are assigned a confidence score based on the strength of their similarity to the test sequences.

The 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 been incorporated into metaSHARK so that gene expression can be visualised on top of metabolic network diagrams (Fig. 1b), which allows 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.

This analysis is being extended to perform systematic analysis of the whole network, and identify all of the pathways of interest for a particular condition. This, combined with network structural analysis, will allow a characterisation of the full metabolic response of the organism in a specific experiment. This data can also be combined with sequence analysis and other genomic data to produce a list of candidate genes for enzymatic functions that appear to be missing from a particular pathway.

metaSHARK will also be expanded to integrate the regulation of metabolic genes at the transcriptional level, in particular the involvement of transcription factors in this process. The analysis will involve the examination of gene expression profiles for coexpression between metabolic enzymes and transcription factor genes, and whether the activation of these factors is life stage specific. This sort of information is vital for the purposes of identifying good drug targets in pathogens such as *Plasmodium falciparum*, the parasite causing the most virulent type of malaria. Analysis of the recently-released complete parasite genome has revealed novel genes and pathways which are being verified by RNAi experiments in Dr McConkey's group. We hope that this work will ultimately lead to the identification of new drug targets for malaria.

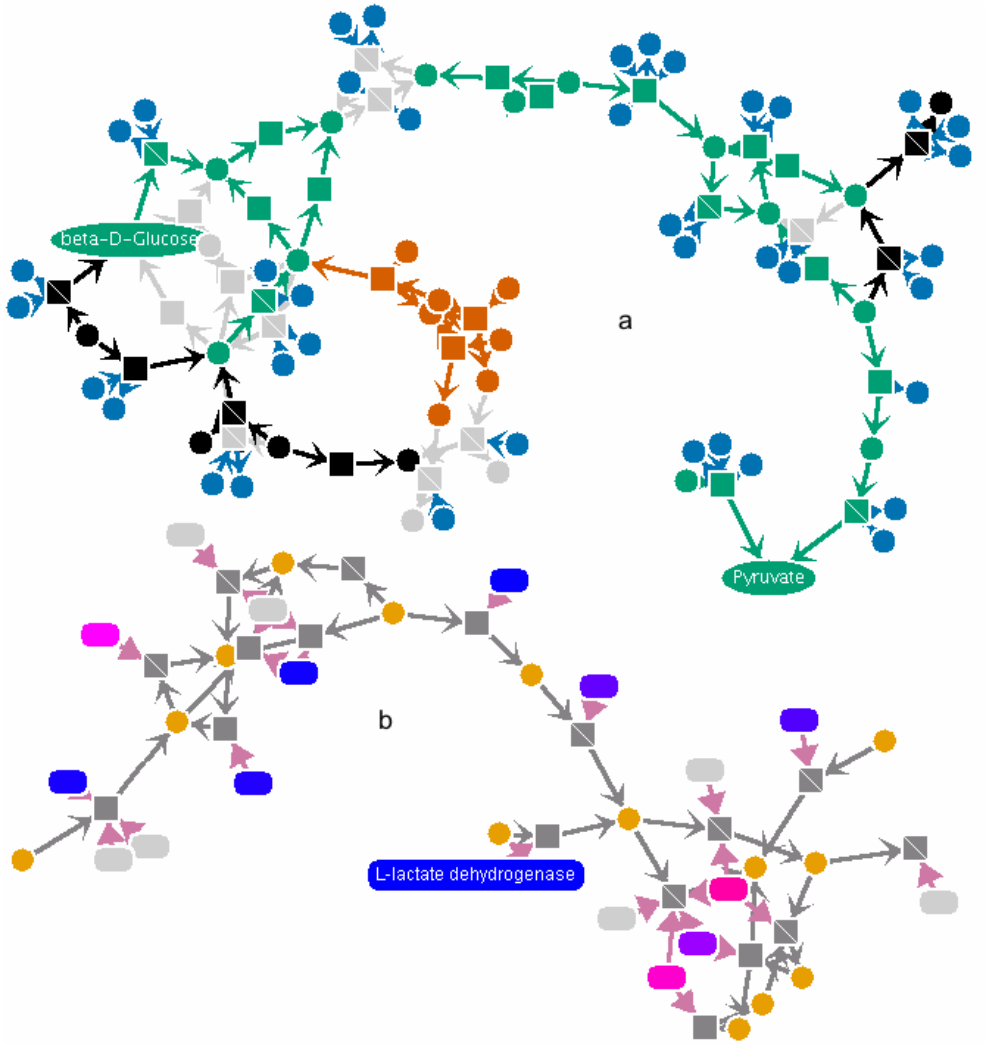



**Fig. 1(a)**: Part of an automated metabolic reconstruction of the avian intracellular parasite *Eimeria tenella*, visualised as a Petri net. Metabolites are represented as circles; reactions as squares. Directed arcs between nodes show the effect of a reaction as the consumption and production of different metabolites. Blue circles represent ubiquitous ("pool") metabolites, such as water, ATP, NADH etc.. Nodes in green show reactions catalysed by enzymes for which good evidence has been found in the *E. tenella* genome. Nodes in red show reactions catalysed by enzymes for which only tentative evidence has been found. Grey nodes show that no evidence for the catalysing enzymes was found, whilst black nodes indicate that no model data was available in our database, hence no gene search could be performed. The reconstruction clearly shows the presence of a glycolytic pathway in this organism as a route in green from beta-D-glucose to pyruvate. **(b)** A section of the glycolysis pathway of *Plasmodium falciparum*, with the enzyme nodes (round-edged rectangles), coloured according to their level of coexpression with lactate dehydrogenase. Blue nodes show a high positive coexpression, whereas a colour towards pink or red shows a high negative coexpression.

## Publications

Pinney, J.W., Shirley, M.W., McConkey, G.A. & Westhead, D.R. (2005) metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Res.* **33**, 1399-1409.

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

## Funding

We thank the MRC and BBSRC for funding this project.

# Herpesvirus-host cell interactions which regulate viral gene expression

Jim Boyne, Rhoswyn Griffiths, Kevin Colgan, James McBride and Adrian Whitehouse

## Introduction

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

## The interaction of the major transcriptional control protein, ORF 50 and viral lytic promoters.

The ORF 50 gene product, also known as the replication and transcription activator (Rta), is an immediate early gene which is well conserved among all gamma-2 herpesvirus and plays a pivotal role in regulating the latent-lytic switch. Herpesvirus saimiri (HVS) ORF 50a functions as a sequence specific transactivator capable of activating delayed early (DE) gene expression via binding directly to an ORF 50 response element (RE) within the respective promoter. Analysis of the ORF 50 REs has identified two distinct types within HVS gene promoters. The first comprises a consensus sequence motif, $CCN_9GG$, the second an AT-rich sequence. Here we demonstrate that ORF 50a is capable of transactivating the DE ORF 9 promoter which encodes the DNA polymerase. Deletion analysis of the ORF 9 promoter mapped the ORF 50 RE to a 95 bp region situated 126 bps upstream of the initiation codon. Gel retardation analysis further mapped the RE to a 28 bp fragment, which was able to confer ORF 50-responsiveness on an enhancerless SV40 minimal promoter. Furthermore, sequence analysis identified multiple CCAAT Enhancer Binding protein alpha (C/EBPα) binding sites within the ORF 9 promoter and specifically two within the close vicinity of the AT-rich ORF 50 RE. Analysis demonstrated that the HVS ORF 50a and C/EBPα proteins associate with the ORF 9 promoter *in vivo*, interact directly and synergistically activate the ORF 9 promoter by binding to adjacent binding motifs. Overall, these data suggest a cooperative interaction between HVS ORF 50a and C/EBPα proteins to activate the DNA polymerase promoter during early stages of the lytic replication cycle. We are presently characterising these protein-DNA interactions using surface plasmon resonance (SPR) and fluorescence resonance energy transfer (FRET).

## The interaction of the nucleocytoplasmic shuttle protein, ORF 57 and viral mRNA.

The ORF 57 protein encodes a nuclear cytoplasmic shuttle protein which mediates the nuclear export of viral mRNAs. We have recent analysis demonstrating that ORF 57 has the ability to bind viral RNA, shuttle between the nucleus and cytoplasm and is required for efficient nuclear export of viral transcripts (Fig. 1b). Moreover, we have shown that ORF57 shuttles between the nucleus and cytoplasm in a CRM-1 independent manner. ORF 57 interacts with the mRNA export factor REF and two other components of the exon-junction complex, Y14 and Magoh. The association of ORF57 with REF stimulates recruitment of the cellular mRNA export factor TAP, and HVS infection triggers the relocalisation of REF and

TAP from the nuclear speckles to several large clumps within the cell. Using a dominant negative form of TAP and RNA interference to deplete TAP, we show that it is essential for bulk mRNA export in mammalian cells and is required for ORF57 mediated viral RNA export. Furthermore, we show that disruption of TAP reduces viral replication. These data indicate that γ-2 herpesviruses utilises ORF57 to recruit components of the exon-junction complex and subsequently TAP to promote viral RNA export via the cellular mRNA export pathway. We now aim to analyse the domains required for these interactions in more detail using structural analysis.

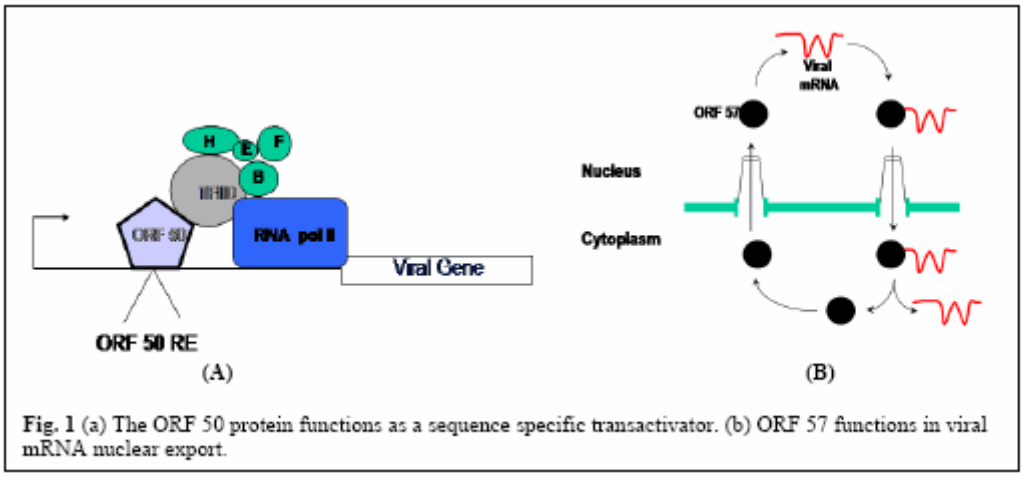



**Fig. 1** (a) The ORF 50 protein functions as a sequence specific transactivator. (b) ORF 57 functions in viral mRNA nuclear export.

**Publications.**


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

Williams, B., Boyne, J.R., Goodwin, D.J., Roaden, L.R., Wilson, S.A. & 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.

Smith, P.G., Burchill, S.A., Brooke, D., Coletta, P.L. & Whitehouse, A. (2005). Efficient infection and persistence of a HVS-based gene delivery vector into human tumour xenografts and multicellular spheroid cultures. *Cancer Gene Therapy,* **12**, 248-256.

Smith, P.G., Oakley, F., Fernandez, M., Mann, D., Lemoine, N.R. & Whitehouse, A. (2005). Herpesvirus saimiri-based vector biodistribution using non-invasive optical imaging. *Gene Therapy,* **12**, 1465-1476.

Calderwood, M.A., White, R.E. Griffiths, R.A. & Whitehouse, A. (2005). Open reading frame 73 is required for Herpesvirus saimiri All-S4 episomal persistence. *Journal of General Virology,* **86**, 2703-2708.

Wakenshaw, L., Walters, M.S. & Whitehouse, A. (2005). Herpesvirus saimiri Rta and cEBPα act synergistically to activate the DNA polymerase promoter. *Journal of Virology,* **79**, 13548-13560.


**Collaborators.**

Stuart Wilson, University of Sheffield; Ren Sun, UCLA, USA; Michael Calderwood, Harvard Medical School, USA; Nicholas Lemoine, Barts Medical School.


**Funding.**

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

# Design and synthesis of ligands for disruption of protein-protein interactions

Fred Campbell and Andrew Wilson

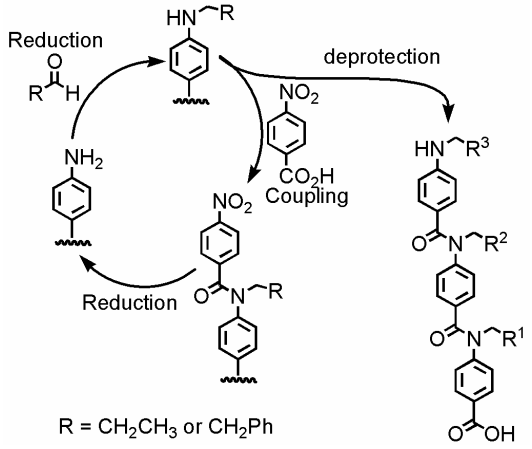
## Introduction

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 to mimic crucial amino acid residues involved in protein-protein interactions.

## Current work

We have developed a procedure for synthesis of trimers comprising *p*-amino benzoic acid alkylated at nitrogen (Fig. 1). This procedure furnishes trimers that are intended to mimic the presentation of i, i + 4 and i +7 residues of an α-helix. So far we have used this ten step procedure to obtain 100's of milligrams of compounds that incorporate simple R-groups e.g. ethyl and phenyl.

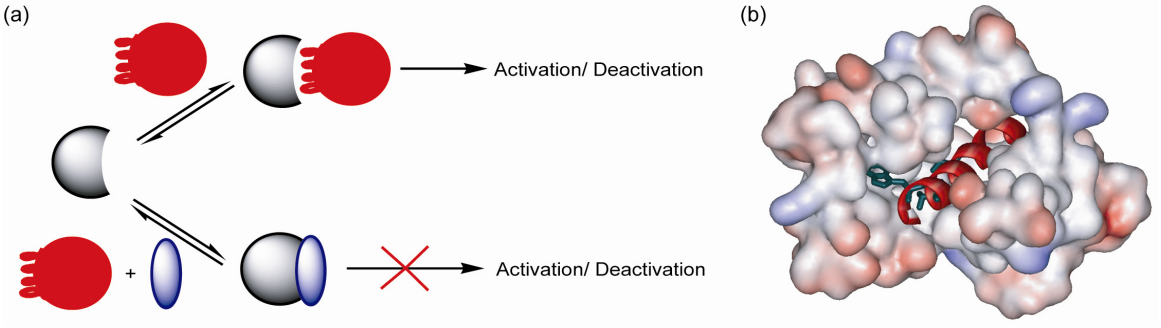
**Scheme 1.** The iterative procedure used for synthesis of *p*-aminobenzoic acid derived protein surface binding ligands




## Future work

We will use this method to synthesise trimers that act as inhibitors of the model protein calmodulin (Fig. 1a). This engages in protein-protein interactions through binding to the α-helices of proteins such as phosphodiesterase and smooth muscle myosin light chain kinase (sMLCK) (Fig. 1b).

**Fig. 1.** (a) Cartoon representation of 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 (Meador *et al Science* **1992**, *257*, 1251-1255) of the α-helical sMLCK peptide (red) bound to calmodulin (Connolly surface).




## Funding

This work is funded by the University of Leeds.

# The Membrane Protein Structure Initiative

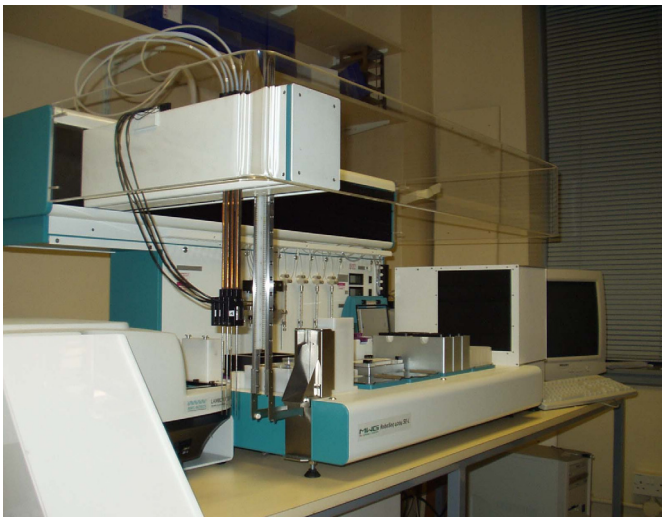
Vincent Postis, Xiaobing Xia, Peter Roach, Jean Ingram, Sophie Foppolo, Karine Deville, Moazur Rahman, Fouzia Ismat, Lijie Sun, Steve Baldwin, John Findlay, Peter Henderson, Paul Knox,
Mike McPherson and Simon Phillips

## Introduction

Genes encoding transmembrane proteins constitute 15-30% of most genomes. Because of their strategic localisation at the interfaces between cells and their environment, and between compartments within cells, these proteins play crucial roles in many cellular and physiological processes, including cell-to-cell signalling events and solute transport. For this reason, they also represent key targets for drug discovery. However, despite this importance, knowledge of their structures and mechanisms of action has lagged far behind our corresponding knowledge of water-soluble proteins. The BBSRC-funded Membrane Protein Structure Initiative (MPSI), a five-year programme of research involving the University of Leeds and seven other UK institutions, aims to rectify this deficiency. Membrane proteins of interest to MPSI include transporters and ion channels from prokaryotes and eukaryotes, with a particular emphasis on proteins related to those of physiological or clinical importance in humans. At Leeds, our focus is the development of high throughput (HTP) methods for the cloning, over-expression and functional characterisation of such proteins and for their solubilisation, purification and crystallization.

## Development of HTP approaches

During the first year of the project, a new laboratory for HTP expression and purification of membrane proteins has successfully been established on level 7 of the Astbury building, and a 100 litre fermenter for large scale cultivation of bacteria has been installed on level 6. Extensive bioinformatic analyses have been employed to identify an initial set of approximately 350 targets, primarily cation-linked transporters, representing a wide variety of protein families from a diversity of bacterial species. About a third of these targets have now been cloned into a family of membrane protein expression vectors designed to allow the rapid testing of multiple types and attachment sites for affinity tags. A key advance has been the introduction of robotic procedures for the parallel assessment of multiple host strains and growth conditions in the search for those yielding highest expression levels. Similarly, robotic approaches have been developed to screen detergent types and concentrations required for optimal solubilisation of membrane proteins. By these means, we can identify the best conditions for protein expression and solubilisation for multiple targets in parallel within a few days. So far, conditions for high level expression have been established for more than 50 targets, and about a dozen of these have been transferred to large scale preparation using our fermenter facility. Following affinity purification, multiple targets are now in crystallization trials, both here at Leeds and in the



Automated workstation for high-throughput screening of membrane protein expression.

laboratories of our MPSI partner institutions. Diffracting 3D and 2D crystals have already resulted in several cases. During the coming year, we plan to extend the development of HTP approaches to target cloning and functional characterization.

**Collaborators**

Neil Isaacs, Richard Cogdell (University of Glasgow), Robert Ford, Jeremy Derrick, Steve Prince (University of Manchester), Per Bullough (University of Sheffield), So Iwata (Imperial College), Mark Sansom (University of Oxford), Bonnie Wallace (Birkbeck College), Miroslav Papiz (CCLRC Daresbury)

**Funding**

This work was funded by the BBSRC under the SPoRT (Structural Proteomics of Rational Targets) initiative.

# Astbury Seminars 2005

**Friday 14th January 2005**
Dr Derek Marsh, Max Planck Institute for Biophysical Chemistry, Gottingen, Germany
*"Membrane lipids and beta-barrel proteins"*

**Thursday 20th January 2005**
Dr Thomas Wileman, Institute for Animal Health
*"Exploitation of Cytoskeletal Motors by very large Viruses: Potential Use of African Swine Fever Virus as a Nano-delivery Machine"*

**Thursday 10th February 2005**
Prof Jonathan Hirst, School of Chemistry, University of Nottingham
*"Quantum Biospectroscopy"*

**Thursday 10th March 2005**
Professor Michael Overduin, Division of Cancer Studies, University of Birmingham
*"Phospholipid Binding Proteins: The Good, The Bad and The Ugly "*

**Thursday 14th April 2005**
Dr Charlie Laughton, School of Pharmacy, University of Nottingham
*"Molecular Flexibility and Molecular Recognition in DNA "*

**Thursday 21st April 2005**
Dr Martin Web, MRC National Institute for Medical Research
*"DNA Helicases: Sensing Movement along DNA"*

**Thursday 5th May 2005**
Dr Jame Parker, Structural Biology Section, Institute of Cancer Research, London
*"Insights into the Mechanism of RNAi from Studies of PIWI Domain - RNA Interactions"*

**Thursday 14th July 2005**
Professor Haiyan Liu, University of Science and Technology of China
*"Computer Simulation of Transition Paths: Peptide Folding in a Continuum Solvent and Enzyme Catalysis Modeled by QM/MM"*

**Tuesday 26th July 2005**
Professor Peter Maloney, Department of Physiology, Johns Hopkins University School of Medicine
*"The Biology and Biophysics of OxlT, a Virtual Proton Pump"*

**Thursday 11th August 2005**
Dr Thomas Edwards, School of Biochemistry & Microbiology, University of Leeds
*"Translational Regulation: The Structure of the Hunchback Repression Complex"*

**Thursday 18th August 2005**
Dr Arwen Pearson, University of Minnesota
*"Probing the Biogenesis of Tryptophan Tryptophylquinone, the Novel Protein Derived Cofactor of Methylamine"*

**Thursday 29 September 2005**
Dr Robert Best, NIH Bethesda, USA
*"Assembling Composite Energy Landscapes from Multiple Experimental Structures: The Helix/Sheet Transition of Arc Repressor"*

**Friday 7th October 2005**
Dr Raghupathy Sarma, Department of Biochemistry and Cell Biology, State University of New York at Stony Brook
*"Fiftieth Anniversary of the Collagen Structure Determination: Ramachandran - Crick Controversy Revisited"*

**Thursday 13th October 2005**
Professor Gideon Davies, Department of Chemistry, University of York
*"Structures and Mechanisms of Glycosyltransferases: What we know, What we don't know and What we think we might know"*

**Monday 24th October 2005**
Dr Roman Tuma, Department of Biosciences & Institute of Biotechnology, University of Helsinki
*"Cooperativity and regulation of a hexameric molecular motor "*

**Thursday 17th November 2005**
Professor Gerhard Klebe, University of Marburg
*"Insights into Binding Affinity by Isothermal Titration Calorimetry, X-ray crystallography and Electrostatic Calculations"*

**Thursday 15th December 2005**
Professor Tony Maxwell, Department of Biological Chemistry, John Innes Centre
*"Passing one piece of DNA through another: the structure and mechanism of DNA topoisomerases"*

# Publications by Astbury Centre Members 2005


Ackley, M.A., Baldwin, S.A. and King, A.E. (2005). "Adenosine contributes to mu-opioid synaptic inhibition in rat substantia gelatinosa *in vitro*." *Neuroscience Letters* **376**, 102-106.

Ali, M.A., Bhogal, N., Findlay, J.B.C. and Fishwick, C.W.G. (2005). "The first *de novo*-designed antagonists of the human NK2 receptor." *Journal of Medicinal Chemistry* **48**, 5655-5658.

Ashcroft, A.E. (2005). "Recent developments in electrospray ionisation mass spectrometry: noncovalently bound protein complexes." *Natural Product Reports* **22**, 452-464.

Ashcroft, A.E., Lago, H., Macedo, J.M.B., Horn, W.T., Stonehouse, N.J. and Stockley, P.G. (2005). "Engineering thermal stability in RNA phage capsids via disulphide bonds." *Journal of Nanoscience and Nanotechnology* **5**, 2034-2041.

Badjic, J.D., Nelson, A., Cantrill, S.J., Turnbull, W.B. and Stoddart, J.F. (2005). "Multivalency and cooperativity in supramolecular chemistry." *Accounts of Chemical Research* **38**, 723-732.

Bakhetia, M., Charlton, W., Atkinson, H.J. and McPherson, M.J. (2005). "RNA interference of dual oxidase in the plant nematode *Meloidogyne incognita*." *Molecular Plant-Microbe Interactions* **18**, 1099-1106.

Bakhetia, M., Charlton, W.L., Urwin, P.E., McPherson, M.J. and Atkinson, H.J. (2005). "RNA interference and plant parasitic nematodes." *Trends in Plant Science* **10**, 362-367.

Baldwin, S.A., Yao, S.Y.M., Hyde, R.J., Ng, A.M.L., Foppolo, S., Barnes, K., Ritzel, M.W.L., Cass, C.E. and Young, J.D. (2005). "Functional characterization of novel human and mouse equilibrative nucleoside transporters (hENT3 and mENT3) located in intracellular membranes." *Journal of Biological Chemistry* **280**, 15880-15887.

Barnes, K., McIntosh, E., Whetton, A.D., Daley, G.Q., Bentley, J. and Baldwin, S.A. (2005). "Chronic myeloid leukaemia: an investigation into the role of Bcr-Abl-induced abnormalities in glucose transport regulation." *Oncogene* **24**, 3257-3267.

Barratt, E., Bingham, R.J., Warner, D.J., Laughton, C.A., Phillips, S.E.V. and Homans, S.W. (2005). "Van der waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water." *Journal of the American Chemical Society* **127**, 11827-11834.

Barrett, S., Bartlett, S., Bolt, A., Ironmonger, A., Joce, C., Nelson, A. and Woodhall, T. (2005). "Configurational stability of bisindolylmaleimide cyclophanes: From conformers to the first configurationally stable, atropisomeric bisindolylmaleimides." *Chemistry-a European Journal* **11**, 6277-6285.

Bartlett, S., Beddard, G.S., Jackson, R.M., Kayser, V., Kilner, C., Leach, A., Nelson, A., Oledzki, P.R., Parker, P., Reid, G.D. and Warriner, S.L. (2005). "Comparison of the ATP binding sites of protein kinases using conformationally diverse bisindolylmaleimides." *Journal of the American Chemical Society* **127**, 11699-11708.

Beccaluva, L., Bianchini, G., Bonadiman, C., Nowell, G.M., Pearson, D.G., Siena, F. and Wilson, A. (2005). "The metasomatic history of the lithospheric mantle beneath NE Spain." *Geochimica Et Cosmochimica Acta* **69**, A787-A787.

Besong, G.E., Bostock, J.M., Stubbings, W., Chopra, I., Roper, D.I., Lloyd, A.J., Fishwick, C.W.G. and Johnson, A.P. (2005). "A *de novo* designed inhibitor of D-Ala-D-Ala ligase from *E. coli*." *Angewandte Chemie-International Edition* **44**, 6403-6406.

Bicknell, K., Brooks, G., Kaiser, P., Chen, H.Y., Dove, B.K. and Hiscox, J.A. (2005). "Nucleolin is regulated both at the level of transcription and translation." *Biochemical and Biophysical Research Communications* **332**, 817-822.

Blades, M.J., Ison, J.C., Ranasinghe, R. and Findlay, J.B.C. (2005). "Automatic generation and evaluation of sparse protein signatures for families of protein structural domains." *Protein Science* **14**, 13-23.

Blanchard, A., Graham, R.S., Heinrich, M., Pyckhout-Hintzen, W., Richter, D., Likhtman, A.E., McLeish, T.C.B., Read, D.J., Straube, E. and Kohlbrecher, J. (2005). "Small angle neutron scattering observation of chain retraction after a large step deformation." *Physical Review Letters* **95**.

Bourque, E., Kocienski, P.J., Stocks, M. and Yuen, J. (2005). "Synthesis of the C14-C28 fragment of tetronasin." *Synthesis-Stuttgart*, 3219-3224.

Bracey, K., Stevens, L. and Wray, D. (2005). "Comparison of the extent of movement of S4 in voltage-gated potassium and calcium channels." *Biophysical Journal* **88**, 200A-200A.

Bradford, J.R. and Westhead, D.R. (2005). "Improved prediction of protein-protein binding sites using a support vector machines approach." *Bioinformatics* **21**, 1487-1494.

Brockwell, D.J., Beddard, G.S., Paci, E., West, D.K., Olmsted, P.D., Smith, D.A. and Radford, S.E. (2005). "Mechanically unfolding the small, topologically simple protein L." *Biophysical Journal* **89**, 506-519.

Bullock, A.N., Debreczeni, J.E., Fedorov, O.Y., Nelson, A., Marsden, B.D. and Knapp, S. (2005). "Structural basis of inhibitor specificity of the human protooncogene proviral insertion site in Moloney murine leukemia virus (PIM-1) kinase." *Journal of Medicinal Chemistry* **48**, 7604-7614.

Calderwood, M., White, R.E., Griffiths, R.A. and Whitehouse, A. (2005). "Open reading frame 73 is required for herpesvirus saimiri A11-S4 episomal persistence." *Journal of General Virology* **86**, 2703-2708.

Callaghan, A.J., Marcaida, M.J., Stead, J.A., McDowall, K.J., Scott, W.G. and Luisi, B.F. (2005). "Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover." *Nature* **437**, 1187-1191.

Callaghan, A.J., Redko, Y., Murphy, L.M., Grossmann, J.G., Yates, D., Garman, E., Ilag, L.L., Robinson, C.V., Symmons, M.F., McDowall, K.J. and Luisi, B.F. (2005). ""Zn-Link": A metal-sharing interface that organizes the quaternary structure and catalytic site of the endoribonuclease, RNase E." *Biochemistry* **44**, 4667-4675.

Caryl, J.A. and Thomas, C.D. (2005). "Initial events in small staphylococcal plasmid transfer in Thomas, C.M. *et al* (eds) Plasmid Biology 2004: International symposium on molecular biology of bacterial plasmids and other mobile genetic elements." *Plasmid* **53**, 47-48.

Cavalli, A., Vendruscolo, M. and Paci, E. (2005). "Comparison of sequence-based and structure-based energy functions for the reversible folding of a peptide." *Biophysical Journal* **88**, 3158-3166.

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

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

Collis, M.W., Lele, A.K., Mackley, M.R., Graham, R.S., Groves, D.J., Likhtman, A.E., Nicholson, T.M., Harlen, O.G., McLeish, T.C.B., Hutchings, L.R., Fernyhough, C.M. and Young, R.N. (2005). "Constriction flows of monodisperse linear entangled polymers: Multiscale modeling and flow visualization." *Journal of Rheology* **49**, 501-522.

Cooksey, J., Kocienski, P. and Li, Y.F. (2005). "A synthesis of the C1-C9 fragment of ionomycin using cationic eta(3)-allylicmolybdenum and -iron complexes." *Collection of Czechoslovak Chemical Communications* **70**, 1653-1668.

Cordier, C. and Nelson, A. (2005). "Reaching the target: Small molecules aim to probe barrier quality." *Chembiochem* **6**, 1953-1955.

Cox, A., Donnelly, D., Kaur, M., Cheetham, S.C., Cockcroft, V.B. and Findlay, J.B.C. (2005). "MTSEA prevents ligand binding to the human melanocortin-4 receptor by modification of cysteine 130 in transmembrane helix 3." *FEBS Letters* **579**, 285-291.

Crampton, N., Bonass, W.A., Kirkham, J. and Thomson, N.H. (2005). "Formation of aminosilane-functionalized mica for atomic force microscopy imaging of DNA." *Langmuir* **21**, 7884-7891.

Cranz-Mileva, S., Friel, C.T. and Radford, S.E. (2005). "Helix stability and hydrophobicity in the folding mechanism of the bacterial immunity protein Im9." *Protein Engineering Design & Selection* **18**, 41-50.

Curcio, R., Caflisch, A. and Paci, E. (2005). "Change of the unbinding mechanism upon a mutation: A molecular dynamics study of an antibody-hapten complex." *Protein Science* **14**, 2499-2514.

Dennis, C.A., Baron, A., Grossmann, J.G., Mazaleyrat, S., Harris, M. and Jaeger, J. (2005). "Co-translational myristoylation alters the quaternary structure of HIV-1 Nef in solution." *Proteins-Structure Function and Bioinformatics* **60**, 658-669.

Di Cola, E., Waigh, T.A., Trinick, J., Tskhovrebova, L., Houmeida, A., Pyckhout-Hintzen, W. and Dewhurst, C. (2005). "Persistence length of titin from rabbit skeletal muscles measured with scattering and microrheology techniques." *Biophysical Journal* **88**, 4095-4106.

Dugan, G.E. and Hewitt, E.W. (2005). "Functional dissection of the human cytomegalovirus immune evasion protein US6." *Immunology* **116**, 68-68.

Espadaler, J., Romero-Isart, O., Jackson, R.M. and Oliva, B. (2005). "Prediction of protein-protein interactions using distant conservation of sequence patterns and structure relationships." *Bioinformatics* **21**, 3360-3368.

Fielding, A.B., Schonteich, E., Matheson, J., Wilson, G., Yu, X.Z., Hickson, G.R.X., Srivastava, S., Baldwin, S.A., Prekeris, R. and Gould, G.W. (2005). "Rab11-FIP3 and FIP4 interact with Arf6 and the Exocyst to control membrane traffic in cytokinesis." *EMBO Journal* **24**, 3389-3399.

Forman, J.R., Qamar, S., Paci, E., Sandford, R.N. and Clarke, J. (2005). "The remarkable mechanical strength of polycystin-1 supports a direct role in mechanotransduction." *Journal of Molecular Biology* **349**, 861-871.

Fry, E.E., Stuart, D.I. and Rowlands, D.J. (2005). The structure of foot-and-mouth disease virus. *Foot and Mouth Disease Virus*. **288:** 71-101.

Garrow, A.G., Agnew, A. and Westhead, D.R. (2005). "TMB-Hunt: a web server to screen sequence sets for transmembrane beta-barrel proteins." *Nucleic Acids Research* **33**, W188-W192.

Garrow, A.G., Agnew, A. and Westhead, D.R. (2005). "TMB-Hunt: An amino acid composition based method to screen proteomes for beta-barrel transmembrane proteins." *Bmc Bioinformatics* **6**.

Gibson, C.T., Smith, D.A. and Roberts, C.J. (2005). "Calibration of silicon atomic force microscope cantilevers." *Nanotechnology* **16**, 234-238.

Gibson, C.W., Thomson, N.H., Abrams, W.R. and Kirkham, J. (2005). "Nested genes: Biological implications and use of AFM for analysis." *Gene* **350**, 15-23.

Gilbert, R.J.C., Beales, L., Blond, D., Simon, M.N., Lin, B.Y., Chisari, F.V., Stuart, D.I. and Rowlands, D.J. (2005). "Hepatitis B small surface antigen particles are octahedral." *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14783-14788.

Gosal, W.S., Morten, I.J., Hewitt, E.W., Smith, D.A., Thomson, N.H. and Radford, S.E. (2005). "Competing pathways determine fibril morphology in the self-assembly of beta(2)-microglobulin into amyloid." *Journal of Molecular Biology* **351**, 850-864.

Governo, R.J.M., Deuchars, J., Baldwin, S.A. and King, A.E. (2005). "Localization of the NBMPR-sensitive equilibrative nucleoside transporter, ENT1, in the rat dorsal root ganglion and lumbar spinal cord." *Brain Research* **1059**, 129-138.

Grewal, S., Herbert, S.P., Ponnambalam, S. and Walker, J.H. (2005). "Cytosolic phospholipase A(2)-alpha and cyclooxygenase-2 localize to intracellular membranes of EA.hy.926 endothelial cells that are distinct from the endoplasmic reticulum and the Golgi apparatus." *Febs Journal* **272**, 1278-1290.

Griffin, S., Clarke, D., McCormick, C., Rowlands, D. and Harris, M. (2005). "Signal peptide cleavage and internal targeting signals direct the hepatitis C virus p7 protein to distinct intracellular membranes." *Journal of Virology* **79**, 15525-15536.

Griffin, S.D.C., McCormick, C.J., Rowlands, D.J. and Harris, M. (2005). "Sub-cellular localisation of native p7 protein in cells expressing the full-length hepatitis C virus polyprotein." *Journal of Virology* **79**, 15525-15536.

Guy, J.L., Jackson, R.M., Jensen, H.A., Hooper, N.M. and Turner, A.J. (2005). "Identification of critical active-site residues in angiotensin-converting enzyme-2 (ACE2) by site-directed mutagenesis." *FEBS Journal* **272**, 3512-3520.

Guy, J.L., Lambert, D.W., Wamer, F.J., Hooper, N.M. and Turner, A.J. (2005). "Membrane-associated zinc peptidase families: comparing ACE and ACE2." *Biochimica et Biophysica Acta-Proteins and Proteomics* **1751**, 2-8.

Hackenberger, C.P.R., Friel, C.T., Radford, S.E. and Imperiali, B. (2005). "Semisynthesis of a glycosylated lm7 analogue for protein folding studies." *Journal of the American Chemical Society* **127**, 12882-12889.

Hawkins, R.J. and McLeish, T.C.B. (2005). "Dynamic allostery of protein alpha helical coiled-colis." *J. Roy. Soc. Interface* **100**, 3143-3148.

Herbert, S.P., Ponnambalam, S. and Walker, J.H. (2005). "Cytosolic phospholipase A(2)-alpha mediates endothelial cell proliferation and is inactivated by association with the Golgi apparatus." *Molecular Biology of the Cell* **16**, 3800-3809.

Homans, S.W. (2005). "Probing the binding entropy of ligand-protein interactions by NMR." *Chembiochem* **6**, 1585-.

Ironmonger, A., Stockley, P. and Nelson, A. (2005). "Asymmetric double ring-opening of a C-2h-symmetric bis-epoxide: improved enantiomeric excess of the product through enantioselective desymmetrisation and 'proof-reading' steps." *Organic & Biomolecular Chemistry* **3**, 2350-2353.

Jahn, T.R. and Radford, S.E. (2005). "The Yin and Yang of protein folding." *FEBS Journal* **272**, 5962-5970.

Jarowicki, K. and Kocienski, P.J. (2005). "1,2-metallate rearrangements of alpha-alkoxyalkyl cuprates." *Synlett*, 167-169.

Jen, C.H., Michalopoulos, I., Westhead, D.R. and Meyer, P. (2005). "Natural antisense transcripts with coding capacity in Arabidopsis may have a regulatory role that is not linked to double-stranded RNA degradation." *Genome Biology* **6**.

Jeuken, L.J.C., Connell, S.D., Nurnabi, M., O'Reilly, J., Henderson, P.J.F., Evans, S.D. and Bushby, R.J. (2005). "Direct electrochemical interaction between a modified gold electrode and a bacterial membrane extract." *Langmuir* **21**, 1481-1488.

Jones, R.P.O., Durose, L.J., Findlay, J.B.C. and Harrison, M.A. (2005). "Defined sites of interaction between subunits E (Vma4p), C (Vma5p), and G (Vma10p) within the stator structure of the vacuolar $H^+$-ATPase." *Biochemistry* **44**, 3933-3941.

Kalverda, A., Thompson, G.S. and Turnbull, W.B. (2005). The importance of protein structural dynamics and the contribution of NMR spectroscopy. *The encyclopedia of genetics, genomics, proteomics and bioinformatics*. M. J. D. e. al, Wiley Intescience.

Kawakami, M., Byrne, K., Khatri, B.S., McLeish, T.C.B., Radford, S.E. and Smith, D.A. (2005). "Viscoelastic measurements of single molecules on a millisecond time scale by magnetically driven oscillation of an atomic force microscope cantilever." *Langmuir* **21**, 4765-4772.

Kennedy, A., Nelson, A. and Perry, A. (2005). "A general, two-directional approach to aza-C-(1 -> 1)-linked disaccharide mimetics." *Chemical Communications*, 1646-1648.

Kennedy, A., Perry, A. and Nelson, A. (2005). "Methods for the synthesis of polyhydroxylated piperidines by diastereoselective dihydroxylation: Exploitation in the two-directional synthesis of aza-C-linked disaccharide derivatives." *Beilstein J. Org. Chem.* **1**, 2-.

Khan, A., Ashcroft, A.E., Higenell, V., Korchazhkina, O.V. and Exley, C. (2005). "Metals accelerate the formation and direct the structure of amyloid fibrils of NAC." *Journal of Inorganic Biochemistry* **99**, 1920-1927.

Knight, P.J., Thirumurugan, K., Xu, Y.H., Wang, F., Kalverda, A.P., Stafford, W.F., Sellers, J.R. and Peckham, M. (2005). "The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head." *Journal of Biological Chemistry* **280**, 34702-34708.

Kocienski, P.J., Christopher, J.A., Bell, R. and Otto, B. (2005). "Nucleophilic addition of alpha-metallated carbamates to planar chiral cationic eta(3)-allylmolybdenum complexes: A stereochemical study." *Synthesis-Stuttgart*, 75-84.

Lambert, D.W., Yarski, M., Warner, F.J., Thornhill, P., Parkin, E.T., Smith, A.I., Hooper, N.M. and Turner, A.J. (2005). "Tumor necrosis factor-alpha convertase (ADAM17) mediates regulated ectodomain shedding of the severe-acute respiratory syndrome-coronavirus (SARS-CoV) receptor, angiotensin-converting enzyme-2 (ACE2)." *Journal of Biological Chemistry* **280**, 30113-30119.

Langton, K.P., Henderson, P.J.F. and Herbert, R.B. (2005). "Antibiotic resistance: multidrug efflux proteins, a common transport mechanism?" *Natural Product Reports* **22**, 439-451.

Laurie, A.T.R. and Jackson, R.M. (2005). "Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites." *Bioinformatics* **21**, 1908-1916.

Leisle, L., Parkin, E.T., Turner, A.J. and Hooper, N.M. (2005). "Angiotensin-converting enzyme as a GPIase: a critical reevaluation." *Nature Medicine* **11**, 1139-1140.

Li, J.Y., Stevens, L. and Wray, D. (2005). "Molecular regions underlying the activation of low- and high-voltage activating calcium channels." *European Biophysics Journal with Biophysics Letters* **34**, 1017-1029.

Liang, W.J., Wilson, K.J., Xie, H., Knol, J., Suzuki, S., Rutherford, N.G., Henderson, P.J.F. and Jefferson, R.A. (2005). "The *gus*BC genes of *Escherichia coli* encode a glucuronide transport system." *Journal of Bacteriology* **187**, 2377-2385.

Lindorff-Larsen, K., Rogen, P., Paci, E., Vendruscolo, M. and Dobson, C.M. (2005). "Protein folding and the organization of the protein topology universe." *Trends in Biochemical Sciences* **30**, 13-19.

Macdonald, A., Chan, J.K.Y. and Harris, M. (2005). "Perturbation of epidermal growth factor receptor complex formation and Ras signalling in cells harbouring the hepatitis C virus subgenomic replicon." *Journal of General Virology* **86**, 1027-1033.

Macdonald, A., Mazaleyrat, S., McCormick, C., Street, A., Burgoyne, N.J., Jackson, R.M., Cazeaux, V., Shelton, H., Saksela, K. and Harris, M. (2005). "Further studies on hepatitis C virus NS5A-SH3 domain interactions: identification of residues critical for binding and implications for viral RNA replication and modulation of cell signalling." *Journal of General Virology* **86**, 1035-1044.

Malham, R., Johnstone, S., Bingham, R.J., Barratt, E., Phillips, S.E.V., Laughton, C.A. and Homans, S.W. (2005). "Strong solute-solute dispersive interactions in a protein-ligand complex." *Journal of the American Chemical Society* **127**, 17061-17067.

Mann, J., Patrick, C.N., Cragg, M.S., Honeychurch, J., Mann, D.A. and Harris, M. (2005). "Functional analysis of HIV type 1 Nef reveals a role for PAK2 as a regulator of cell phenotype and function in the murine dendritic cell line, DC2.4." *Journal of Immunology* **175**, 6560-6569.

Maxwell, K.L., Wildes, D., Zarrine-Afsar, A., de los Rios, M.A., Brown, A.G., Friel, C.T., Hedberg, L., Horng, J.C., Bona, D., Miller, E.J., Vallee-Belisle, A., Main, E.R.G., Bemporad, F., Qiu, L.L., Teilum, K., Vu, N.D., Edwards, A.M., Ruczinski, I., Poulsen, F.M., Kragelund, B.B., Michnick, S.W., Chiti, F., Bai, Y.W., Hagen, S.J., Serrano, L., Oliveberg, M., Raleigh, D.P., Wittung-Stafshede, P., Radford, S.E., Jackson, S.E., Sosnick, T.R., Marqusee, S., Davidson, A.R. and Plaxco, K.W. (2005). "Protein folding: Defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins." *Protein Science* **14**, 602-616.

McLeish, T.C.B. (2005). "Protein folding in high-dimensional spaces: Hypergutters and the role of nonnative interactions." *Biophysical Journal* **88**, 172-183.

Morten, I.J., Radford, S.E. and Hewitt, E.W. (2005). "Investigation into the role of lysosomes in dialysis related amyloidosis." *Immunology* **116**, 63-63.

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

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

Murphy, J.E., Tedbury, P.R., Homer-Vanniasinkam, S., Walker, J.H. and Ponnambalam, S. (2005). "Biochemistry and cell biology of mammalian scavenger receptors." *Atherosclerosis* **182**, 1-15.

Ng, S.P., Rounsevell, R.W.S., Steward, A., Geierhaas, C.D., Williams, P.M., Paci, E. and Clarke, J. (2005). "Mechanical unfolding of TNfn3: The unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation." *Journal of Molecular Biology* **350**, 776-789.

Paci, E., Greene, L.H., Jones, R.M. and Smith, L.J. (2005). "Characterization of the molten globule state of retinol-binding protein using a molecular dynamics simulation approach." *FEBS Journal* **272**, 4826-4838.

Paci, E., Lindorff-Larsen, K., Dobson, C.M., Karplus, M. and Vendruscolo, M. (2005). "Transition state contact orders correlate with protein folding rates." *Journal of Molecular Biology* **352**, 495-500.

Paci, E. and Vendruscolo, M. (2005). "Detection of non-native hydrophobic interactions in the denatured state of lysozyme by molecular dynamics simulations." *Journal of Physics-Condensed Matter* **17**, S1617-S1626.

Patching, S.G., Baldwin, S.A., Baldwin, A.D., Young, J.D., Gallagher, M.P., Henderson, P.J.F. and Herbert, R.B. (2005). "The nucleoside transport proteins, NupC and NupG, from *Escherichia coli:* specific structural motifs necessary for the binding of ligands." *Organic & Biomolecular Chemistry* **3**, 462-470.

Perrazzollo, C., Wist, J., Poggi, L., Homans, S.W. and Bodenhausen, G. (2005). "Effects of protein-pheromone complexation on correlated chemical shift modulations." *J. Biomol. NMR* **33**, 233-242.

Pinney, J.W., Shirley, M.W., McConkey, G.A. and Westhead, D.R. (2005). "metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*." *Nucleic Acids Research* **33**, 1399-1409.

Plattt, G.W., McParland, V.J., Kalverda, A.P., Homans, S.W. and Radford, S.E. (2005). "Dynamics in the unfolded state of beta(2)-microglobulin studied by NMR." *Journal of Molecular Biology* **346**, 279-294.

Radford, S.E., Gosal, W.S. and Platt, G.W. (2005). "Towards an understanding of the structural molecular mechanism of beta(2)-microglobulin amyloid formation *in vitro*." *Biochimica et Biophysica Acta-Proteins and Proteomics* **1753**, 51-63.

Redzic, Z.B., Biringer, J., Barnes, K., Baldwin, S.A., Al-Sarraf, H., Nicola, P.A., Young, J.D., Cass, C.E., Barrand, M.A. and Hladky, S.B. (2005). "Polarized distribution of nucleoside transporters in rat brain endothelial and choroid plexus epithelial cells." *Journal of Neurochemistry* **94**, 1420-1426.

Robinson, C., Connell, S., Brookes, S.J., Kirkham, J., Shore, R.C. and Smith, D.A.M. (2005). "Surface chemistry of enamel apatite during maturation in relation to pH: implications for protein removal and crystal growth." *Archives of Oral Biology* **50**, 267-270.

Robinson, C., Connell, S.D., Kirkham, J. and Smith, D.A.M. (2005). "Regular nano-domains in apatite crystals from developing enamel: Implications for crystal assembly from subunits and protein binding sites for growth modulation." *Geochimica et Cosmochimica Acta* **69**, A66-A66.

Rodriguez-Mendieta, I.R., Spence, G.R., Gell, C., Radford, S.E. and Smith, D.A. (2005). "Ultraviolet resonance Raman studies reveal the environment of tryptophan and tyrosine residues in the native and partially folded states of the E colicin-binding immunity protein Im7." *Biochemistry* **44**, 3306-3315.

Sadowski, M.I., Parish, J.H. and Westhead, D.R. (2005). "Automated derivation and refinement of sequence length patterns for protein sequences using evolutionary computation." *Biosystems* **81**, 247-254.

Sanderson, J.A., Findlay, J.B.C. and Fishwick, C.W.G. (2005). "Preparation of novel 3H-trifluoromethyldiazirine-based photoactivatable potassium channel antagonists." *Tetrahedron* **61**, 11244-11252.

Sharma-Oates, A., Quirke, P. and Westhead, D.R. (2005). "TmaDB: a repository for tissue microarray data." *Bmc Bioinformatics* **6**.

Smith, K.M., Slugoski, M.D., Loewen, S.K., Ng, A.M.L., Yao, S.Y.M., Chen, X.Z., Karpinski, E., Cass, C.E., Baldwin, S.A. and Young, J.D. (2005). "The broadly selective human $Na^+$/nucleoside cotransporter (hCNT3) exhibits novel cation-coupled nucleoside transport characteristics." *Journal of Biological Chemistry* **280**, 25436-25449.

Smith, P.G., Burchill, S.A., Brooke, D., Coletta, P.L. and Whitehouse, A. (2005). "Efficient infection and persistence of a herpesvirus saimiri-based gene delivery vector into human tumor xenografts and multicellular spheroid cultures." *Cancer Gene Therapy* **12**, 248-256.

Smith, P.G., Oakley, F., Fernandez, M., Mann, D.A., Lemoine, N.R. and Whitehouse, A. (2005). "Herpesvirus saimiri-based vector biodistribution using noninvasive optical imaging." *Gene Therapy* **12**, 1465-1476.

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

Stockley, P.G., Ashcroft, A.E., Francese, S., Thompson, G.S., Ranson, N.A., Smith, D.A.M., Homans, S.W. and Stonehouse, N.J. (2005). "Dissecting the fine details of assembly of a T=3 capsid." *Journal of Theoretical Medicine* **6**, 119-125.

Street, A., Macdonald, A., McCormick, C. and Harris, M. (2005). "Hepatitis C virus NS5A-mediated activation of phosphoinositide 3-kinase results in stabilization of cellular beta-catenin and stimulation of beta-catenin-responsive transcription." *Journal of Virology* **79**, 5006-5016.

Taib, M., Pinney, J.W., Westhead, D.R., McDowall, K.J. and Adams, D.J. (2005). "Differential expression and extent of fungal/plant and fungal/bacterial chitinases of *Aspergillus fumigatus*." *Archives of Microbiology* **184**, 78-81.

Taylor, D.R., Watt, N.T., Perera, W.S.S. and Hooper, N.M. (2005). "Assigning functions to distinct regions of the N-terminus of the prion protein that are involved in its copper-stimulated, clathrin-dependent endocytosis." *Journal of Cell Science* **118**, 5141-5153.

Tezel, A.K., Leal, L.G. and McLeish, T.C.B. (2005). "Rheo-optical evidence of CCR in an entangled four-arm star." *Macromolecules* **38**, 1451-1455.

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

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

Thomson, N.H. (2005). "The substructure of immunoglobulin G resolved to 25 kDa using amplitude modulation AFM in air." *Ultramicroscopy* **105**, 103-110.

Torrance, G.M., Gilbert, D.R., Michalopoulos, I. and Westhead, D.W. (2005). "Protein structure topological comparison, discovery and matching service." *Bioinformatics* **21**, 2537-2538.

Tskhovrebova, L. and Trinick, J. (2005). "Muscle disease: a giant feels the strain." *Nature Medicine* **11**, 478-479.

Tsuboi, T., Ravier, M.A., Xie, H., Ewart, M.A., Gould, G.W., Baldwin, S.A. and Rutter, G.A. (2005). "Mammalian exocyst complex is required for the docking step of insulin vesicle exocytosis." *Journal of Biological Chemistry* **280**, 25565-25570.

Uguru, G.C., Stephens, K.E., Stead, J.A., Towle, J.E., Baumberg, S. and McDowall, K.J. (2005). "Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*." *Molecular Microbiology* **58**, 131-150.

Urquhart, P., Pang, S. and Hooper, N.M. (2005). N-glycans as apical targeting signals in polarized epithelial cells. *Lipids, Rafts and Traffic*. **72:** 39-45.

Vardy, E., Catto, A.J. and Hooper, N.M. (2005). "Proteolytic mechanisms in amyloid-beta metabolism: therapeutic implications for Alzheimer's disease." *Trends in Molecular Medicine* **11**, 464-472.

Visser, F., Baldwin, S.A., Isaac, R.E., Young, J.D. and Cass, C.E. (2005). "Identification and mutational analysis of amino acid residues involved in dipyridamole interactions with human and *Caenorhabditis elegans* equilibrative nucleoside transporters." *Journal of Biological Chemistry* **280**, 11025-11034.

Visser, F., Zhang, J., Raborn, R.T., Baldwin, S.A., Young, J.D. and Cass, C.E. (2005). "Residue 33 of human equilibrative nucleoside transporter 2 is a functionally important component of both the dipyridamole and nucleoside binding sites." *Molecular Pharmacology* **67**, 1291-1298.

Wakenshaw, L., Walters, M.S. and Whitehouse, A. (2005). "The herpesvirus saimiri replication and transcription activator acts synergistically with CCAAT enhancer binding protein alpha to activate the DNA polymerase promoter." *Journal of Virology* **79**, 13548-13560.

Wallace, E.J., Hooper, N.M. and Olmsted, P.D. (2005). "The kinetics of phase separation in asymmetric membranes." *Biophysical Journal* **88**, 4072-4083.

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

Wang, J.J., Saito, Y., Batchelder, D.N., Kirkham, J., Robinson, C. and Smith, D.A. (2005). "Controllable method for the preparation of metalized probes for efficient scanning near-field optical Raman microscopy." *Applied Physics Letters* **86**.

Warner, F.J., Lew, R.A., Smith, I.A., Lambert, D.W., Hooper, N.M. and Turner, A.J. (2005). "Angiotensin-converting enzyme 2 (ACE2), but not ACE, is preferentially localized to the apical surface of polarized kidney cells." *Journal of Biological Chemistry* **280**, 39353-39362.

Weiss, J.L., Evans, N.A., Ahmed, T., Wrigley, J.D.J., Khan, S., Wright, C., Keen, J.N., Holzenburg, A. and Findlay, J.B.C. (2005). "Methionine-rich repeat proteins: a family of membrane- associated proteins which contain unusual repeat regions." *Biochimica et Biophysica Acta-Biomembranes* **1668**, 164-174.

Whitehouse, C., Fang, J.Y., Aggeli, A., Bell, M., Brydson, R., Fishwick, C.W.G., Henderson, J.R., Knobler, C.M., Owens, R.W., Thomson, N.H., Smith, D.A. and Boden, N. (2005). "Adsorption and self-assembly of peptides on mica substrates." *Angewandte Chemie-International Edition* **44**, 1965-1968.

Whyteside, G., Meek, P.J., Ball, S.K., Dixon, N., Finbow, M.E., Kee, T.P., Findlay, J.B.C. and Harrison, M.A. (2005). "Concanamycin and indolyl pentadieneamide inhibitors of the vacuolar $H^+$-ATPase bind with high affinity to the purified proteolipid subunit of the membrane domain." *Biochemistry* **44**, 15024-15031.

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

Williams, G.J., Woodhall, T., Nelson, A. and Berry, A. (2005). "Structure-guided saturation mutagenesis of N-acetylneuraminic acid lyase for the synthesis of sialic acid mimetics." *Protein Engineering Design & Selection* **18**, 239-246.

Wood, J.P.A., Capaldi, S.A., Robinson, M.A., Baron, A.J. and Stonehouse, N.J. (2005). "RNA multimerisation in the DNA packaging motor of bacteriophage phi29." *Journal of Theoretical Medicine* **6**, 127-134.

Woodhall, T., Williams, G., Berry, A. and Nelson, A. (2005). "Creation of a tailored aldolase for the parallel synthesis of sialic acid mimetics." *Angewandte Chemie-International Edition* **44**, 2109-2112.

Woodhall, T., Williams, G., Berry, A. and Nelson, A. (2005). "Synthesis of screening substrates for the directed evolution of sialic acid aldolase: towards tailored enzymes for the preparation of influenza A sialidase inhibitor analogues." *Organic & Biomolecular Chemistry* **3**, 1795-1800.

Woodhall, T., Williams, G.J., Berry, A. and Nelson, A. (2005). "Directed evolution of an aldolase for the parallel synthesis of sialic acid mimetics." *Angewandte Chemie-International Edition* **44**, 2109-2112.

You, J.W., Dove, B.K., Enjuanes, L., DeDiego, M.L., Alvarez, E., Howell, G., Heinen, P., Zambon, M. and Hiscox, J.A. (2005). "Subcellular localization of the severe acute respiratory syndrome coronavirus nucleocapsid protein." *Journal of General Virology* **86**, 3303-3310.

Zhang, J., Smith, K.M., Tackaberry, T., Visser, F., Robins, M.J., Nielsen, L.P.C., Nowak, I., Karpinski, E., Baldwin, S.A., Young, J.D. and Cass, C.E. (2005). "Uridine binding and transportability determinants of human concentrative nucleoside transporters." *Molecular Pharmacology* **68**, 830-839.