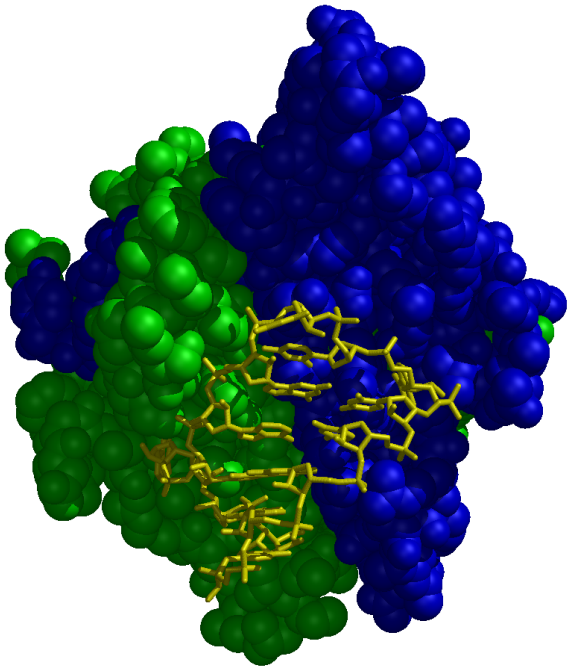
Astbury Centre for Structural Molecular Biology

University of Leeds



Annual Report 2000

Front cover illustration:

RNA aptamer complexed to MS2 bacteriophage coat protein

The image shows an RNA aptamer molecule (gold), corresponding to a stem-loop structure, complexed to two coat protein subunits (blue and green) of the icosahedral bacteriophage MS2. The structure was determined by X-ray crystallography using synchrotron radiation collected at the Daresbury Laboratory (UK) and the ESRF (Grenoble, France). The two subunits shown correspond to a small part of the spherical virus shell that consists of a total of 180 subunits. The view is from the inside of the virus and the RNA is clinging to the inside surface of the shell. The structure is one of a series determined to elucidate rules for protein-RNA recognition in the system.

Convery, M.A., Rowsell, S., Stonehouse, N.J., Ellington, A.D., Hirao, I., Murray, J.B., Peabody, D.S., Phillips, S.E.V. and Stockley, P.G. (1998) "The crystal structure of an RNA aptamer-protein complex at 2.8Å resolution" *Nature Struct. Biol.*, **5**, 133-139.

Rowsell, S., Stonehouse, N.J., Convery, M.A., Adams, C.J., Ellington, A.D., Hirao, I., Peabody, D.S., Stockley, P.G. and Phillips, S.E.V. (1998) "Crystal structures of a series of RNA aptamers complexed to the same protein target" *Nature Struct. Biol.*, **5**, 970-975.

Acknowledgement

The Astbury Centre for Structural Molecular Biology thanks its many sponsors for support of the work, its members for writing these reports and Jenny Walker for her hard work in collating and type-setting this report.

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

Mission Statement

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

Introduction

The Astbury Centre for Structural Molecular Biology (ACSMB) is an interdisciplinary research centre of the University of Leeds. It was founded in 1999 to carry out research in all aspects of structural molecular biology. ACSMB brings together over fifty academic staff from the faculties of Biological Sciences, Chemistry, Physics and Medicine to examine the molecular mechanisms of life itself. It is named after W.T. (Bill) Astbury, a biophysicist who laid many of the foundations of the field during a long research career at the University of Leeds (1928-1961). Astbury originally identified the two major recurring patterns of protein structure (α and β), took the first X-ray fibre diffraction pictures of DNA (in 1938) and is widely credited with the definition of the field of molecular biology.

The ACSMB today continues this tradition and focuses on the determination of the atomic structures of biological macromolecules using all of the current major techniques (X-ray diffraction, nuclear magnetic resonance spectroscopy and electron microscopy). ACSMB also specialises in new biophysical techniques such as measuring the mechanical folding properties of individual protein molecules using atomic force microscopy. These studies are combined with analyses of biological function with the ultimate aim of understanding the molecular basis of life itself. During the year, ACSMB was designated a Marie Curie Research Training Centre by the European Union, allowing young scientists from all over Europe to visit us and receive training in our structural techniques.

Structural molecular biology plays a pivotal role in modern biology, both in the fundamental understanding of living things and in the design of new treatments for disease. It is currently undergoing a revolution in both scale and scope. The publication of the human genome sequence holds out the prospect of using structural molecular biology techniques to determine the exact molecular structure of every protein in our bodies. New drugs are frequently designed by analysing the molecular structures of the proteins they target, so this promises to generate dramatic advances in healthcare. It is notable that the original observation by Astbury of transitions in fibrous proteins between α and β forms remains relevant today since devastating amyloid diseases such as Alzheimer's, BSE and CJD are linked to such transitions in brain proteins.

In the pages that follow, you will find highlights of our work over the last year. The reports have largely been written by our younger researchers, whose enthusiasm for their work augurs well for our future. Noteworthy examples include the first observation of the beginning of the power stroke in myosin, the motor protein that drives muscle, the structure determination of RNA polymerase from Hepatitis C as a target for antiviral drug design,

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.

Simon E.V. Phillips
Director, Astbury Centre for Structural Molecular Biology
Leeds, April 2001
http://www.astbury.leeds.ac.uk

INDEX

Mass spectrometry facility

Alison E. Ashcroft

Overview of facility

The mass spectrometry facility (http://www.astbury.leeds.ac.uk/Facil/mass.htm) has a Platform II single-quadrupole electrospray instrument with on-line HPLC capabilities, a Q-Tof orthogonal acceleration quadrupole - time of flight tandem instrument with nanospray ionisation and a TofSpec matrix assisted laser desorption ionisation mass spectrometer. The facility runs an analytical service as well as being actively involved in several research projects with groups within the Astbury Centre for Structural Molecular Biology and the Faculty of Biological Sciences, and also with external collaborators.

Research projects

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

i. Protein folding

Protein folding is an intriguing area in biochemistry and protein mis-folding is thought to be responsible for several diseases. Working with Prof. S. E. Radford's group, electrospray mass spectrometry is being used to monitor folding profiles and intermediates for the protein β_2-microglobulin (a 99-residue β-sheet protein) using hydrogen-deuterium exchange in order to gain an insight into folding, mis-folding and hence fibril formation.

ii. Protein - ligand non-covalent interactions

In collaboration with the groups of Prof. P. G. Stockley, Prof. S. E. Radford and Dr N. J. Stonehouse, controlled mass spectrometric conditions that preserve non-covalent interactions are being employed to investigate and characterise macromolecular structures. Such studies include protein-peptide, protein-protein, and protein-RNA complexes. The latter are particularly important in virus assembly, an area we are investigating with respect to the MS2 and Qβ systems. By preserving virus assembly and disassembly intermediates in the gas phase we are using mass spectrometry to determine their composition and stoichiometry. Protein-protein macromolecular complexes are important species in fibril formation and are under investigation as an integral part of our β_2-microglobulin folding studies.

iii. Reaction monitoring

Reaction monitoring by mass spectrometry is being employed in areas such as the thrombin activation of Factor XIII A-subunit with Prof. P. J. Grant and Dr R. A. S. Ariens of the Medical School, and for monitoring enzyme-substrate formation with Dr C. W. Wharton, University of Birmingham.

Factor XIII (FXIII) is a transglutaminase that plays an important role in blood coagulation. It is converted by thrombin to its activated form which catalyses the covalent cross-linking of fibrin to increase the stability of the fibrin clot. Deficiency of FXIII results in uncontrolled bleeding, poor wound healing and miscarriage in pregnancy. Figure 1 shows the mass spectrum of FXIII A-subunit (83,192-3Da) which was mass measured to an accuracy of ±0.01% before being treated with thrombin. Both the activated A-subunit (79,256.7 Da) and the activated peptide (3951.2 Da) were identified after proteolytic cleavage, thus substantiating the proposed amino acid sequences.

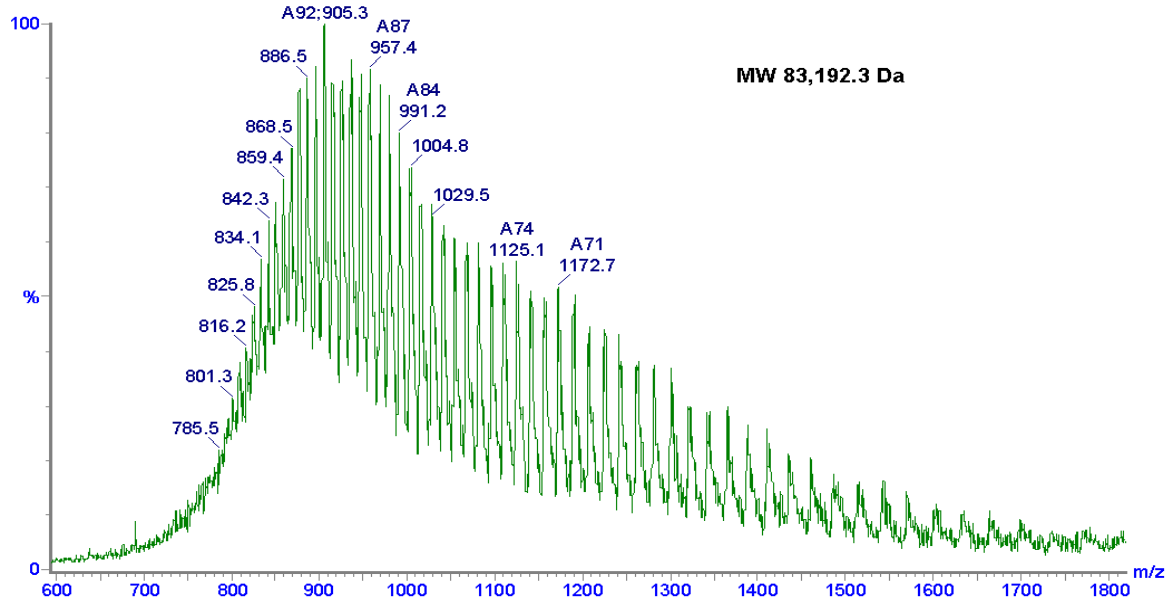


Figure 1: Mass spectrum of Factor XIII A-subunit, molecular weight measured 83,192.3 Da (calculated 83, 192.4 Da).

iv. Structural elucidation

Tandem mass spectrometric sequencing of proteins is useful for the confirmation of amino acid sequence. This technique is being used to characterise membrane proteins, after digestion with a variety of enzymes to ensure a high sequence coverage, in collaboration with Prof. P. J .F. Henderson and Dr R. B. Herbert's group. We have also entered the field of proteomics working with external collaborators (Prof. R. A. Wilson, University of York; Dr R. E. Banks, St. James' Hospital) to generate *de novo* mass maps and sequence tags from 2-D gel digests.

References

A.M. Parrott, H. Lago, C.J. Adams, A.E. Ashcroft, N.J. Stonehouse, P.G. Stockley. 2000 "RNA Aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour". *Nucleic Acids Research*, **28**, 2, 489-497.

S. Jones, J.S. Reader, M. Healy, A.E. Ashcroft, A.P. Kalverda, D.A. Smith, S.E. Radford. 2000 "Unfolding of the Greek Key Protein Y74W Apo-pseudoazurin: Evidence of an Equilibrium Intermediate Species". *Biochemistry*, **39**, 5672-5682.

A.E. Ashcroft, P.J. Grant, R.A.S. Ariens. 2000 "A study of Human Coagulation Factor XIII A-subunit by Electrospray Ionisation Mass Spectrometry." *Rapid Communications in Mass Spectrometry*, **14**, 1607-1611.

Funding

We wish to acknowledge financial support from The Wellcome Trust, BBSRC, Micromass UK Ltd and The University of Leeds.

Molecular studies on active nucleoside transport proteins

Steve Baldwin, Dawn Hadden, Ralph Hyde

Introduction

Nucleoside uptake is essential for the synthesis of nucleotides by salvage pathways in many mammalian cells that lack *de novo* purine biosynthetic pathways, such as bone marrow cells, enterocytes and some brain cells. By regulating the concentration of adenosine available to cell-surface receptors, nucleoside transporters in mammals also influence a wide variety of physiological processes including neurotransmission and cardiovascular activity. Moreover, the transporters represent the route of uptake for a variety of nucleoside analogue drugs used in the treatment of cancer and viral infections. Gaining a better understanding of the molecular mechanism of nucleoside transporters is therefore of both physiological and clinical interest.

A novel concentrative nucleoside transporter from human tissues

Active, sodium-linked uptake systems for nucleosides are found in many mammalian tissues, but are especially abundant in intestine, kidney and liver. These transport activities have been divided into three classes on the basis of their substrate-specificity – the *cit* class are pyrimidine-selective, the *cif* class are purine-selective and the *cib* class exhibit broad substrate selectivity. The proteins that correspond to the first two classes were cloned several years ago by our collaborators Jim Young and Carol Cass at the University of Alberta in Edmonton, and were found to belong to a novel protein family designated the CNT (concentrative nucleoside transporter) family. Representatives of this family have since been identified in many eukaryotes and also in prokaryotes (Fig. 1). However, the identity of the protein responsible for *cib*-type activity remained elusive. We decided to explore the possibility that the *cib* transporter was also a member of the CNT family, by searching the EST and the high-throughput genome sequence databases for possible candidate genes. In collaboration with Jim Young and Carol Cass, this approach led to the successful cloning of the first examples of *cib*-type transporters, from human and mouse, during the summer of 2000. The novel transporters, which we designated hCNT3 and mCNT3 respectively, proved to be members of a new subfamily of the CNTs, most closely related to a transporter from the most primitive living vertebrate, the hagfish (Fig. 1).

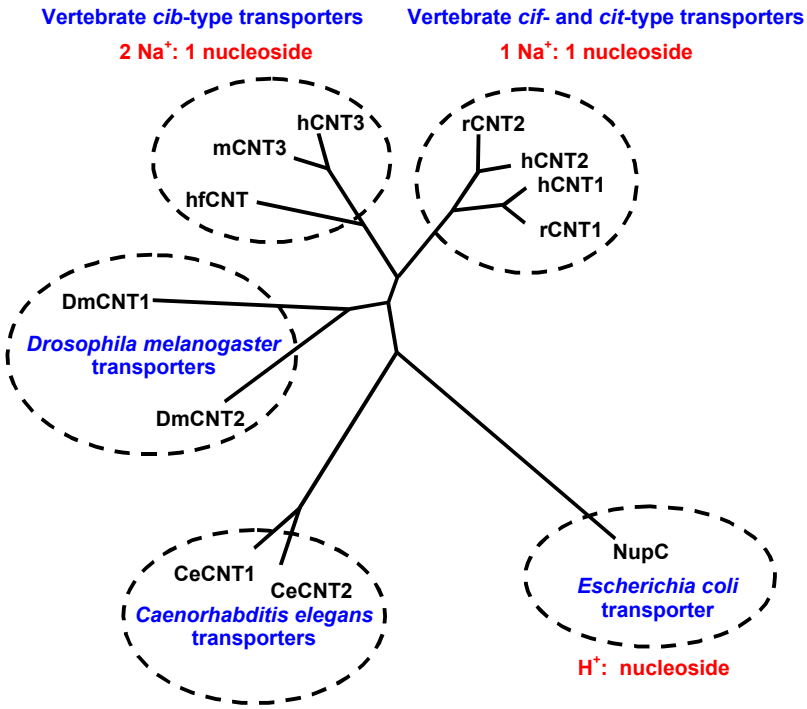


Figure 1. Phylogenetic tree of the CNT family
(r = rat, h = human, m = mouse, Dm = *Drosophila melanogaster*, Ce = *Caenorhabditis elegans*, Hf = hagfish)

Structure/function studies on mammalian CNT family members

Through the use of glycosylation scanning mutagenesis and site-directed antibodies, we have recently shown that mammalian CNTs probably possess a rather unusual topology with 13 transmembrane (TM) segments (Fig. 2), while their bacterial counterparts lack the first three

of these TMs. Interestingly, different CNT subfamilies exhibit different stoichiometries of ion symport: CNT1 and CNT2 transport 1 Na^+ per nucleoside, CNT3 transports 2 Na^+ ions and the bacterial transporter NupC transports H^+ ions rather than sodium (Fig. 1). In collaboration with Jim Young and Carol Cass, we are currently exploiting these different ion dependencies and the differing substrate selectivities of CNT family members to identify regions of the transporters involved in substrate/ion recognition. For example, the kinetic properties of chimaeras and point mutations of the human CNTs expressed in *Xenopus* oocytes have revealed that residues in TM7 and 8 play a critical role in substrate selectivity (Fig. 2).

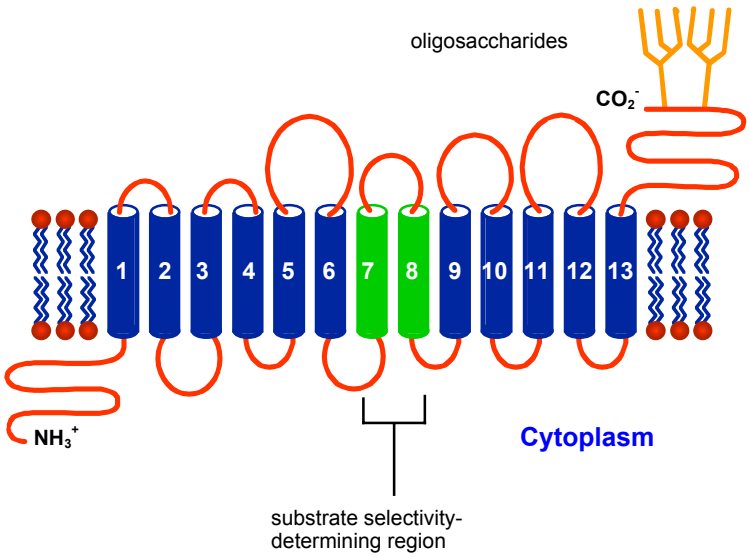


Figure 2. Predicted membrane topology of the human pyrimidine-selective nucleoside/Na^+ transporter hCNT1

Bacterial CNTs as model systems for structure/function investiga-tions

Although site-directed mutagenesis followed by oocyte expression has proven to be a powerful approach for exploring the mechanism of mammalian CNTs, detailed structural studies have been hampered by the difficulty of expressing these proteins at high level. In parallel with studies on the mammalian transporters, we are therefore also working on their bacterial counterparts, which are predicted to be easier to over-express. This is being done in collaboration with Maurice Gallagher at the University of Edinburgh, who cloned the first example of a bacterial CNT, NupC from *Escherichia coli*. Working with Peter Henderson at Leeds, we have been able to over-express NupC at levels up to ~25% of the total membrane protein. Excitingly, such amounts have proven sufficient for examination of the nucleoside binding site by solid-state NMR approaches, which are being undertaken in collaboration with Richard Herbert, Peter Henderson and Adrian Brough at Leeds, together with Tony Watts and Paul Spooner at the National Biological Solid State NMR Facility, University of Oxford. Most recently, we have been able to purify the protein to homogeneity in functional form on a large scale (tens of mg) in the form of a fusion to maltose binding protein. This advance should enable 2-D and 3-D crystallisation trials to begin in the near future.

Collaborators

Peter Henderson, Simon Phillips, Richard Herbert, Adrian Brough; University of Leeds
Maurice Gallagher; University of Edinburgh
Jim Young, Carol Cass; University of Alberta, Edmonton, Canada

References

Loewen, S.K., Ng, A.L., Yao, S.M., Cass, C.E., Baldwin, S.A., & Young, J.D. (1999) Identification of amino acid residues responsible for the pyrimidine and purine nucleoside specificities of human concentrative Na^+ nucleoside cotransporters hCNT1 and hCNT2. *J. Biol. Chem.* **274**, 24475-24484.

Ritzel, M.W.L., Ng, A.M.L., Yao, S.Y.M., Graham, K., Loewen, S.K., Smith, K.M., Ritzel, R.G., Mowles, D.A., Carpenter, P.-A., Chen, X.-Z., Karpinski, E., Hyde, R.J., Baldwin, S.A., Cass, C.E. and Young, J.D. (2001) Molecular identification and characterization of novel human and mouse concentrative Na^+-nucleoside cotransporter proteins (hCNT3 and mCNT3) broadly selective for purine and pyrimidine nucleosides (system *cib*). *J. Biol. Chem.* **276**, 2914-2927.

Funding
We gratefully acknowledge the support of BBSRC, MRC and The Wellcome Trust.

Purine transporters in humans and protozoan parasites

Steve Baldwin, Marie Parker, Ralph Hyde

Introduction

Malaria, resulting from infection by protozoa of the genus *Plasmodium*, causes more than 300 million clinical cases per annum and leads to more than 1.5 million deaths world-wide in tropical and subtropical areas. Unfortunately, resistance to existing antimalarial drugs, including those most recently introduced such as artemisinin (Fig. 1), is rapidly spreading. Therefore the identification of new drug targets is an important goal. One such target might be purine salvage, because the intraerythrocytic stages of *Plasmodium* (Fig. 3) lack the enzymes required for *de novo* purine nucleotide biosynthesis and so are absolutely dependent upon host purines. Many mammalian cells similarly utilise salvage pathways, and take up the necessary purines via nucleoside transport proteins in the plasma membrane. In the hope of exploiting differences in the substrate and inhibitor specificity of human and parasite transporters for therapeutic purposes, a parallel program of research into the structure and function of mammalian and protozoan purine transporters is therefore underway in our laboratory.

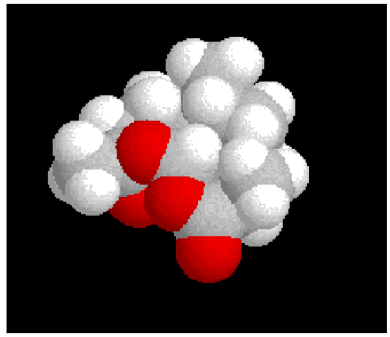


Figure 1. The antimalarial drug artemisinin

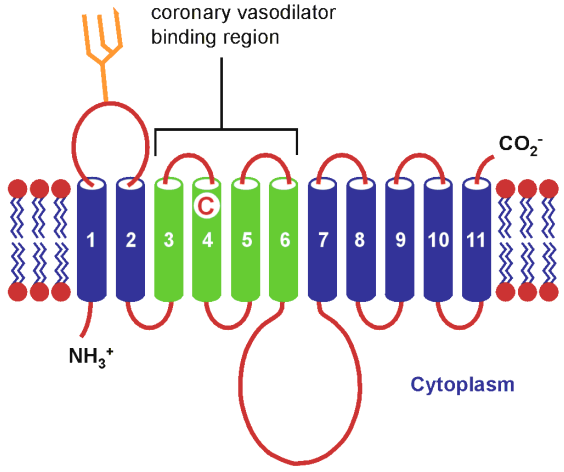


Figure 2. Predicted membrane topology of mammalian members of the ENT family (C = cysteine 140 of rENT2)

Humans possess at least three homologous equilibrative nucleoside transporters

Mammalian cells possess both active, sodium-linked nucleoside transporters (concentrative nucleoside transporters, CNTs) and passive, equilibrative nucleoside transporters. The latter are the most widespread and can be further subdivided into two subclasses on the basis of their sensitivity to the transport inhibitor nitrobenzylthioinosine (NBMPR). Transporters of the *es* (equilibrative sensitive) subclass are inhibited by nanomolar concentrations of NBMPR (K_i 0.1-10 nM). In contrast, transporters of the *ei* (equilibrative insensitive) subclass are relatively insensitive to NBMPR even at micromolar concentrations. In general, transporters of the *es*-type are also potently inhibited by the coronary vasodilators dipyridamole, dilazep and lidoflazine analogues such as draflazine, while *ei*-type transporters are less sensitive to these inhibitors. Our laboratory cloned the first examples of *es*- and *ei*-type transporters, hENT1 and hENT2 respectively, from human placenta in 1997, in collaboration with Jim Young and Carol Cass at the University of Alberta in Edmonton, Canada. Both hENT1 and hENT2 transport a broad range of purine and pyrimidine nucleosides, while hENT2 is also capable of transporting nucleobases. These proteins turned out to be members of a novel family that we have designated the equilibrative nucleoside transporter (ENT) family. Very recently, we have cloned a third member of the family, hENT3, but its substrate selectivity remains to be determined. All three transporters are predicted to possess 11 transmembrane (TM) segments, and we now have direct experimental evidence for much of the topology shown in Fig. 2.

Such evidence has included the results of glycosylation scanning mutagenesis experiments, and accessibility of cysteine residues to membrane-impermeable reagents. For example, with our Canadian collaborators we have recently shown that cysteine 140 in TM4 of the rat *ei*-type transporter rENT2, is responsible for inhibition of transport by extracellular p-chloromercuribenzene sulphonate, and so must be exofacial (Fig. 2). Interestingly, substrate protects rENT2 against inhibition, implying that TM4 contributes to the substrate translocation channel of the transporter. Through examination of the properties of chimeric transporters we had previously established that the TM3-6 region contains residues responsible for interactions of the ENTs with NBMPR and coronary vasodilators (Fig. 2).

Cloning of a novel ENT family member from *Plasmodium falciparum*

In an attempt to identify the transporter(s) responsible for purine uptake by the causative agent of malaria, *Plasmodium falciparum*, we searched the unfinished sequence data arising from the Malaria Genome Sequencing Project for homologues of mammalian active and passive nucleoside transporters. This approach led to identification of fragments encoding a putative ENT homologue, and thence to isolation of a complete coding sequence for this protein, which we termed PfENT1. The 422-residue protein is predicted to adopt an 11-TM topology similar to that of its human counterparts and is expressed during the intraerythrocytic stage of the *Plasmodium* life cycle. Expression of PfENT1 in *Xenopus* oocytes showed that it was indeed a transporter, catalysing the saturable uptake of a wide range of both nucleosides and nucleobases. Interestingly, PfENT1 was found to differ profoundly from its mammalian counterparts in several respects. For example, it was able efficiently to transport 3´-deoxynucleoside analogues such as the antiviral drug AZT, which are poor substrates for human ENTs. Similarly, it was not inhibited by characteristic inhibitors of the mammalian transporters such as NBMPR and coronary vasodilators. These findings suggest that it will be possible to develop inhibitors specific for the parasite transporter for use in anti-malarial chemotherapy. The dependence of *Plasmodium* on purines from the host, and our observation that PfENT1 appears to represent the sole route for uptake of purines by the parasite, suggest that the transporter itself may be a worthwhile chemotherapeutic target. Alternatively, its unique specificity might be exploited to deliver cytotoxic purine analogues selectively into the parasite cytosol while sparing human cells.

Collaborators

Glenn McConkey; University of Leeds
Jim Young, Carol Cass; University of Alberta, Edmonton, Canada

References

Parker, M.D., Hyde, R.J., Yao, S.Y.M., McRobert, L., Cass, C.E., Young, J.D., McConkey, G.A. & Baldwin, S.A. (2000) Identification of a nucleoside/nucleobase transporter from *Plasmodium falciparum*, a novel target for antimalarial chemotherapy. *Biochem. J.* **349**, 67-75.

Yao, S.Y.,M., Sundaram, M., Chomey, E.G., Cass, C.E., Baldwin, S.A. and Young, J.D. (2001) Identification of Cys^{140} in helix 4 as an exofacial cysteine residue within the substrate translocation channel of rat equilibrative nitrobenzylthioinosine (NBMPR)-insensitive nucleoside transporter rENT2. *Biochem. J.* **353**, 387-393.

Funding

We acknowledge the support of BBSRC, MRC and The Wellcome Trust

Biophysical studies of the dynamics and mechanism of Class II fructose 1,6-bisphosphate aldolase

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

Introduction

The importance of protein dynamics in enzyme catalysis and specificity has been previously highlighted in a number of proteins, such as hexokinase and carboxypeptidase A. We know that motions in proteins, ranging from fluctuation of atoms or side chains to closure of entire loops, are a prerequisite for substrates to bind and for reaction products to leave. For example, the role of such flexible loops has been studied in the archetypal $(\alpha/\beta)_8$-barrel protein, triose phosphate isomerase, where a loop, between residues 166-176, in the protein moves from an open position, when substrate is not bound, to a closed position when the substrate binds, stabilizing the reaction intermediate.

Another important member of the $(\alpha/\beta)_8$-barrel family is fructose-1,6-bisphosphate aldolase (FBP-aldolase), which catalyses the reversible aldol condensation of glyceraldehyde 3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP) to form FBP. The FBP-aldolases can be divided into Class I and II enzymes, where the latter is homodimeric and requires one zinc ion per monomer for catalysis.

Motions in the mechanism of Class II FBP-aldolase

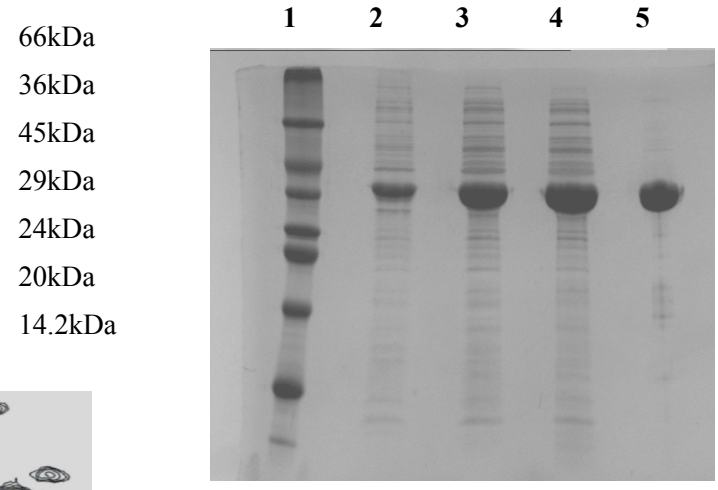
The availability of several crystal structures of the Class II FBP-aldolase has revealed differences in zinc binding in each distinct structure. Comparison of the structures in the absence or presence of the substrate analogue, phosphoglycolohydroxamate, suggests that rotation of His-110 and His-264 upon substrate binding moves the zinc from a buried site to a more exposed one where it binds to the substrate. Other motions in the enzyme on substrate binding include the movement of the loop containing the highly conserved Glu-182, the catalytic base responsible for proton abstraction from DHAP.

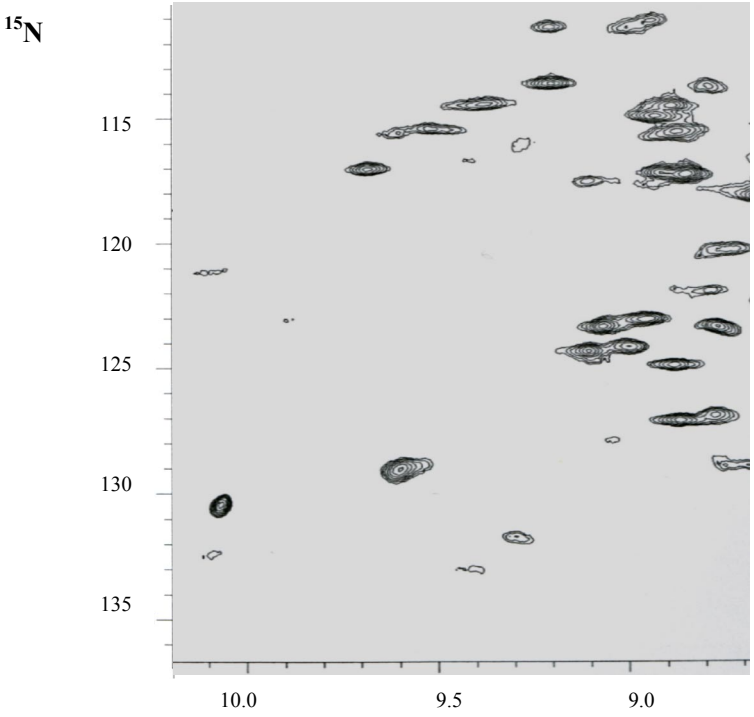
Multi-dimensional NMR spectroscopy

A programme of high-field NMR spectroscopy has been initiated to study the movement of this flexible loop and the role of motion in enzyme catalysis and specificity. The challenging aspect of this project is to work with such a large molecule, a dimer of 78kDa. A number of labelled samples have been prepared containing ^{2}H, ^{15}N and ^{13}C. Over-expression studies show that high levels of FBP-aldolase over-expression, similar to that obtained in rich media, can be achieved in the labelling media with D_2O. Both [^{1}H, ^{15}N]-HSQC and [^{1}H, ^{15}N]-TROSY spectra have been collected and we are now using TROSY type triple-resonance NMR experiments to determine the backbone chemical shifts of the residues of interest in the active site and in the flexible loop.

SDS-PAGE analysis of the various stages of the purification of [^{15}N, ^{13}C, ^{2}H]-FBP-aldolase.

Lane 1: Molecular weight markers.
Lane 2: Whole cell extract.
Lane 3: After the French Press.
Lane 4: 40-80% ammonium sulphate precipitate.
Lane 5: DE-52 anion exchange eluate.
The protein of interest is shown with an arrow.





Portion of the 600 MHz [^{1}H, ^{15}N]-TROSY spectrum of [^{15}N, ^{13}C, ^{2}H]-FBP-aldolase.

References

Zgiby, S. M., Thomson, G.J., Qamar, S. & Berry, A. (2000) Exploring substrate binding and discrimination in fructose-1,6-bisphosphate and tagatose-1,6-bisphosphate aldolases. *Eur. J. Biochem*, **267**, 1858-1868.

Hall, D. R., Leonard, G.A., Reed, C.D., Watt, C.I., Berry, A. & Hunter, W.N. (1999) The crystal structure of *Escherichia coli* Class II fructose-1,6-bisphosphate aldolase in complex with phosphoglycolohydroxamate reveals details of mechanism and specificity. *J. Mol. Biol.*, **287**, 383-394.

Plater, A. R., Zgiby, S.M., Thomson, G.J., Qamar, S., Wharton, C.W. & Berry, A. (1999) Conserved residues in the mechanism of the *E. coli* Class II FBP-aldolase. *J. Mol. Biol.*, **285**, 843-855.

Funding: BBSRC and the Wellcome Trust are gratefully acknowledged.

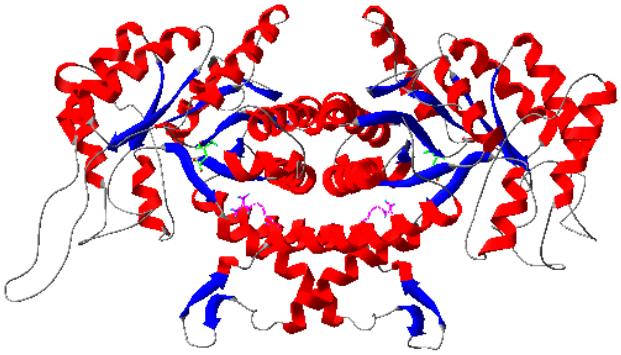
Protein engineering of aldolase and the sialic acid synthesising enzymes

Alan Berry, Gavin Williams, Jijun Hao, Silvie Domann and Venty Suryanti

Introduction

Organic chemists are increasingly turning to the use of enzymes in organic synthesis, and as such the aldolases, which catalyse carbon-carbon formation, represent an important class of biocatalyst. However, the natural substrate specificity of these enzymes limits their wider use in synthesis. Thus, we aim to use protein engineering to alter the substrate specificity of these enzymes.

Mechanistic studies of fructose-1,6-bisphosphate aldolase (FBP-aldolase)



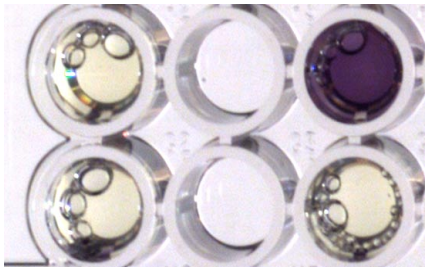
The crystal structure of FBP-aldolase. The enzyme belongs to the TIM barrel family of protein fold.

This important glycolytic enzyme provides a framework for our protein engineering experiments. FBP-aldolase catalyses the reversible condensation of dihydroxyacetone phosphate (DHAP) and glyceraldehyde phosphate (G3P) to form fructose bisphosphate (FBP). Aldolases are of two types, the Class I aldolases which use an active site lysine and the zinc dependent Class II aldolases. The Class II enzyme from *E.coli* has been cloned, overexpressed and a crystal structure with bound substrate analogue determined. During the past year a comprehensive program of site-directed mutagenesis coupled with detailed steady-state kinetics, inhibition studies, chemical modification and Fourier-transform infra red spectroscopy has delineated the role of residues involved in substrate binding and catalysis. We have shown aspartate-109 to be responsible for proton donation to the substrate G3P, and that serine-61 and asparagine-35 play roles in substrate binding.

Rational redesign of aldolase and directed evolution

Our knowledge gathered by mechanistic and structural studies provided us with an excellent opportunity to use protein engineering to alter the specificity of the aldolase. To complement these studies we have cloned and overexpressed the *E.coli* Class II tagatose-1,6-bisphosphate (TBP) aldolase and a crystal structure determined in collaboration with Dr WN Hunter (Dundee). TBP is a diastereoisomer of FBP. The TBP and FBP-aldolase are both $(\alpha/\beta)_8$-barrels and share the same catalytic mechanism, yet each enzyme is highly specific for its natural substrate. Rational attempts at converting the FBP-aldolase to a TBP-aldolase with a total of nine mutations failed to switch substrate specificity, highlighting the complexity of substrate binding and catalysis. We are now using directed evolution to convert the TBP-aldolase activity into FBP-aldolase. DNA shuffling is being used to randomly mutate the TBP-aldolase gene, a colorimetric screen has been used to identify those mutants with increased FBP-aldolase activity. Active mutants are purified and DNA sequencing of the genes reveals the mutations present. The mechanism of the evolved aldolase(s) is determined with the help of detailed steady state kinetics and the wild-type crystal structure. We are currently modifying this methodology to evolve other properties such as thermal stability or

alternative mechanisms. These experiments may give insights into how these enzymes evolved in nature. We are also evolving new catalysts for the synthesis of sialic acid analogues. The enzymes sialic acid aldolase and sialic acid synthase have been cloned from a variety of microbial sources. These are being purified and characterised for evaluation as biocatalysts. The substrate specificity of these enzymes will be altered by rational engineering and directed evolution.



Detection of aldolase activity from colonies grown in microtitre plates. The purple colour indicates an active aldolase is present. This is used to identify active mutants from libraries created by DNA shuffling.

Complementary to this work is the development of new chromogenic and fluorogenic substrates/sensors for the aldolase reaction. Collaboration with the chemistry department at the University of Leeds is leading towards the synthesis and use of novel compounds for directed evolution. In addition, the synthesis of substrate analogues and enzyme inhibitors as well as novel substrates is providing additional mechanistic information.

Collaborators
Dr. Adam Nelson, University of Leeds, Chemistry
Dr. Bill Hunter, University of Dundee

References
Zgiby, S. M., Thomson, G.J., Qamar, S. & Berry, A. (2000) Exploring substrate binding and ciscrimination in fructose-1,6-bisphosphate and tagatose-1,6-bisphosphate aldolases. *Eur. J. Biochem,* **267**, 1858-1868.

Hall, D. R., Leonard, G.A., Reed, C.D., Watt, C.I., Berry, A. & Hunter, W.N. (1999) The crystal structure of Escherichia coli Class II fructose-1,6-bisphosphate aldolase in complex with phosphoglycolohydroxamate reveals details of mechanism and specificity. *J. Mol. Biol.,* **287**, 383-394.

Plater, A. R., Zgiby, S.M., Thomson, G.J., Qamar, S., Wharton, C.W. & Berry, A. (1999) Conserved residues in the mechanism of the E. coli Class II FBP-aldolase. *J. Mol. Biol.,* **285**, 843-855.

Funding
Financial support from BBSRC and The Wellcome Trust is gratefully acknowledged.

Structure-function analysis of G protein- coupled receptors

Nirmala Bhogal, Alan Cox, Mark Wigglesworth, Mohammed Sheikh Taj and John Findlay.

Introduction

Our work centres around Family A (rhodopsin-like) and Family D (STE2p yeast pheromone) peptide and Family B (secretin-like) glycoprotein G protein-coupled receptors (GPCRs)**.** GPCRs possess 7 transmembrane α-helices (TMs), connected by hydrophilic loop regions and have their N- and C-termini exposed to the extracellular and intracellular environments, respectively. It is this serpentine arrangement which facilitates recognition of an extracellular biomolecule and initiation of intracellular signalling events (Fig. 1). The involvement of dysfunctional GPCRs in disease pathology emphasises the need for a greater understanding of the mechanisms of ligand interaction, receptor activation and receptor structure.

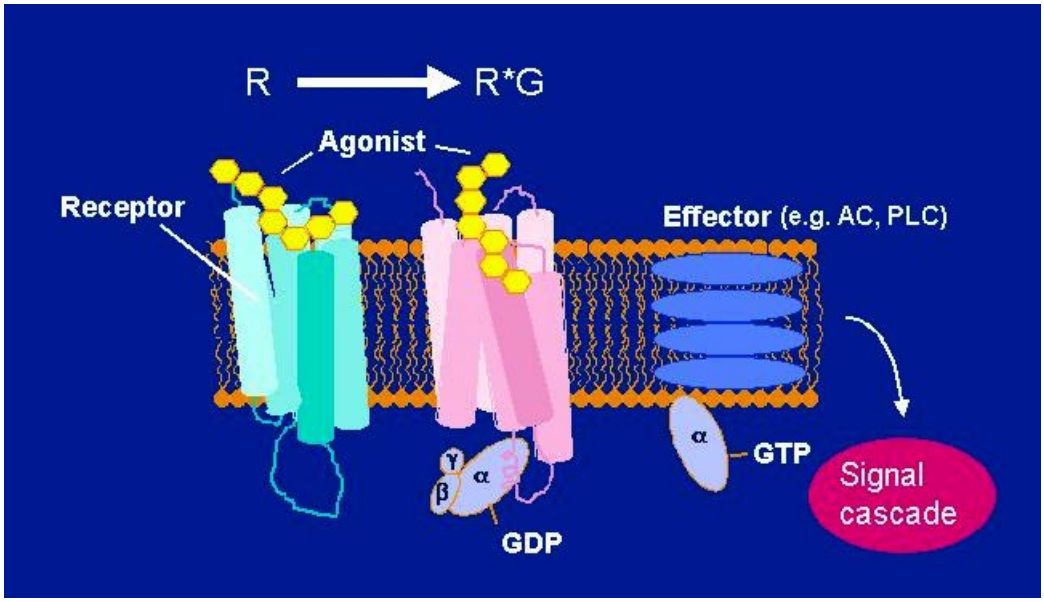


Figure 1. GPCR signalling. Upon peptide agonist binding, both the receptor and agonist undergo conformational changes culminating in an increase of R* sites which associate with an intracellular G protein. Downstream signalling involves the exchange of GDP for GTP by the α-subunit which initiates the signal cascade via one or more effectors.

Agonist binding sites

Physiological ligands for GPCRs are agonists - i.e. they bind to and activate GPCRs. Agonists vary in size from small ions and chemicals, to peptides and glycoproteins. Generally, small agonists for Family A GPCRs interact with receptor side chains in the extracellular third of the TM segments. However, a growing body of evidence suggests that, with the exception of tripeptides, neuropeptides make contact sites in the extracellular *and* intramembranous segments of their receptors via an induced fit mode of interaction, thus complicating the interpretation of mutagenesis data: Residues which appear to be involved in agonist binding may do so by maintaining the environment around the binding crevice rather than participate in direct ligand interaction. We have developed an approach whereby disulphide bond formation between Cys-substituted receptors and probe-labelled Cys-peptide analogues can be used to distinguish conformational determinants from receptor-peptide contacts. Our data support the concept of transmembrane and extracellular contacts for peptide agonists.

The binding environment: Consequences of receptor activation

Environment-sensitive thiol agents are being used to investigate the accessibility of cysteine residues in wild-type and mutant GPCRs. Positively charged reagents seem to inhibit radioligand binding to two neuropeptide receptors which implies the presence of at least one cysteine residue in, or near, a charged ligand binding site. We are currently systematically replacing cysteine residues within these receptors with alanine or serine in order to pinpoint modified cysteines.

GPCRs resonate between at least one inactive form (R) and one active form (R*): Only R* sites can associate with *and* activate G-proteins even though the R form predominates. Hence, in the absence of an agonist, a small population of R* gives rise to basal activity and only agonist binding shifts the equilibrium towards R* and increased G-protein activity. Natural and engineered receptor-activating mutants are thought to push the equilibrium towards R*. We are looking at changes in accessibility of native and engineered TM cysteines in Family A, B and D receptors in response to the introduction of activating mutations to assess whether receptors from different families are activated by similar shifts/rotations in the TM segments and explore the possible existence of multiple R* forms induced by mutations in distinct receptor regions.

Structure-activity relationships (SAR)

The poor bioavailibility of peptide ligands has stimulated interest in small organic compounds which can mimic or inhibit binding of the receptor-specific biomolecule. Consistent with studies on other receptors, the binding sites for the inhibitors (antagonists) only partially overlap with that of peptide agonists. Non-peptidic antagonists appear to make fewer contacts than the peptides and, unlike the peptide agonists, appear to bind predominantly to the TM segments. Hence, SAR is a suitable approach to defining antagonist-receptor contacts. Libraries of receptor-specific antagonist analogues are being used in conjunction with site-directed receptor mutagenesis to establish specific contacts between piperidine antagonists with TM6 of a Family A peptide receptor.

Structural analysis

Engineered fragments of GPCRs are known to spontaneously fold and associate. This has been exploited to examine the orientation of putative protein-buried receptor sites and establish possible helix-helix contacts. Expression of complementary pairs of Cys-substituted receptor fragments and assessment of the ligand binding, signal transduction and catalysed disulphide bond formation has revealed that (in Family A receptors) ligand binding residues in TMs 3 and 7 and TMs 5 and 6 lie in close proximity.

Direct structural work requires the isolation of approx. 10 mg of >90% homogenous, functional receptors. We are able to express a peptide receptor at 13 pmol/mg crude sf9 insect cell membranes and recover 30% of these receptors in a functional soluble form. Our next goal is to scale-up expression for NMR studies using spin-labelled ligands to probe their binding environments.

Collaborators: Heptagen; SmithKline Beecham; BASF; Novo Nordisk.

Funding: We acknowledge the support of BBSRC.

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

Caitriona Dennis, Gemma Dixon, Nicky Kingswell, Matt Bentham, Joachim Jäger and Mark Harris

Nef is a 205 amino acid regulatory protein that plays a critical role in viral pathogenesis. The protein is both phosphorylated and N-terminally myristoylated. The structure of part of the Nef molecule has been solved - a C-terminal core domain of Nef (amino acids 71-206) lacking a disordered loop (residues 156-173) (see Figure 1 below). As yet there is no information about the structure of the full-length myristoylated form. Our preliminary Circular Dichroism data suggest that myristoylated Nef (expressed in the eucaryotic baculovirus system as an N-terminal GST fusion) has a higher alpha-helical content than a non-myristoylated derivative (see Figure 2), suggesting that myristoylation plays a structural role. Myristoylation is an eucaryotic specific co-translational modification that is catalysed by a single-subunit ribosomal associated enzyme - N-myristoyltransferase (NMT). In order to generate large amounts of myristoylated Nef for structural studies we are funded by the BBSRC to co-express Nef in *E.coli* with human NMT. In a related BBSRC funded project we are analysing the mechanism of interaction between Nef and cellular membranes - by analogy to other cellular myristoyl-proteins this is likely to involve not only myristate but additional interactions, eg electrostatic interactions between basic amino acids and acid phospholipids.

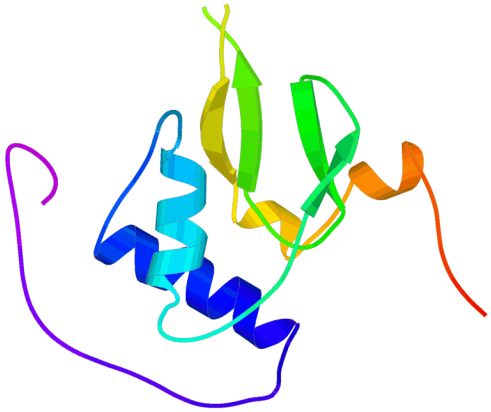


Figure 1: Structure of the conserved C-terminal Nef 'core' fragment.

One of the functions of Nef is to down-modulate the cell surface glycoprotein CD4 (which also functions as the viral receptor). This involves a direct physical interaction between myristoylated Nef and the cytoplasmic domain of CD4. We are extending previous work in the laboratory by undertaking a detailed analysis of the interaction between Nef and CD4. This work is funded by the EU and the MRC (the latter as part of the new Co-operative Group entitled 'Enabling Technologies in the Life Sciences'). The long term goal of this project is to develop assays for the interaction that can be used for drug screening.

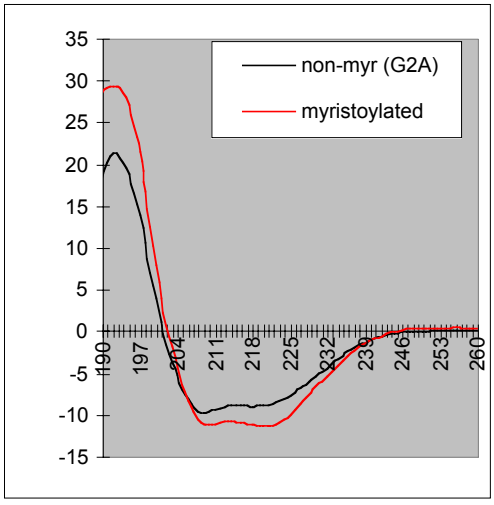


Figure 2: Circular Dichroism analysis of myristoylated and non-myristoylated NefGST fusion protein.

Collaborators
EU Framework Five 'Targeting Nef'

Funding
We acknowledge the support of BBSRC and the MRC

Structural and functional studies on Hepatitis C virus non-structural proteins.

Mustapha Aoubala, Katherine Crowder, John Holt, Andrew Macdonald, Mick Miller and Mark Harris

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

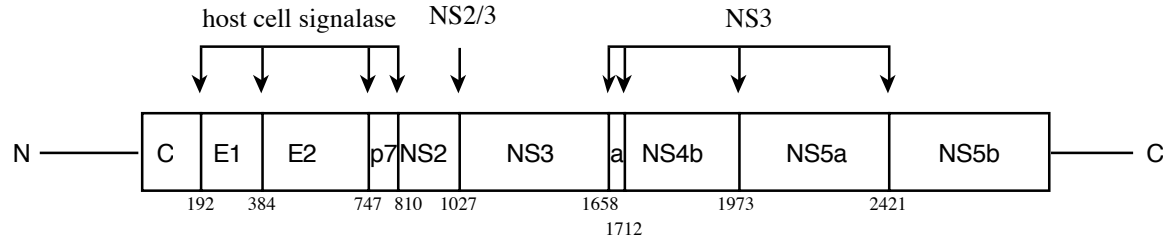


Fig. 1: Polyprotein cleavage strategy of HCV

The non-structural protein NS3 has three enzymatic activities: a proteinase and a helicase/NTPase. Recently, catalytically inactive NS3 fragments containing an arginine-rich motif have been reported to interact with, and inhibit, the catalytic subunit of cAMP-dependent protein kinase (PKA C-subunit). We have used recombinant baculoviruses to express full-length, catalytically-active NS3 and shown that it is also able to inhibit PKA C-subunit *in vitro*. However both mutational analysis and experiments in which a constant ATP concentration was maintained by the addition of an ATP regeneration system, demonstrated that the ability to inhibit PKA was due to ATPase activity. We are currently pursuing the functional consequences of NS3 expression in mammalian cell lines to determine whether ATPase activity might play a role in pathogenesis of this virus.

NS5A contains a poly-proline motif reported to interact with the SH3 domain of the intracelllular adaptor protein Grb2. Further to this study we have shown that NS5A is also able to interact with the SH3 domains derived from members of the Src family of protein tyrosine kinases. We are currently using phage display, surface plasmon resonance and *in vitro* binding assays to more precisely understand these interactions. Additionally we are expressing NS5A transiently in mammalian cells to identify the functional consequences of these interactions for signal transduction pathways within the cell.

Collaborators:
Roger Clegg, Hannah Research Institute; Derek Mann, University of Southampton;
Kalle Saksela, University of Tampere, Finland.

Funding: This work is funded by the BBSRC and the MRC.

Investigation of the ligand binding site in the glucuronide-H^+ symporter, *Gus*B, in *Escherichia coli*, using chemical and solid state NMR spectroscopy approaches

Anton Rajakarier, Rietie Venter, Adrian Brough, Fiona Melvin, Richard Herbert and Peter Henderson

Introduction

Alkyl or aryl β-D-glucuronides (**Figure 1**) are important natural compounds, often being the conjugates by which xenobiotics or drugs are excreted from the human body. The drug connection provides a pharmaceutical interest for the synthesis of these compounds.

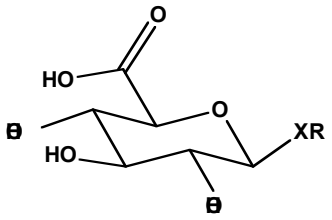


Figure 1: β-D-glucuronide where X= O, S or NH; R= alkyl or aryl group

The source of all glucuronides in the mammals is UDP-glucuronide, which possesses the capacity to conjugate with a wide range of molecular groups namely R-OH, R-COOH, R-NH_2 or R-SH. Glucuronides are produced in the liver cells and are then excreted into the bile and removed by the urinary tract or large intestine. *Escherichia coli* in the human gut uses the glucuronides as a carbon source, assimilating them through the cell membrane by agency of membrane protein *Gus*B, one of a family of important bacterial transport proteins.

We have cloned, sequenced and over expressed the *Gus*B protein to high levels. The *gus*B gene encodes 457 amino acid residues with a theoretical molecular weight of 49,982Da. Based on the amino acid sequence, *Gus*B is deduced to consist of 12 trans membrane α-helices with *N*- and *C*- terminals on the cytoplasmic side of the membrane. Each helix contains approximately 20-25 amino acid residues. *Gus*B is a member of the galactose-pentose-hexuronide (GPH) transport family and shares significant homology with other members within this family.

Aim of the research

We are applying novel cross polarisation-magic angle spinning solid state NMR (CP-MAS NMR) methods to determine the 3-D structure of the binding site of *Gus*B in its native membrane state. This technique is based on the interaction of stable isotope labels (^{13}C, ^{15}N, ^{19}F) introduced at specific sites in the ligand (glucuronide) and the protein (*Gus*B) by means of short-range (5-15Å) polarisation transfer. By making quantitative distance measurements a picture of the ligand-binding site can be constructed.

Chemical synthesis of substrates for CP-MAS NMR and bioassay

Starting with [1-^{13}C]-labelled glucose (which is readily available commercially) we have developed a high yielding route to prepare methyl [1-^{13}C]-β-D-glucuronide over five steps. The overall reaction yield is 70% (**Figure 2**).

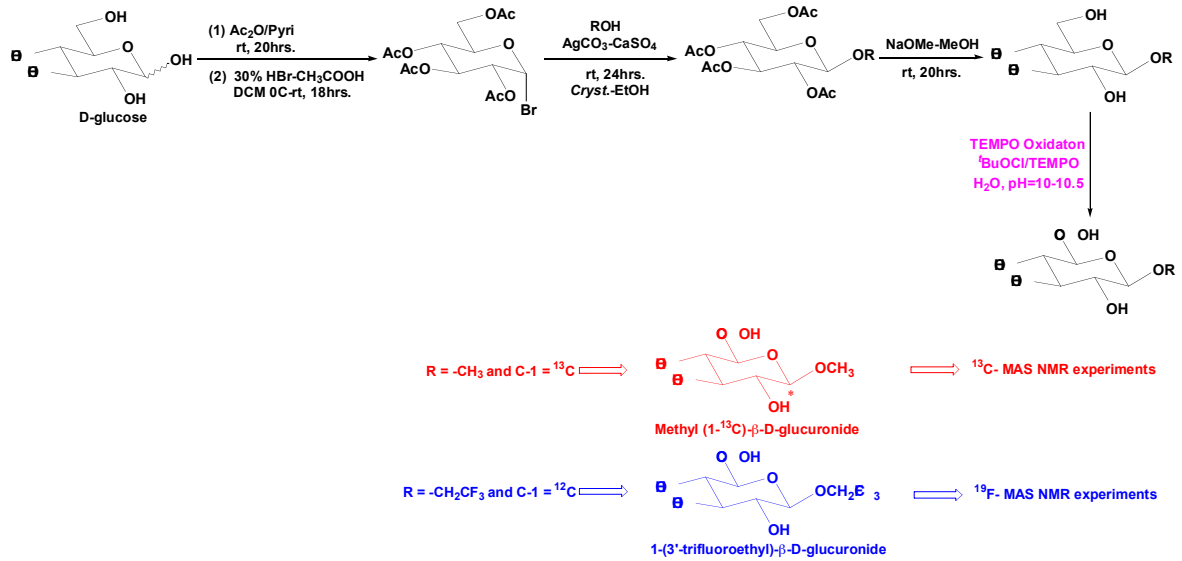


Figure 2: Total synthesis of methyl [1-^{13}C]-β-D-glucuronide and 1-(3'-trifluoroethyl)-β-D-glucuronide for carbon and fluorine solid state NMR experiments, respectively

The pivotal step in the synthesis is the oxidation of methyl [1-^{13}C]-β-D-glucoside to methyl [1-^{13}C]-β-D-glucuronide by tetramethylpiperidinyl-1-oxy (TEMPO) and t-butyl hypochlorite. This oxidation method is readily applicable to the preparation of a wide range of glucuronides from other glucoside derivatives. All the products obtained were spectroscopically characterised.

Similarly, various glucuronide derivatives such as fluoro-deoxy, epimers, deoxy and carbonyl protected glucuronides were chemically synthesised and tested in biological assays to establish the importance of the corresponding hydroxyl groups in the substrate recognition by *Gus*B.

CP-MAS solid-state NMR experiments

CP-MAS solid-state NMR was carried out at 75.46MHz for ^{13}C nuclei (300.13MHz for ^{1}H) using a Bruker (Billerica, MA) MSL-300 spectrometer in Leeds.

The proton-decoupled ^{13}C CP-MAS NMR spectra that were recorded on inner membrane vesicles containing 0.15μmol (8mg) of *Gus*BH protein is shown in **Figure 3A**. The spectrum is dominated by the signal from the natural abundance ^{13}C nuclei of the alkyl groups (δ_C 0– 50ppm) from the phospholipids, but the region expected to accommodate the resonance from the substrate (δ_C 103ppm) is free of any natural abundance contributions.

After methyl [1-^{13}C]-β-D-glucuronide (6mM) was added to the membranes, the CP signal arising from a motion-restricted part of this substrate emerges in the anticipated spectral region as shown in **Figure 3B**. It could be shown that the signal at δ_C 103ppm arises from the specific binding of methyl [1-^{13}C]-β-D-glucuronide to the protein, as opposed to non-specific binding to the lipid bi-layer, by competing out the methyl glucuronide with another substrate. p-Nitrophenyl β-D-glucuronide is such a substrate that binds to *Gus*BH with a similar affinity to methyl β-D-glucuronide. The methyl β-D-glucuronide signal disappeared when unlabelled p-nitrophenyl β-D-glucuronide was added to the *Gus*BH containing membranes in ten-fold excess (60mM) over the amount of methyl β-D-glucuronide added (**Figure 3C**).

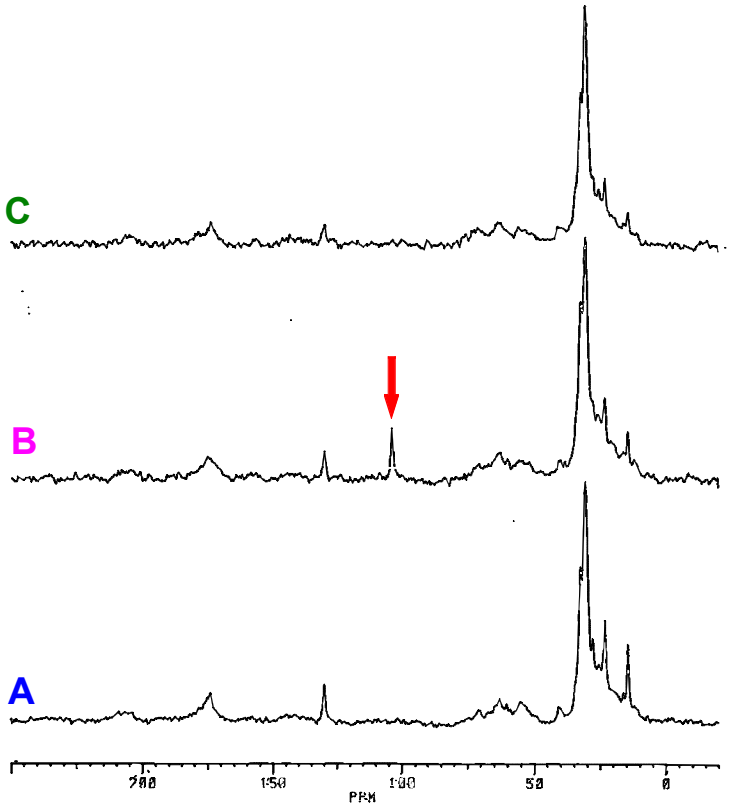


Figure 3: Proton decoupled ^{13}C CP-MAS solid-state NMR spectra from inner membranes containing 0.15µmol of *Gus*B(His)$_6$: without any substrate (**A**); with methyl [1-^{13}C]-β-D-glucuronide (6mM) added (**B**) showing intensity from bound substrate (at δ_C 103ppm); and when the glucuronide was competed out by treatment with unlabelled p-nitrophenyl β-D-glucuronide (60mM) (**C**).

As further proof that the substrate signal arises from methyl [1-^{13}C]-β-D-glucuronide motionally restricted in the binding site of *Gus*B, negative control membranes were prepared. The plasmid for over-expression of *Gus*BH (pWJL24H) is derived from the high copy number plasmid pTTQ18. A suitable negative control would therefore be membranes from NM554 (pTTQ18) prepared with similar conditions of growth and induction as for *Gus*BH-containing membranes. The CP-MAS NMR spectrum recorded from inner membrane vesicles without any *Gus*B, but with the same amount of total protein (35mg) as the *Gus*BH containing membranes, is shown in **Figure 4A.** No substrate signal was observed at δ_C 103ppm when labelled methyl glucuronide was added to the negative control membranes (**Figure 4B**), showing that the signal observed in the *Gus*BH containing membranes arises from substrate bound to *Gus*B.

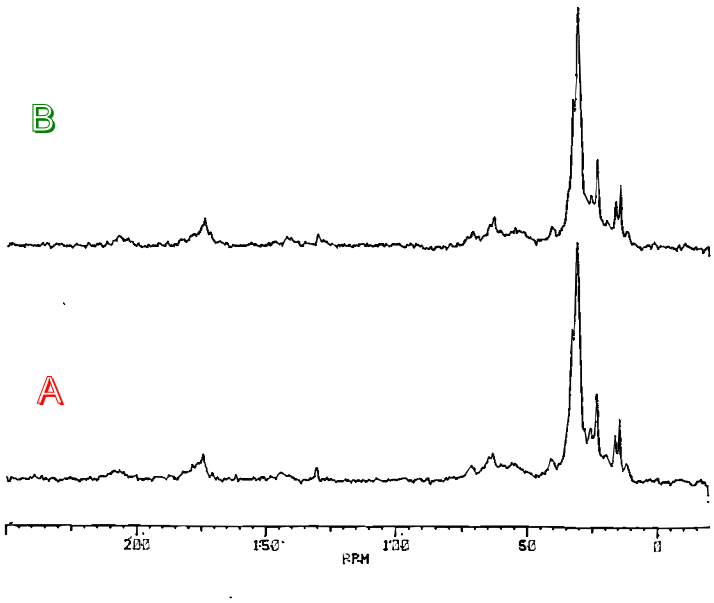


Figure 4: Proton decoupled ^{13}C CP-MAS solid-state NMR spectra from inner membranes without *Gus*B expressed. Membranes without any substrate (**A**) and after addition of methyl [1-^{13}C]-β-D-glucuronide (6mM) (**B**).

The initial results showed that methyl β-D-glucuronide can be observed specifically bound to *Gus*B and it is possible to displace the peak by competitive displacement. Furthermore, variable contact time (CT) and de-phasing delayed cross polarisation (DDCP) NMR techniques were used to confirm the peak observed at δ_C 103ppm was due to methyl [1-^{13}C]-β-D-glucuronide bound to the *Gus*B (experimental details are not included).

Biological assays

The structural requirements for glucuronide binding to the *Gus*B protein were explored by testing various synthetic glucuronide derivatives as inhibitors for acetamidophenyl [1-^{14}C]-glucuronide uptake in *E. coli* cells. It was found that all the substituents needed to be equatorial for efficient inhibition. Further and importantly, the C-3-OH is essential and is a H-bond donor; also the C-4-OH is important and acts as an acceptor; and the C-2-OH is not important. Naturally the carboxyl group (C-6) is essential.

Conclusions

- CP-MAS solid state NMR experiments have been carried out in order to identify the specific binding of the methyl [1-^{13}C]-β-D-glucuronide to *Gus*B.
- A variety of methyl β-D-glucuronide derivatives have been chemically synthesised and the importance of the hydroxyl groups, carboxyl group and stereochemistry for binding to *Gus*B have been established in biological assays.

Funding

We acknowledge financial support from BBSRC.

NMR facility

Arnout Kalverda, Steve Homans

Equipment

The NMR facility has benefited from the installation of two new NMR spectrometers, a 500 MHz and a 600 MHz VARIAN INOVA, in February 2000. The spectrometers are housed in a purpose built extension to the Astbury building. The facility is now equipped with one 600 MHz and two 500 MHz NMR spectrometers with identical setup. Each spectrometer has four frequency channels while a fifth channel (dedicated to deuterium) is usually installed on the 600 MHz system but can be moved to the 500 MHz systems when needed. This arrangement allows the study of singly, doubly and triply labelled proteins and nucleic acids.

Towards solution NMR of larger systems

With the development of relaxation optimised spectroscopy (TROSY) there has been a considerable extension of the size limits for solution NMR. The improvement in sensitivity and resolution that the TROSY method provides is particularly relevant for triple resonance experiments as they include extended periods where transverse ^{15}N magnetisation is present. For instance, three dimensional triple resonance experiments like HNCA and HN(CO)CA which are essential for obtaining sequential assignments in bigger proteins show a large benefit from using TROSY methodology. A full suite of TROSY based triple resonance experiments has been implemented on the 600 MHz system. Their performance has been tested with a sample of ^{2}H, ^{13}C, ^{15}N ubiquitin dissolved in 50% glycerol and a ^{2}H, ^{13}C, ^{15}N labelled 78 kDa dimeric aldolase. The results show that even at 600 MHz the TROSY based triple resonance experiments show a considerable improvement for proteins larger than 20 kDa. The experiments on aldolase show the feasibility of obtaining sequential assignments for multimeric system in the 40 – 80 kDa range.

NMR of protein folding and misfolding

NMR is one of the few techniques that can provide residue specific information about partially folded states of proteins. Since many of these partially folded states are not directly accessible because they have very broad NMR spectra or are only transiently populated, more indirect methods like hydrogen exchange and denaturant titrations need to be used. We are working on a variety of systems (D1 domain of amine oxidase, Colicin Immunity proteins, pseudoazurin, β_2-microgobulin) for which the protein folding is also being studied with other spectroscopic and kinetic techniques. As a step towards investigating partially folded states of the amyloidogenic β_2-microgobulin, ^{15}N assignments of the native state and ^{1}H, ^{15}N assignments of the urea denatured state have been completed using ^{15}N correlated 3D spectroscopy.

Publications

Jones, S., Reader, J.S., Healy, M., Capaldi, A.P., Ashcroft, A.E., Kalverda, A.P., Smith, D.A. & Radford, S.E. (2000) Partially unfolded species populated during equilibrium

denaturation of the beta-sheet protein Y74W apo-pseudoazurin. *Biochemistry*, **39**, 5672-5682.

McParland, V.J., Kad, N.M., Kalverda, A.P., Brown, A., Kirwin-Jones, P., Hunter, M.G., Sunde, M. & Radford, S.E. (2000) Partially unfolded states of beta(2)-microglobulin and amyloid formation in vitro. *Biochemistry*, **39**, 8735-8746.

Morris, L.A., Van den Bosch, J.J.K., Versluis, K., Thompson, G.S. & Jaspars, M. (2000) Structure determination and MSn analysis of two new lissoclinamides isolated from the Indo-pacific oscidian *Lissoclinum patella*: NOE restrained molecular dynamics confirms the absolute sterechemistry derived by degradative methods. *Tetrahedron* **56**, 8345-8353.

Thompson, G.S. Leung, Y.C., Ferguson, S.J., Radford, S.E. & Redfield, C. (2000) The structure and dynamics in solution of Cu(I) pseudoazurin from *Paracoccus pantotrophus. Protein Science* **9**, 846-858.

Thompson, G.S., Shimizu, H., Homans, S.W. & Donohue-Rolfe, A. (2000) Localisation of the binding site for the oligosacharide mooiety of GB(3) on verotoxin 1 using NMR residual dipolar coupling measurements. *Biochemistry* **39**, 13153-13156.

Funding
We acknowledge the support of the Wellcome Trust.

NMR studies of the structure of the Verotoxin B subunit from *E. coli* O157 with its receptor Gb_3.

Gary Thompson, Hiroki Shimizu, Steve Homans.

Introduction

Since the discovery of *E.coli* 0157 in Argentina in 1977 this strain of *E. coli* has been a present and increasing cause of death and disease in both the first and the third Worlds. During the largest outbreak in the United Kingdom at Wishaw in Scotland 180 people were infected and 17 people died. Furthermore infection by *E. coli* O157 is currently both a major cause of fatal diarrhoea in the third world and the main cause of renal failure in infants.

In common with a number of other pathogens (e.g. cholera) *E. coli* O157 produces an AB5 class toxin. These bear their name because they contain a single catalytic A unit that prevents protein sythesis in the cell by N ribosylation of the ribosome. However, translocation of the A subunit into the cell and the subsequent disease is not possible without the B5 subunit (VTB). The B5 subunit has been shown to bind to a cell surface receptor globotriceramide (Gb3; Galα1-4Galβ1-4Glc-Cer) via the Gal-Gal-Glc oligosaccharide Gb3-OS. This ligand shows weak binding in the millimolar range ($K_d = 1x10^{-3}$) and X-Ray crystallography showed three equally occupied binding sites (1-3)[i]. The uncertainty in the major the binding site of Gb3-OS with VTB obviously makes the design of inhibitors for the binding of Gb3 a major problem.

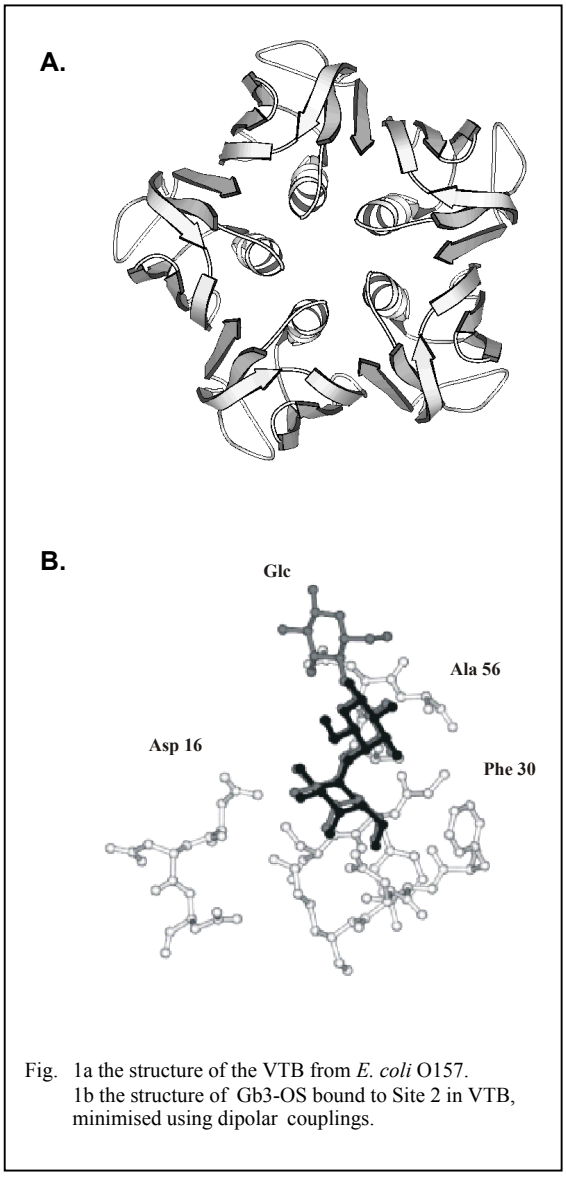


Fig. 1a the structure of the VTB from *E. coli* O157.
1b the structure of Gb3-OS bound to Site 2 in VTB, minimised using dipolar couplings.

Use of dipolar couplings.

Dipolar couplings are a new and extremely powerful method for obtaining orientational information about bonds from NMR spectra of molecules in partially aligned media such as liquid crystals. Using a DHPC:DMPC medium and a fully ^{13}C labelled (0.70mM) sample of Gb3-OS we have measured ^{1}H-^{13}C transferred dipolar coupling over a range of protein:sugar concentrations giving very accurate dipolar couplings for the bound state of Gb3OH with VTB. As VTB is an axially symmetric molecule the principle axis of alignment in the partially aligned medium must be along the axis of symmetry[ii]. Using this fact we have been able to minimise the structure of Gb3-OS in the three sites in the VTB crystal structure against the observed dipolar couplings. The exceedingly good match obtained for the ligand in site 2 (E_{dip} =0.00 kcal $mole^{-1}$, RMSD for heavy atoms 0.01 ± 0.01Å) clearly confirm it as the major (>99%) binding site.

Current work on the structure of the VTB:GB_3-OS complex

Having shown unequivocally that site 2 is the major binding site for Gb3-OS to VTB in solution a question remains open: is it possible to fully calculate the structure of Gb3-OS bound to VTB using residual dipolar couplings and the small number of available transferred nOes (TRNOEs)[iii] currently available? This is especially complicated in the case of VTB as not only are there only three TRNOES available (which are all to one alanine methyl group and so are not independent) but also the axial symmetry of the alignment of Gb3-OS greatly decreases the resolving power of the dipolar couplings.

Using XPLOR we have performed docking calculations using a new protocol based on rigid body torsion angle dynamics. With one independent TRNOE and 10 dipolar coupling restraints to dock VTB to Gb3-OS the protocol is able to find the correct binding site. Unfortunately, the orientation of Gb3-OS in the final stage of the protocol cannot be achieved, as the Gb3-OS VTB binding site is quite 'flat'. However we have been able to show that with 1 extra independent restraint a well resolved structure with a good match between the mean structure and the crystal structure (RMSD 1.72 Å) can be calculated.

References

Thompson G. S., Shimizu H., Homans S.W., & Art Donohue-Rolfe. (2000) Localization of the Binding Site for the Oligosaccharide Moiety of Gb3 on Verotoxin 1 Using NMR Residual Dipolar Coupling Measurements, *Biochemistry*; **39** 13153-13156.

Collaborators

Art Donohue-Rolfe. Dept. of Comparative Medicine, Tufts University School of Veterinary Medicine, 200 Westboro Road, N. Grafton, MA 01536, USA.

Funding

GST is supported by the BBSRC.

---

[i] H. Ling, A Boodhoo, B. Hazes, M. D. Cummings, G. D. Armstrong, J. L. Brunton, R. J. Read. Biochemistry, 37 1777-1788 (1998).

[ii] Al Hashimi, H. M. Bolon, P. J, Prestegard, J. Magn. Reson. 142, 153-158 (2000).

[iii] H. Shimizu, A. D. Donohue Rolfe, S. W. Homans. J. Am. Chem. Soc. 121, 5815- 5816 (1999).

The structure of *Mycobacterium tuberculosis* lipoarabinomannan (LAM)

Achim Treumann, Steve Homans

Introduction

Tuberculosis is an increasing and major world-wide problem, especially in Africa where the spread is facilitated by AIDS. It is estimated that nearly 1 billion people will become infected, 200 million will become sick, and 70 million will die world-wide between now and 2020. In 1998, approximately 8 million cases and two million deaths were attributed to TB; 100,000 of those two million deaths occurred among children.

Mycobacterium tuberculosis, the bacterium that causes this disease, is protected from the host by a unique cell wall. One of the main components of this cell wall is lipoarabinomannan (LAM), a glycoconjugate that is involved in many interactions with the host's immune system and therefore likely to contribute to the pathogen's survival.

The structure of LAM has been investigated extensively in the early '90s by Delphi Chatterjee and coworkers in Pat Brennan's laboratories. The backbone of LAM consists of a linear α1-6 mannosyl polymer that is linked to a phosphatidylinositol anchor. Most mannosyl residues are substituted on their C-2 with another single mannose. This mannosyl backbone carries an unknown number of branched arabinosyl side chains, which consist of the rare α-D-arabinofuranose. The arabinosyl branches are based on a linear α1-5 linked backbone. Branching residues carry an additional α1-3 linked arabinofuranose. At their non-reducing ends, the α-arabinofuranans are capped with a β1-2 linked arabinofuranose. Slow-growing mycobacteria like *Mycobacterium tuberculosis* attach one or more α-mannoses to these β-arabinose termini. In total, one molecule of LAM contains about 30 mannose units and 50 arabinofuranoses.

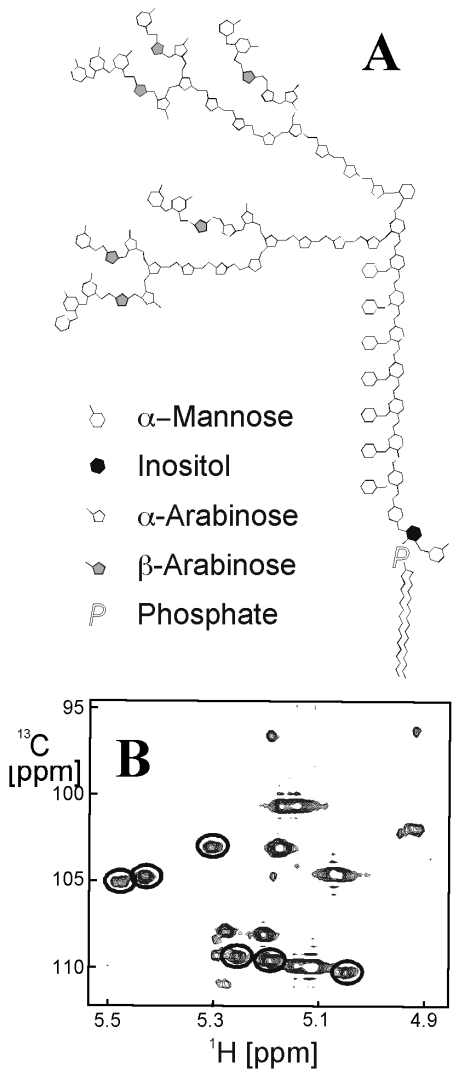


Figure 1. (**A**) Current (incomplete) model of the structure of (LAM). (**B**) Contour diagram of the anomeric region of a constant-time HSQC NMR spectrum of 13C-labelled LAM. Major peaks that cannot be assigned to the model in (**A**) and are therefore indicative of new structural features are marked with circles.

NMR studies of LAM

We are using high field NMR studies on LAM that has been purified from bacteria that were grown in ^{13}C enriched media to obtain a complete structure of this clinically important molecule. The stable isotope labelling enables us to perform highly resolved heteronuclear experiments on very small amounts of sample that would otherwise be impossible. In two-

dimensional NMR spectra, we can identify peaks that indicate new structural elements of LAM, features that will be important to explain the interactions of LAM with patients' immune systems.

Enzymatic cleavage of LAM

As it is impossible to make a complete structural assignment based on studies of the complete molecule, we are breaking it down into smaller fragments, using highly specific glycosidases. These fragments are separated chromatographically and can subsequently be analysed using NMR spectroscopy.

Tandem mass spectrometric studies of LAM fragments

In addition to NMR spectroscopy, we are also using tandem mass spectrometry to obtain more information about the fragments generated by digestion of LAM with endo- and exoglycosidases.

Collaborators

Delphi Chatterjee, Jordi Torrelles, Michael McNeil, Patrick Brennan, University of Colorado, Ft. Collins.

Funding

We acknowledge the support of the Wellcome Trust

Efficient resonance assignment and global fold determination of backbone labelled proteins

Alexander Giesen, Steve Homans

Introduction

As the result of a massive effort several genomes have been sequenced and soon the sequence of all human genes will be known. In most cases it will be impossible to devise the structure of newly discovered proteins from their sequence alone. However, the number of experimentally determined structures is small compared to the number of known protein sequences. This is due to the fact that determining a structure by x-ray crystallography or NMR spectroscopy is a long and tedious process. A rapid determination of the protein fold, even at low resolution, will be useful for characterizing the structure of proteins and finally their function.

Multidimensional NMR

With the development of multidimensional triple resonance NMR methods, the structure of proteins of significant size and spectral complexity can now be determined. A fundamental limitation on the size of proteins whose resonances can be assigned by triple resonance methods that transfer magnetisation through alpha-carbon nuclei (e.g. HNCA, HN(CO)CA) is the rapid transverse relaxation of these nuclei. The relaxation time of these carbons can be increased by substitution of deuterons for protons. A second problem is the homonuclear C-alpha-C-beta coupling. For fully ^{13}C-labelled proteins one works either with limited resolution in the ^{13}C dimension to prevent the resolution of this coupling, or a constant time (CT) period is applied during ^{13}C frequency labelling. The first approach frequently does not allow unambiguous identification of the connectivities, whereas a CT experiment increases the time period during which transverse C-alpha magnetisation is present to approximately 27 msec. This means that even for deuterated proteins a significant loss of magnetisation will occur. In principle, selective decoupling of C-beta overcomes this limitations. However this approach introduces other undesirable complications. Firstly, it is in general impossible to selectively decouple serine C-beta atoms, and secondly, the composite pulse train introduces an undesirable Bloch-Siegert shift on C-alpha resonances.

New isotopic enrichment strategies

An alternative approach, which does not suffer from these disadvantages, involves a protein that is isotopically enriched exclusively in the backbone N, C-alpha and C-carbonyl atoms and deuterated in the C-alpha position. This permits optimal resolution and sensitivity to be obtained in triple resonance experiments, since the undesirable C-alpha-C-beta scalar coupling is absent. We synthesised backbone ^{13}C, ^{15}N, ^{2}H(50%)-labelled amino acids by chemical means starting from ^{13}C, ^{15}N-labelled glycine. Partially backbone labelled human ubiquitin was then obtained from these selectively labelled amino acids. Backbone labelled proteins can be used for acquisition of triple resonance correlation spectra with high resolution in the C-alpha dimension and optimal sensitivity. Recently devised NMR methods of measuring residual dipolar couplings in liquid crystalline media provide a new rapid route to the determination of protein fold. For proteins that are isotopically enriched exclusively in their backbone nuclei, residual dipolar couplings can be measured effectively with HSQC-based techniques. Optimal sensitivity is maintained due to the absence of the constant-time building blocks that are required for fully labelled proteins.

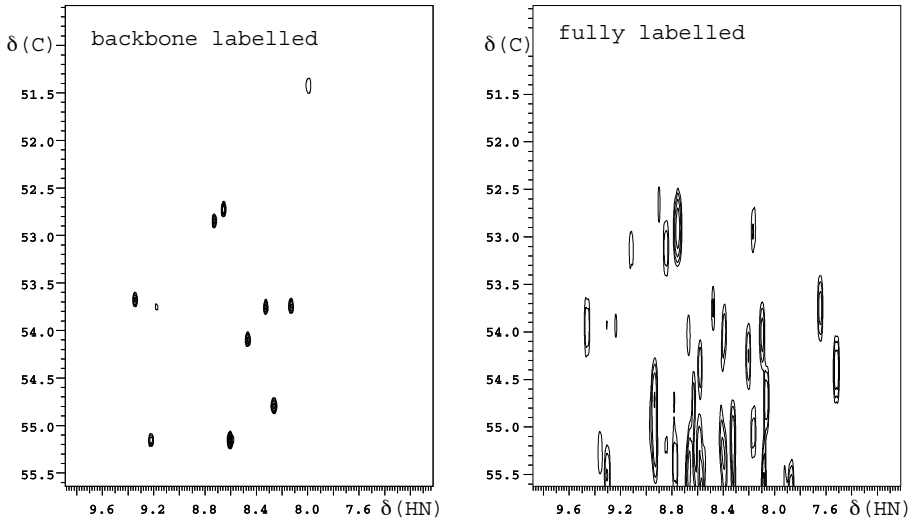


Figure 1 : 2D-HNCA of VFL backbone labelled ubiquitin (left) and fully labelled ubiquitin (right). Note the high resolution obtained for the backbone labelled sample.

Collaborators

Jonathan M. Brown, Martek Biosciences Corporation, Columbia, Maryland USA

References

Giesen, A.W., Bae, L.C., Barrett, C.L., Chyba, J.A., Chaykovsky, M.M., Cheng, M.-C., Murray, J.H., Oliver, E.J., Brown, J.M., Dahlquist, F.W. & Homans, S.W. (2001) Measurement of One-Bond $^{1}H\alpha$-$^{13}C\alpha$ Couplings in Backbone-Labelled Proteins. *J. Biomol. NMR* in press

Funding

We acknowledge the support of the Wellcome Trust

Subtle changes in the fingers domain of DNA polymerase ß influence nucleotide discrimination and fidelity

Danyal Conn, Andrew Capaldi, Joanne Sweasy[#], and Joachim Jäger

[#]Departments of Therapuetic Radiology and Genetics, Yale University School of Medicine.

DNA polymerase ß

DNA polymerase ß (pol ß) is one of the smallest eukaryotic DNA polymerases and offers a simple system to examine the correlation between polymerase structure and fidelity of DNA synthesis. Polymerase ß plays a major role in base excision repair (BER) as it removes 5'-deoxyribose phosphate and fills short lesions in DNA (short patch synthesis). Furthermore, *pol* ß is implicated in nucleotide excision repair. There is also evidence to suggest a meiotic function for *pol* ß. During meiosis, *pol* ß localizes to the synaptonemal complexes in the first prophase of meiosis.

Selection of mutator mutants

In this study, the M282L variant of *pol* ß (M282Lß) was identified using an *in vivo* genetic screen based on the replacement of the DNA *pol* A gene in *E.coli* with the mammalian *pol* ß gene. Mutator phenotypes were isolated using a Trp^+ reversion assay from a library of random *pol* ß mutants. *Pol* ß variants that produce Trp^+ revertants at significant frequencies over WT *pol*ß were isolated.

M282Lß shows increased mutagenesis in both *in vivo* and *in vitro* assays. DNA from the *pol* ß mutant library was transfected into the *E.coli* SC18-12 strain for further kinetic and biophysical studies.

Protein stability and polymerase fidelity

To confirm the intrinsic mutator activity of M282Lß, we used a forward mutation assay to determine the *in vitro* spontaneous mutation frequencies of wild-type and mutant *pol* ß proteins. In this assay, WTß or M282Lß fills a 203 nucleotide gap of the HSV-*tk* gene at the ATP binding site. Errors committed by the polymerase during DNA synthesis can inactivate the HSV-*tk* gene.

Transient-state kinetic methods were used to investigate the mechanism of the intrinsic mutator activity of M282Lß. Results show an 11-fold decrease in dNTP substrate discrimination at the level of ground-state binding. In contrast, during the protein conformational change (structural step) and/or phosphodiester bond formation (chemical step), the nucleotide discrimination is improved by 3-fold. However, there is an overall reduction in fidelity for M282Lß.

To investigate the possibility of any global structural changes in M282Lß, we used circular dichroism spectroscopy to determine the helical content as a function of temperature and urea. The T_m values are 42°C for WTß and 48°C for M282Lß (Figure 1). This suggests that the amino acid alteration stabilized the *pol* ß mutant protein. To verify the observed protein stability for M282Lß, circular dichroism spectra were collected in the presence of a chemical denaturant, urea, which indicate that the reduced fidelity may possibly be related to a distinct increase in protein stability.

Structural studies

X-ray crystallography was utilized to determine the structural basis of the decreased DNA synthesis fidelity. The crystal structure was solved and refined against data extending to 2.3Å resolution. Met282, which does not directly interact with the DNA template or the incoming dNTP substrate, is located on helix N of *pol* ß. In the mutant structure (Figure 2), nearly all of the structural changes map to site 282 and the surrounding region.

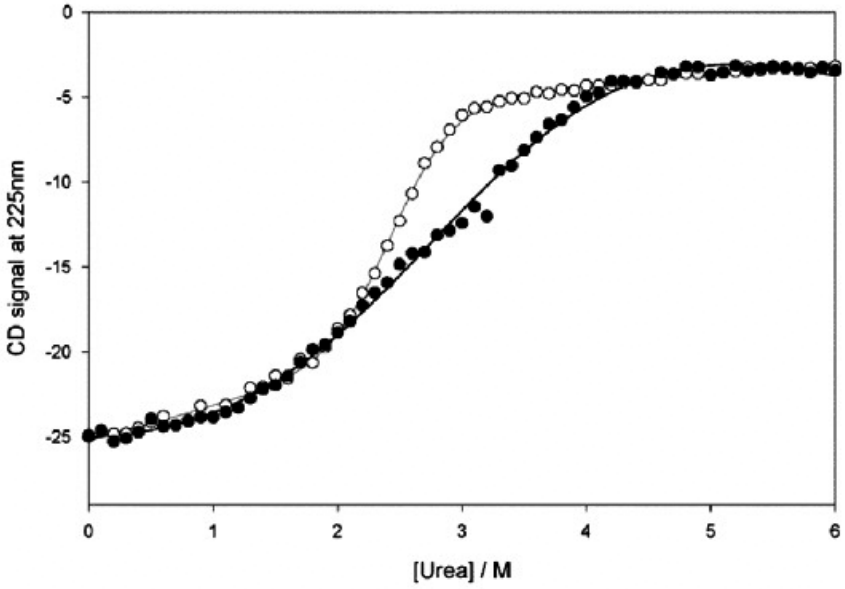


Fig 1

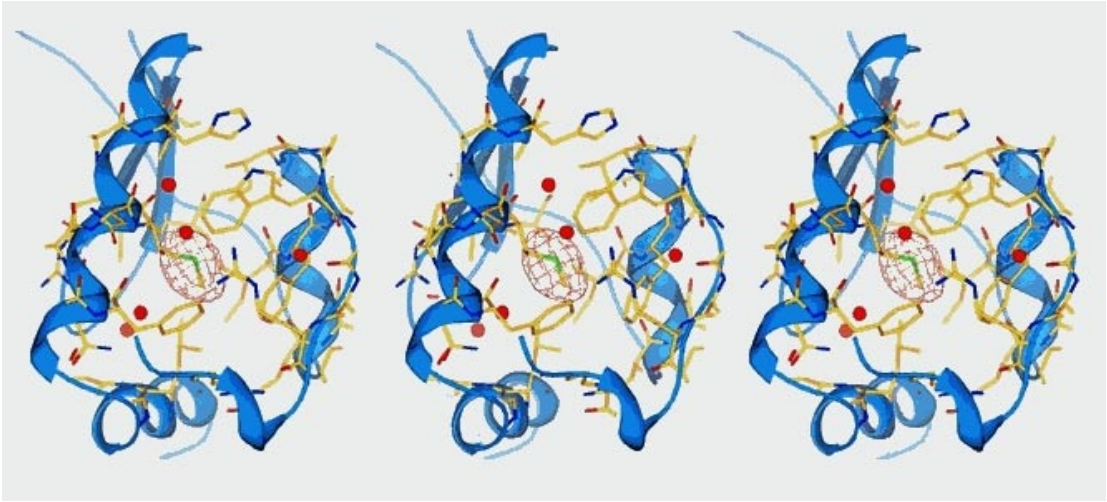


Fig 2

Repositioning of amino acid residues 283, 293, and 295 near Met282 results in a collapsed, more densely packed hydrophobic core, which in turn produces a polymerase with enhanced protein stability.

The structural results indicate that repositioning of amino acid residues surrounding site 282 appears to affect the fidelity properties of the mutant polymerase. Thus, Met282 and the neighboring amino acid residues appear to play an important direct (and indirect) role in the mechanism of selecting the correct dNTP substrates over the incorrect ones.

Publications

Biochemistry, manuscript submitted

Funding

We acknowledge the support of Yorkshire Cancer Research and National Cancer Institute for this work.

The crystal structure of Hepatitis C virus RNA polymerase complexed with nucleotides and metal ions

Damien O'Farrell, Rachel Trowbridge, Nicola Stock, Henric Ekstrand, Dave Rowlands & Joachim Jäger

Introduction

The Hepatitis C virus (HCV) is a human pathogen associated with the majority of cases of transfusion associated and sporadic non-A, non-B hepatitis. Most, if not all infections become chronic and lead to various clinical conditions such as an inapparent carrier state with almost normal liver functions, acute hepatitis and, in about 50% of cases, chronic hepatitis. Approximately 20% of the chronic cases develop liver cirrhosis, which leads to liver failure in about 25% of cirrhotic cases. Furthermore, patients with chronic HCV are at a high risk of developing hepatocellular carcinoma. The virus is transmitted primarily by the parenteral route, and many HCV-infected individuals are intravenous drug users or recipients of blood products. So far, treatment of chronic HCV infections is limited to interferon-α therapy which is successful in only 40% of treated patients. After cessation of therapy, about 70% of these responders relapse and only 25% of patients show a long term prognosis. The response can possibly be improved using combination therapy with ribavirin. Even in this case 60% of patients do not show a long-term response, substantiating the need for a more effective therapy against chronic hepatitis C. Therefore, the RNA dependent RNA polymerase of HCV has been selected as a key target for developing specific antiviral therapy.

Structural studies

Full length HCV polymerase (NS5b, residues 2420-2990, Mr 66,000) is expressed from recombinant baculovirus in Sf9 cells. Yields of 400μg of purified full-length protein, sufficient for kinetic studies, can be readily obtained. For structural work however, a bacterial expression system has been established. Purification is based on affinity chromatography using Ni-NTA superflow resin followed by second step using poly(U) sepharose. This protocol yields pure NS5B (>99%) in large quantities (12mg/l culture). The protein is fully active and crystals can be grown within four days to two weeks.

The crystal structure of HCV RNA polymerase has been solved and refined against data extending to 2.35Å resolution. The residues in both subunits are well ordered except for those at the immediate C-terminus near the hexa-histidine tag. The polymerase folds into the characteristic fingers, palm and thumb subdomains. The overall structure of the NS5b palm domain is practically

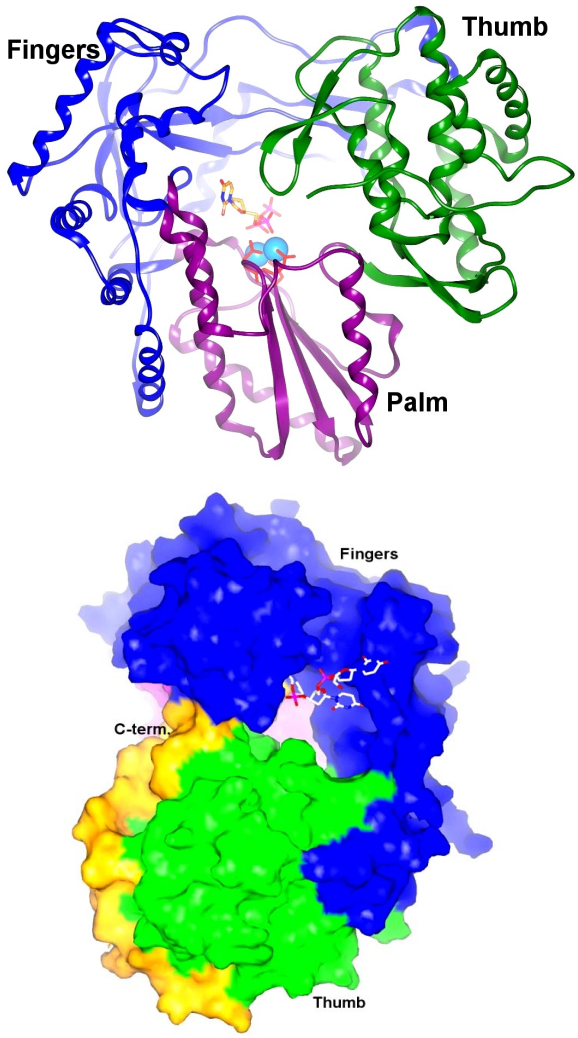


Figure 1:
(a) Ribbon digram of HCV RNA polymerase (NS5B) complexed with UTP. (b) molecular surface rendering of NS5B showing a uracil trimer soaked into the template binding site.

identical to that of other polymerases. The "fingertips" are in close contact with the thumb subdomain. Thus, NS5b has adopted the closed conformation, similar to the catalytically relevant HIV-1 reverse transcriptase/DNA complex.

Soaking experiments with Mg, Mn, ATP, UTP and $r(U)_5$ have revealed the binding sites for the catalytic metal ions, incoming nucleotide-triphosphate and the position of the viral RNA template (See Figure 1). Using HIV-1 RT as a reference, modeling of dsRNA primer-template into the NS5b polymerase active site reveals that the thumb subdomain has to be displaced, possibly indicating how the polymerase translocates on the viral genome. We have now identified a large number of residues which are presumably involved in controlling processivity and fidelity of HCV RNA polymerase. Mutagenesis and kinetic studies are currently underway.

Structure-based drug design and high-throughput screening

Using the programs SPOCK (J.Christopher, Texas A&M) and SPROUT (AP Johnson, Leeds), several small cavities in the vicinity of the active site have been characterised with regard to possible binding sites for highly specific antivirals. Approximately 20 suitable compounds have been chosen, two of which have shown sub-millimolar IC50s. Soaking experiments with pre-grown apo-polymerase crystals have revealed a possible binding site less than 8Å away from the catalytic residues.

In parallel, a highly successful screening program using a non-radioactive, antibody-based platform technology has yielded two families of compounds and several unique inhibitors with near micromolar IC50s. Crystallographic, toxicological and pharmacokinetic studies are currently underway.

References

7th International Meeting on viral Hepatitis, Surfer's Paradise Gold Coast, p 24.

Collaborators

Tobias Restle, MPI Dortmund, FRG
Jens Bukh, NIH, Bethesda, USA
Ken Powell, Arrow Pharamceuticals, London
Berwyn Clarke, Virko Ltd., Cambridge

Funding

We akcnowledge the support of Action Research, BBSRC, BBSRC-CASE (Microsciences Ltd), The Royal Commission of the Exhibition 1851 and RepliZyme Ltd for this work.

Further development and applications of computer programs for *de novo* ligand design

Erika Biro, Krisztina Boda, Michael Briggs, David Cosgrove, Szabolcs Czepregi, Jeff Marchaland, Mark Stewart, Zsolt Szabo, Zsuzsanna Szabo, Attila Ting, Son Van, Miklos Vargyas, and Peter Johnson.

Software Development

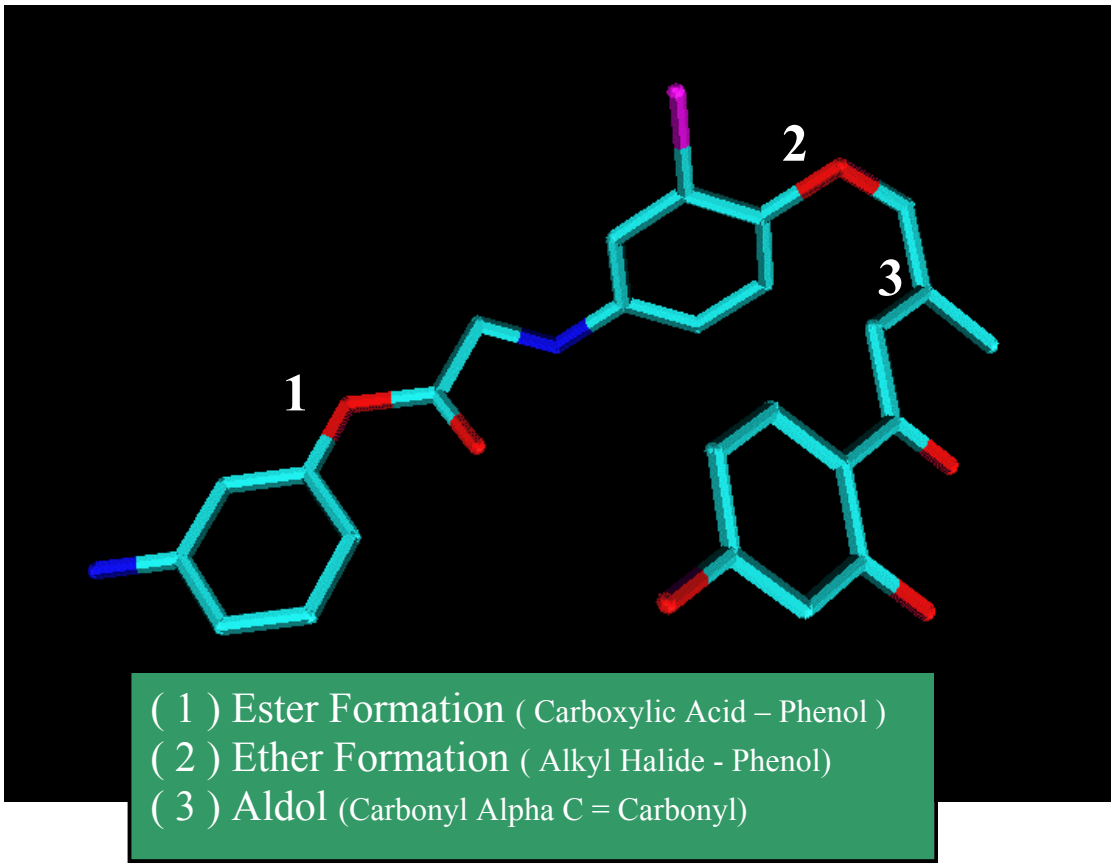
SPROUT and its progeny

Given a 3-D protein structure or a pharmacophore hypothesis, the SPROUT program designs from scratch putative ligands which are complementary to the protein in shape and electrostatics and hence are predicted to bind strongly. Over the past year, development of the program has been concentrated in the following areas:

(a) Modification of the program to build synthetic accessibility into the design process. SPROUT is capable of generating thousands of hypothetical ligands. These can be ranked by estimated binding affinity, but even more important is their synthetic accessibility. With SYNSPROUT (Krisztina) ease of synthesis is built into the design process, because the program uses as building blocks structures of readily compounds and joins them together by processes which mimic specific reactions described in a knowledge base created by the user. Hence putative ligands suggested by the system are accompanied by a synthetic plan. VLSPROUT (Virtual Library Sprout) deals with the same problem in a different way – it docks members of a large virtual combinatorial library to a protein structure in order to identify those members which are likely to bind most strongly and hence should be included in smaller real libraries which are synthesised and screened (Jeff).

(b) Parallelisation of the code (Erika and Zsolt) for the most compute intensive modules so that runs which would previously have taken weeks are now normally completed in less than a day, using the ICAMS Beowulf cluster (22 Pentium III processors running linux and linked by fast Ethernet). As a result of the successful Molecular Interactions JIF bid, we should have an even faster system in place in mid 2001.

(c) Modification of the program to allow it to design *synthetic receptors* instead of ligands (Erika). Since SPROUT is concerned with molecular interactions in general, only minor modifications of the program were needed for the inverse process. RECSPROUT has been used for the design of a receptor for the peptide D-ala-D-ala, which is the target for the antibiotic vancomycin (which binds to this peptide). This receptor is currently being synthesised by Mark. It should be noted that although there has been extensive research into synthetic receptors, most of them only work well in non aqueous solution as they rely almost exclusively on hydrogen bonding to bind the substrate. We hope to do better by building hydrophobic interactions into our design.



Putative inhibitor of CDK2 designed by SYNSPROUT and predicted to bind at the low nanomolar level

Other computational projects

Progress has been made in the application of a novel mathematical technique, *Intetrval Analysis,* to a number of problems in computational chemistry (Miklos and Zsuzsanna). It promises to be a very powerful technique for global optimisation. A novel method of defining shape similarity based on surface curvature has shown promise in predicting the relative activities of a series of inhibitors (David).

Applications

Use of SPROUT with a pharmacophore hypothesis for analogues of the potent antihelminthic agent, paraherquamide, has led to the design of a number of analogues of which six have been synthesised (Michael). Three of these have been found to be active at the low micromolar level by our collaborators (Pfizer) and are currently being patented. SPROUT designed potential inhibitors of TNF Convertase (TACE) have been synthesised (Son Van) and are being screened.

Collaborators

Andrew Leach (GSK)
Kuen Yeap and Chris Dutton (Pfizer)
Ross McGuire (Organon)

Publications

Cosgrove D A, Bayada DM, Johnson A P (2000) A novel method of aligning molecules by local surface shape similarity. *J Comput Aid Mol Des*, **14**, 573-591,

Szabo Z, Vargyas M, Johnson A P (2000) Novel treatment of conformational flexibility using interval analysis. *J Chem Inf Comp Sci*, **40**, 339-346,

Ting A, McGuire R, Johnson A P, et al. (2000). Expert system assisted pharmacophore identification. *J Chem Inf Comp Sci*, **40**, 347-353

Funding
We acknowledge the support of GSK, Pfizer and Organon.

Proteomics and allied technologies

Jeff Keen

Introduction

State-of-the-art instrumentation for the characterisation of proteins is an essential component of an active research environment in structural biology. Over the past few years Leeds has been extremely successful in acquiring the equipment and expertise required to achieve detailed characterisation of proteins, including a Q-Tof™ tandem mass spectrometer (JREI 1998) and an extensive protein chemistry facility (JREI 1999). We have also been awarded a grant to establish a proteomics facility for biological and medical research (JREI 2000) to enhance high-throughput protein analysis in the post-genomic era.

Proteomics

Proteomics is used for the global analysis of protein expression patterns. The term arises from the word proteome - the ***prote***in complement of the gen***ome***, *i.e*. the complete set of proteins expressed from the corresponding genetic information in any given circumstance. It enables a "snapshot" to be taken of the current state of protein expression, providing both qualitative and quantitative information and consequently can be used to investigate changes in protein expression patterns between differing situations. Differences in expressed protein profiles can be used to study the effects of genetic changes (*e.g*. mutation, gene knockout), of environmental challenge (*e.g*. pollution, disease, drug intervention) or to investigate linked biochemical pathways.

The central technology generally involves 2-dimensional (2D-) PAGE for high-resolution separation of proteins, coupled to mass spectrometry for high-throughput protein identification. The aim is to separate complex protein mixtures into discrete spots, which can then be analysed independently to provide identification. This usually involves digestion of the protein and analysis of the peptides using mass spectrometry. The results are used to search databases to provide the identity of the parent protein.

a

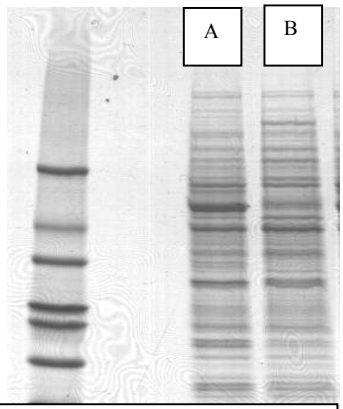


Fig.1. SDS-PAGE of proteins extracted from induced (track A) and control (track B) bacterial cells.

During the course of the last year, we have begun trials with number of systems to elucidate the changes that occur in various circumstances. In one such project, we are attempting to identify genes within a metabolic pathway induced by growing a micro-organism in the presence of an industrial pollutant. Growth conditions have been optimised and enzyme extracts obtained which show marked differences by 1D-PAGE between induced and control states (Fig. 1). We are currently using MALDI-MS mass mapping to identify some of these proteins.

With the award of the JREI grant, a comprehensive facility can be established for more extensive investigations. Typically, samples of proteins are separated by 2D-PAGE in parallel, the protein patterns are visualised through staining and digital images acquired. Sophisticated software can then be used to compare images from a database of different samples to establish standard and altered expression patterns and select proteins showing significantly different levels of expression for identification. The software can drive robotics to excise protein spots, treat them individually with protease to generate peptide mixtures and prepare these mixtures for analysis by mass spectrometry. MALDI-MS is used to screen these digests and produce lists of peptide masses that can be used to search databases to identify the parent protein. A single experiment could generate the identities of dozens or

even hundreds of proteins exhibiting varied expression levels, providing numerous clues to what is happening at the cellular level and identifying targets for detailed structural and functional characterisation.

Protein sequencing

During this last year, we have seen the installation of new state-of-the-art instruments for both N- and C-terminal sequencing (funded through JREI 1999), the latter providing a facility unique within UK academia. The N-terminal instrument is capable of providing sequence at the single picomole level and can routinely generate 10-20 residues from 10-20 pmol starting material. It has been used in a number of internal and external projects for the confirmation of recombinant protein fidelity, the identification of unknown proteins and the provision of *de novo* information for cloning projects. The C-terminal instrument works at the nanomole level, generating a few residues of information. Thus far, it has been used to investigate the C-terminal integrity of a small number of recombinant proteins.

Biochemistry of the hair fibre surface

The technologies associated with proteomics are being applied to an investigation of the biochemical properties of the hair fibre surface. The outermost layer of the fibre, the cuticle, provides an impervious barrier, but little is known of its detailed biochemical composition. A Unilever-funded project (in collaboration with Prof. John Findlay) is using a proteomics-based approach, utilising 1D and 2D electrophoresis and mass spectrometric peptide mapping, to separate and identify protein constituents of this layer. Evidence suggests that specific cytoskeletal keratins are major components of the cuticle. In addition, a targeted approach is being used to specifically label lipid-modified proteins located at the fibre surface in order to identify those components that contribute to barrier protection.

References

Keen, J.N. and Ashcroft, A.E. (1999) Sequence analysis of expressed proteins. *Post-Translational Processing - A Practical Approach.* Eds. S.J. Higgins and B.D. Hames, pp. 1-42, Oxford University Press.

Collaborators

John Findlay, School of Biochemistry and Molecular Biology, Leeds.
Unilever Research, Port Sunlight.

Funding

We acknowledge the support of Applied Biosystems, Dionex and the University of Leeds.

A general, two-directional synthesis of *C*-α(1→6)-linked disaccharides

Michael Harding, Robert Hodgson and Adam Nelson

Introduction

C-Linked glycosides are a particularly interesting class of carbohydrate mimetics; these mimetics are resistant to enzymatic degradation, have potential as inhibitors of glycosidases and glycosyl transferases and often have biological activity and conformational properties which are similar to natural oligosaccharides.

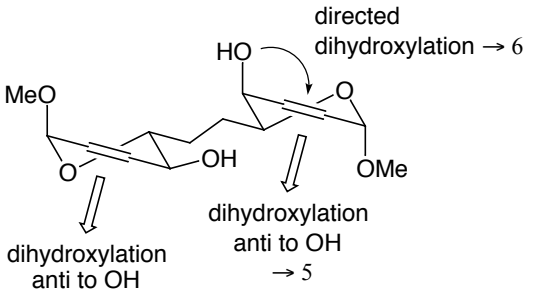
A two-directional synthesis of *C*-linked disaccharides

We have developed a general synthesis of *C*-α(1→6)-linked disaccharides (see Scheme). The power of our approach lies in the ability to choose at a late stage in the synthesis which diastereomeric analogue is prepared. In each of the syntheses, the diol **1** was a key intermediate.

Dihydroxylation of both of the alkenes of **2** under Upjohn conditions (cat. OsO_4/NMO) occurred opposite to the adjacent hydroxyl groups to give, after acetylation, the hexaacetate **7**. This two-directional approach is very efficient indeed: in just two steps, six new stereogenic centres have been introduced with almost complete stereocontrol. The hexaacetate **7** is a protected *C*-linked Alt-α(1→6)-Man mimetic in which C-6 of the altrose ring has been replaced with a methoxy group.

In a similar manner, the diastereomeric diol **4**, synthesised by Mitsunobu inversion of **2** and hydrolysis, was converted into the protected *C*-linked disaccharides **8** and **9**. Hence, double dihydroxylation of **4**, either opposite to, or directed by, the axial hydroxyl groups, gave, after acetylation, **8** and **9**. The di-THPs **8** and **9** are protected Gal-α(1→6)-Gul and All-α(1→6)-Tal mimetics respectively.

A two-directional synthetic strategy does not, of course, restrict our approach to the synthesis of C_2-symmetric disaccharide mimetics. For example, inversion of one of the homotopic alcohols of **2**, and hydrolysis, gave **3** in which the dihydropyran rings were stereochemically differentiated. Dihydroxylation of **3**, *anti* to both of the hydroxyl groups gave the protected Gal-α(1→6)-Man mimetic **5** (Figure).

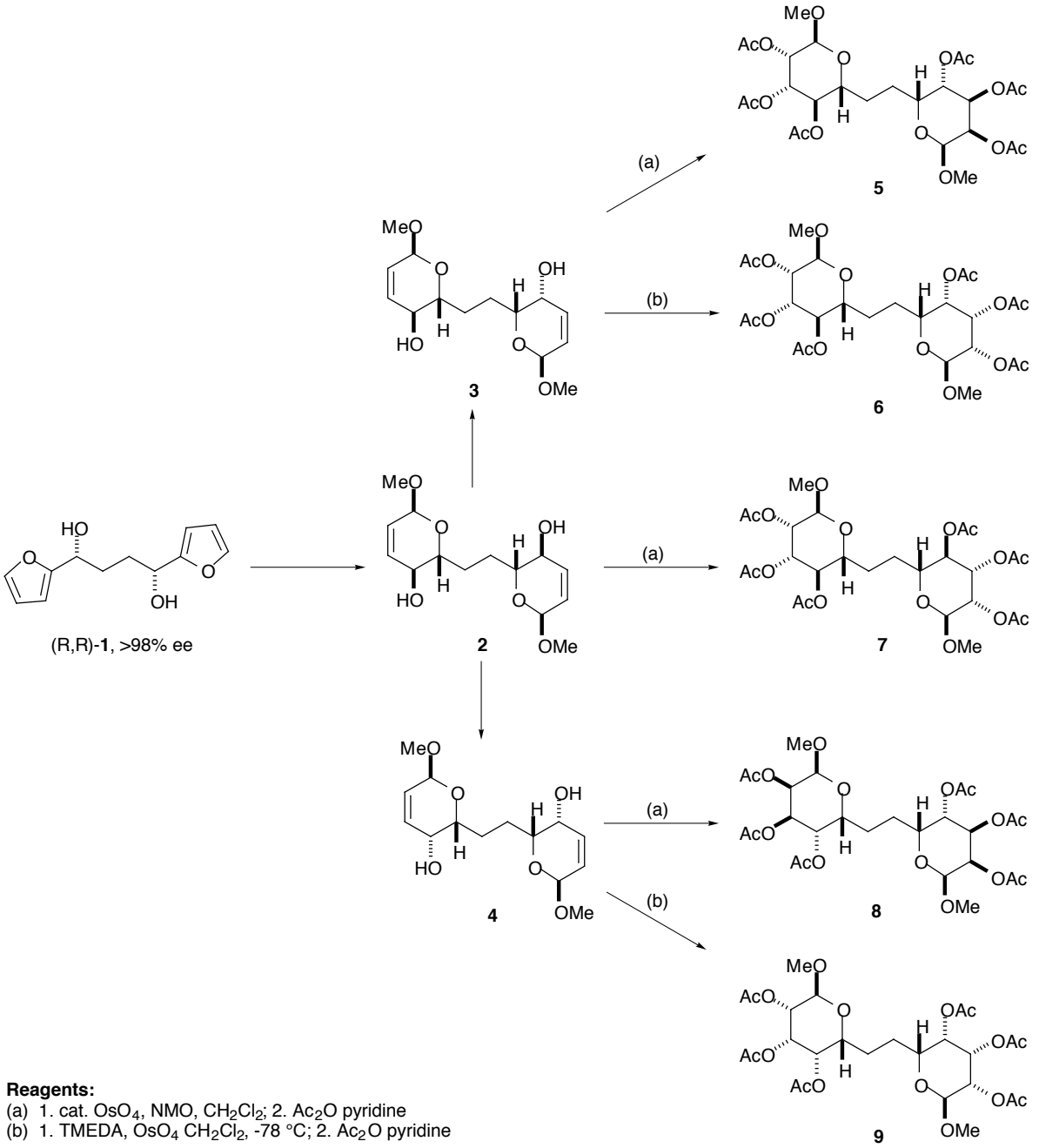


Figure

More remarkably, the diol **3** could be elaborated in a two-directional fashion such that the stereochemical outcome of dihydroxylation was different in each of the rings. Hence, dihydroxylation of **3** under Donohoe's conditions (TMEDA, OsO_4, CH_2Cl_2, −78 °C) was directed by the axial alcohol but occurred *anti* to the equatorial alcohol to give, after acetylation, the protected All-α(1→6)-Man mimetic **6** (Figure).

Summary

This work is the first synthesis of *C*-linked disaccharides entirely from non-carbohydrate based precursors. Most other syntheses rely on the coupling of sugar derivatives, for example using Suzuki, Ramberg-Bäcklund, Wittig, Henry and metathesis reactions. A particular merit of our approach, which makes it amenable to the synthesis of libraries of stereoisomeric carbohydrate mimetics, is that several diastereomeric *C*-linked disaccharides may be prepared by minor variation of a general reaction sequence.



Collaborators

Tajir Mahid, Aventis, New Jersey, USA

References

Nelson, A. and Harding, M. (2001) A general, two-directional synthesis of *C*-α(1→6)-linked carbohydrate mimetics: the first synthesis of *C*-linked disaccharides from non-carbohydrate based starting materials *Chem. Comm.*, in preparation.

Funding: Sponsored by the Leverhulme Trust

Biology of bacterial peptidoglycan synthesis

Lars Hesse, Chris Storey, Simon Phillips, Ian Chopra

Introduction

Peptidoglycan (PG), which forms a sacculus around the bacterial cell, is an essential cell wall polymer since interference with its synthesis or structure leads to loss of cell shape and integrity followed by bacterial death. Peptidoglycan contains chains of alternating N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) residues, cross-linked through pentapeptide side chains attached to MurNAc. The pentapeptide side chain sequence is L-Ala (occasionally L-Gly)-D-Glu-X-D-Ala-D-Ala where X is usually a diamino acid and often meso-diaminopimelic acid (m-Dpm) or L-Lys. The biosynthesis of peptidoglycan is a complex three stage process. The first stage involves the assembly of the disaccharide-peptide monomer unit by enzymes located in the cytoplasm, or at the inner surface of the cytoplasmic membrane. The second stage involves transfer of the monomer unit across the cytoplasmic membrane and the third stage polymerisation of the monomer unit on the outer surface of the membrane and concomitant binding of the nascent peptidoglycan to the pre-existing cell wall.

Drug targets

The synthesis of the pentapeptide chain is facilitated by enzymes of the Mur pathway and involve MurA-F the D-Ala-D-Ala ligases (Ddl) and alanine racemace (Alr). MurA and –B convert UDP-N-acetylglucosamine into UDP-N-acetylmuramic acid, the nucleotide sugar used by the ligases MurC-F to produce the pentapeptide chain which is later transferred across the cytoplasmic membrane to be incorporated in the growing cell wall. In contrast to well established approaches exploiting the later stages of PG synthesis (e.g. the discovery and development of β-lactam and glycopeptide antibiotics) we are using structure-based approaches for the design and synthesis of selective inhibitors of early stage Mur enzymes in collaboration with the Chemistry Department in Leeds. The most promising targets are enzymes using D-amino acids or substrates with no eukaryotic counterparts providing good prospects for discovering selective bacterial inhibitors.

Rationale of drug screening

For the drug design it was necessary to use already available crystal structures and clone and overexpress the target enzymes. For the first approaches of drug design and synthesis we used MurD, the D-glutamate adding ligase of *E. coli* (Fig. 1). The enzyme was cloned, overexpressed and purified to homogeneity using standard techniques like ion exchange and gel filtration chromatography.

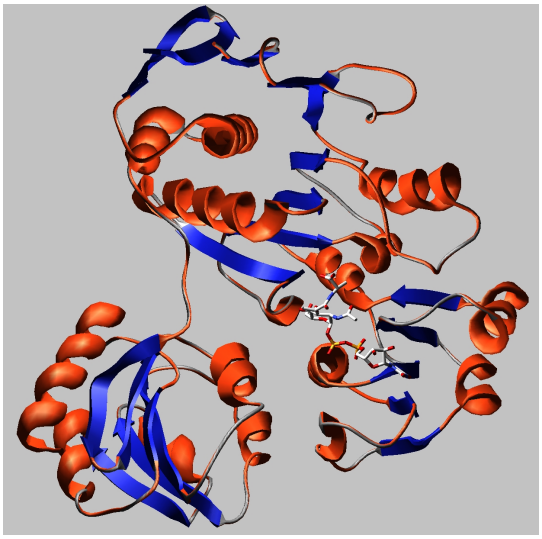


Figure 1: *E. coli* MurD with substrate UDP-MurNAc-L-Ala bound (PDB: 1EEH)

The substrate for the target enzymes is not commercially available and had to be generated using the enzymes MurA-C and was purified by ion exchange chromatography. Compounds based on the structure based design were synthesised and provided to us by our collaborators at the School of Chemistry, Leeds. For validation of the enzyme's kinetic parameters and drug screening process we used an endpoint and a coupled continous

enzyme assay and applied the endpoint assay in a high throughput screening (HTS). We hope to be able to co-crystallise MurD with compounds showing inhibitory effects after the final evaluation of the HTS results. This structure would then be used to refine the drug design and improve binding of the putative inhibitor to its target.

Peptidoglycan paradox in *Chlamydiae*

We also have strong interests in the physiology of Chlamydia, especially *C. trachomatis*, a gram-negative intracellular pathogen causing serious infections in humans. Analysis of the *C. trachomatis* genome shows the presence of most of the necessary enzymes for the synthesis of PG but no PG or PG-like polymer was detected so far. We therefore started to clone the *C.t.* Mur pathway genes and overexpress them in *E. coli*. MurC, the first enzyme to add an amino acid to UDP-MurNAc, shares one open reading frame with the D-Ala-D-Ala ligase, two steps involved in the same pathway but not in consecutive reactions. So far also nothing is known about the sequence of amino acids in the putative pentapeptide of Chlamydia. The recombinantly expressed Mur enzymes from *C.t.* might be used to identify their substrate specificity. Using this approach we will be able to produce a series of supposed pentapeptides *in vitro* to generate antibodies. These might finally enable us to identify PG in living chlamydial cells.

Collaborators

SmithKline Beecham Pharmaceuticals, Upper Providence, Pensylvania, USA and Harlow, UK
Prof. Peter J.F. Henderson, Astbury Centre, Leeds
Dr. David J. Adams, Division of Microbiology, Leeds
Prof. Ron Grigg, Prof. Peter Johnson, Dr. Colin Fishwick, School of Chemistry, Leeds

References

Chopra, I., Storey, C., Falla, T.J. & Pearce, J.H. (1998) Antibiotics, peptidoglycan synthesis and genomics: the chlamydial anomaly revisited. *Microbiology* **144**:2637-2678

Storey, C. & Chopra, I. (2001) Affinities of β-Lactams for Penicillin Binding Proteins of Chlamydia trachomatis and Their Antichlamydial Activities. *Antimicrobial Agents and Chemotherapy*. **45**: 303-305

Hesse, L., Storey, C., Kallender, H., Sylvester, D.R., Lonsdale, J.T., Phillips, S.E.V., Henderson, P.J.F., & Chopra, I. (2001) Enzymatic activity and properties of MurC from *Chlamydia trachomatis*, an enzyme with UDP-N-acetylmuramoyl:L-Alanine Ligase Activity (in preparation).

Funding

We acknowledge the support of SmithKlineBeecham.

Crystallographic studies of copper amine oxidase

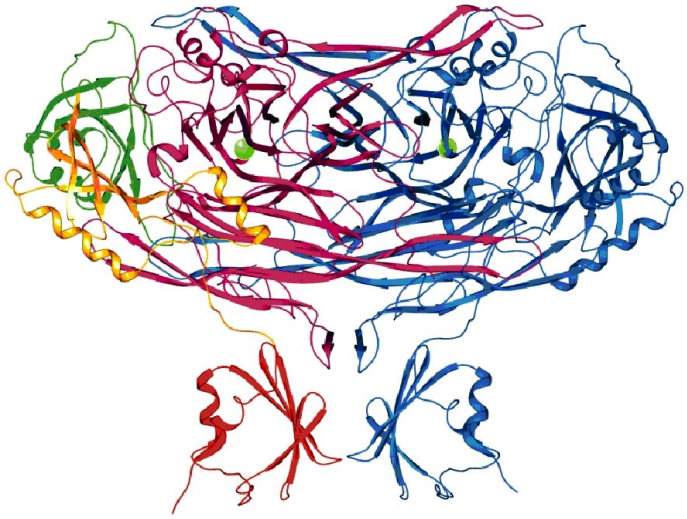
Helen Dawkes, Peter Knowles, Christian Kurtis, Mike McPherson, Simon Phillips, Carrie Wilmot

Introduction

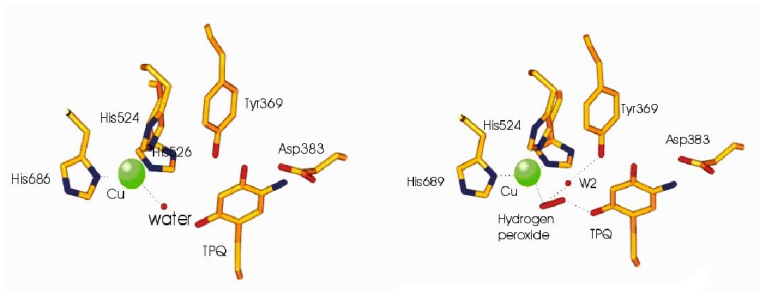
Copper amine oxidases are ubiquitous metalloenzymes. Their function in prokaryotes and lower eukaryotes is to utilise amines as a source for carbon and nitrogen. In higher eukaryotes, their roles are less well understood, but have been linked with cell signalling, growth and development, and cell death. Despite this apparent diversity, their fundamental role is to catalyse the oxidation of primary amine substrates to their corresponding aldehydes. Amine oxidases are homodimeric enzymes ranging in size from 140-180 kDa. We work on copper amine oxidase from *Escherichia coli* (ECAO); the crystal structure of which has been determined to 2.0 Å. Each subunit consists of four distinct domains and contains a single copper ion and a quinone co-factor, 2,4,5-trihydroxyphenylalanine quinone (TPQ) at the active site. Also found at the active site are three His residues and one or two water molecules in an approximate square pyramidal arrangement around the copper. The TPQ cofactor is generated from an intrinsic tyrosine in the amino acid sequence by a self-processing event that requires only the bound copper ion and molecular oxygen.

The catalytic mechanism of ECAO

The mechanism of ECAO involves a two-electron redox reaction. The reaction is complex and can be divided into two half-reactions: reductive and oxidative. During the reaction, the enzyme is reduced via the formation of a Schiff base between TPQ and the substrate amine, followed by liberation of the product aldehyde. The oxidative half of the reaction involves the copper ion and the activation of molecular oxygen to recycle the reduced enzyme back to its



(a) ECAO homodimer



(b) substrate reduced ECAO

(c) after prolonged aerobic exposure to substrate

Figure 1 (a) ECAO homodimer, one monomer is coloured blue and the other is coloured by domains. (b) and (c) essential features of the active sites of crystals soaked in an amine substrate and flash frozen during the catalytic turnover. Copper ions are shown as green spheres.

oxidised resting state. This is accompanied by the release of ammonia and hydrogen peroxide. In recent years this mechanism has been intensively studied, but the exact details of oxygen activation, the fundamental process of aerobic biology, remain poorly understood.

Visualising catalytic intermediates

By employing the technique of flash-freezing, we have trapped reaction intermediates in catalytically competent crystals. This technique 'freezes' the motions of this enzyme so snapshots of its activity can be observed by solving the structures by X-ray crystallography. To view the short-lived biochemical reaction, however, requires that these steps be synchronised in some way so that all the molecules in the crystal are trapped at the same stage of the reaction. Crystals have been freeze-trapped after exposure to substrate, either anaerobically or aerobically, and studies of intermediates have been determined (Figure 1, b and c). Oxygen mimics, such as nitric oxide, have also been used to investigate the site of initial dioxygen binding. We have found that nitric oxide replaces the position of the water molecule found in Figure 1b.

Conveniently, the oxidation states of TPQ have distinctive visible spectra. Thus, we have been using single-crystal spectrophotometry to help identify the accumulation of an intermediate during catalytic turnover. We are currently using this technique to trap various intermediates with the overall aim of mapping the active reaction cycle.

References

Wimot, C.M., Hajdu, J., McPherson, M.J., Knowles, P.F., Phillips, S.E.V. (1999) Visualisation of Dioxygen Bound to Copper During Enzyme Catalysis. *Science* **286** 1724-1728

Funding

We acknowledge the support of the BBSRC

Crystal structure of the Holliday junction- resolving enzyme T7 endonuclease I at 2.1Å resolution

Jon Hadden, Maire Convery and Simon Phillips

Introduction

Homologous genetic recombination is important in DNA rearrangements and in the repair of double-strand breaks in DNA. The central DNA intermediate in this process is the four-way (Holliday) junction, and recognition and manipulation of this structure by proteins are important elements of the mechanism. The penultimate stage of recombination requires the resolution of the junction into component duplex species by junction-specific nucleases. In general these relatively small, basic proteins bind to DNA junctions in a highly structure-selective manner. Structure-selective interactions are fundamental to the mechanism of genetic recombination.

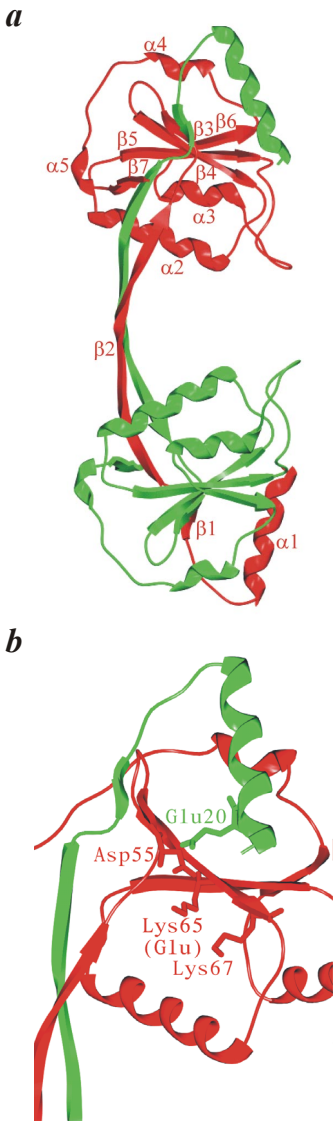


Figure1
a) Overall structure of an endonuclease I homodomer. Individual monomers are shown in red and green.
b) Close up view of the active site of endonuclease I.

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. The crystal structures of two junction-resolving enzymes have been previously reported. The structure of RuvC has been determined at 2.5 Å resolution, while that of T4 endonuclease VII has recently been described at 2.4 Å resolution, together with an inactive mutant at 2.1 Å resolution. No structural similarity is discernible between these two enzymes. We have determined the crystal structure of a third junction-resolving enzyme T7 endonuclease I at 2.1 Å resolution. This is the first structure of a junction-resolving enzyme that falls into the superfamily of proteins that includes the restriction enzymes.

Endonuclease I forms an intimately associated symmetrical homodimer comprising two domains (Fig. 1*a*). Each domain is composed of residues 17-44 from one subunit and residues 50-145 from the other. The two domains are connected by a bridge that forms part of an extended ß-sheet (ß2). Endonuclease I has an extensive dimer interface with many interactions along the ß-sheet bridge and between the N-terminal region (17-44) and the body of the other subunit. This is consistent with our observation that in free solution, subunit exchange in endonuclease I is undetectable in the absence of denaturing agents.

Each domain comprises a central five-stranded mixed ß-sheet, flanked by five ß-helices with one strand and one helix contributed by the other subunit in the dimer. The compact nature of each domain suggests a stable structure that may function independently.

Mutagenesis experiments have previously identified five acidic residues which could

potentially be involved in catalysis in endonuclease I, Glu 20, Glu 35, Asp 55, Glu 65 and Asp 74. Glu 20, Asp 55 and Glu 65 (mutated to Lys 65 in the crystallised protein) are clustered and form an acidic cluster on the surface of the protein and we suggest that these residues constitute the active site of the nuclease. Closer examination of this region (Fig. 1*b*) reveals that the arrangement of residues is closely similar to that found in the active sites of a number of well-characterised restriction endonucleases, including, *Bgl*I, *Eco*RV, *Eco*RI and *Fok*I. The restriction endonucleases normally possess three or four active site residues, typically two (*Bam*HI, *Eco*RI, *Fok*I) or three (*Bgl*I, *Eco*RV) acidic amino acids and a lysine residue are involved in catalysis. The acidic amino acids are thought to be involved in chelating one or two metal ions, whilst the lysine residue probably plays a role in stabilising the transition state or the product of the cleavage reaction. A structural alignment of a number of restriction enzymes suggests that the active site of endonuclease I belongs to the same family as *Bgl*I and *Eco*RV, and that lysine 67 could also be involved in catalysis.

T7 endonuclease I shows little structural similarity with T4 endonuclease VII. The latter is almost totally α-helical except for a section of β-sheet organised as part of a zinc domain that carries a number of acidic and histidine side chains required for catalysis. RuvC has an α/β fold, but with different topology from endonuclease I. The protein domain-recognition program DALI failed to identify any significant structural similarity between the three proteins.

By contrast, there is clearly a significant structural similarity between the active site of endonuclease I and those of a number of restriction enzymes, discussed above. Moreover, the *archaeal* junction-resolving enzyme Hjc has an active site that is predicted, by sequence motif and site-directed mutagenesis, to be very similar to that of endonuclease I and the restriction enzymes. These enzymes, together with MutH and _exonuclease, are therefore likely to be related to an ancestral nuclease, forming a superfamily of nucleases. The essential nuclease function has acquired different types of specificity in the various enzymes, so that the restriction enzymes exhibit high sequence specificity, while the resolving enzymes such as endonuclease I are highly selective for the branched structure of DNA junctions.

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

References

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

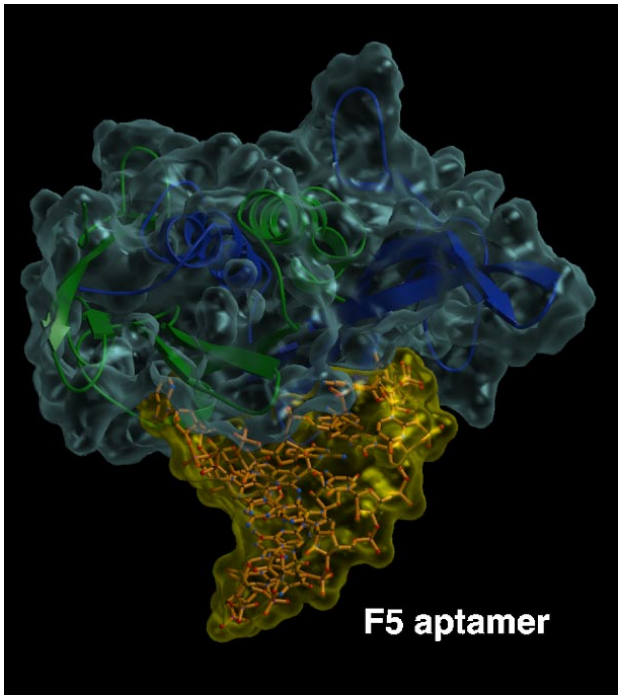
Funding: We acknowledge the support of The Wellcome Trust

Structural determination of a chemically modified RNA translational operator complexed with bacteriophage MS2 coat protein

Wilf Horn, Maire Convery, Nicola Stonehouse, Chi Trinh, Peter Stockley and Simon Phillips.

Introduction

MS2 is a T=3 icosahedral bacteriophage with a single stranded RNA genome that infects *E.coli.* The MS2 capsid structure has been determined to a resolution of 2.8 Å by Lars Liljas' group in Uppsala, Sweden and has an approximate diameter of 275 Å. The coat protein subunit exists as three conformers, (A, B and C), with the same amino acid sequence but with distinct tertiary structures. Two types of protein dimer (AB and CC) are present within the capsid shell. As a consequence of conformer folding and packing, pores connecting the capsid interior to the external environment are present at both the 3-fold and 5-fold symmetry axes. The pores are large enough in size to allow the passage of single stranded nucleic acid molecules in to and out of the capsid. *In vivo,* a 19nt RNA stem loop within the viral genome binds to a specific site on a coat protein dimer, acting both to initiate phage self assembly and to inhibit viral replicase gene translation. This translational repression complex formed between the RNA stem loop and the MS2 coat protein has been utilised for many years here in Leeds to study sequence specific protein-RNA interactions. A series of RNA aptamer families, generated via SELEX, have been shown to bind, with varying affinity, to the capsid dimers at the same site as the wild-type stem loop. One of these families, (F5) (Figure 1), has recently been used to investigate the effects on binding affinity when nucleotides with modified functional groups are incorporated at defined positions within the stem loop[1].



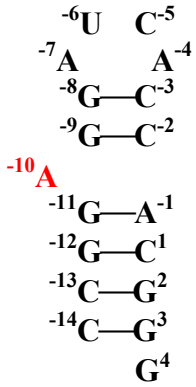


Figure 1;
(Top), Structure of F5 RNA aptamer-coat protein dimer complex showing van der Waals surfaces of protein (blue) and RNA (yellow).
(Bottom), Secondary structure of F5 RNA stem loop aptamer, Adenine at –10 position that is substituted with 2'-deoxy-2aminopurine (2AP) is indicated in red.

Structural determination of modified RNA operator

The unpaired adenine at the –10 position has been shown to be an important determinant of binding affinity, forming three putative hydrogen bonds with the MS2 capsid. In order to determine the role of this nucleotide with regard to binding affinity the adenine at this

position was substituted for 2'-deoxy-2aminopurine (2AP) (Figure 1). This change was predicted to lead to the abolition of a protein-RNA hydrogen bond with subsequent reduction in affinity for coat protein. In fact the binding affinity of this modified F5 operator was higher than any previously characterised operator[1]. In order to explain this seeming paradox the crystal structure of the operator-capsid complex was determined to a resolution of 2.8 Å. The F5 RNA operator with 2AP incorporated at the –10 position, produced via solid phase chemical synthesis, was soaked into pre-crystallised RNA free capsids entering by the 3 and 5-fold capsid pores. This allows the determination of the RNA structure without the need for RNA crystals. Surprisingly the crystal structure indicates that the modified operator is still able to make the same protein-RNA contacts as the F5 operator. The increase in affinity however may be explained by the formation of an additional hydrogen bond and other more subtle changes in the protein-RNA interface.

References

1) Parrott, A.M., Lago, H., Adams, C.J., Ashcroft, A.E., Stonehouse, N.J. & Stockley, P.G. (2000) RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour. *Nucleic Acids Research,* **28,** 489-497

Funding

We acknowledge the support of BBSRC.

Structural studies of AhrC, the arginine activator/repressor from *Bacillus subtilis*

Caitriona Dennis, James Garnett, Mark Parsons, Simon Phillips.

Introduction

The concentration of the amino acid L-arginine within the gram positive bacterium *Bacillus subtilis* is controlled by a regulatory protein, AhrC, the product of the *ahr*C gene. This protein is active as a hexamer and has the ability to repress and activate genes involved in arginine metabolism.

Biosynthesis of arginine in *Bacillus subtilis* occurs in 7 steps. The genes of the enzymes involved are located in 2 clusters within the genome (*arg*CJBD-*cpa*-F and *arg*GH). Upon an elevated concentration of arginine within the cell, AhrC binds arginine and interacts with operators within the promoter regions of the gene clusters to repress transcription of the genes. The regions of DNA recognised by AhrC are termed ARG boxes and are 18 bp pseudo-palindromic sequences. Examination of the *arg*C promoter and gene has revealed two AhrC binding sites. The first, which is the higher affinity site, contains two ARG boxes separated by 11 bp and lies within the *arg*C promoter whilst the second is a lower affinity site and contains a single ARG box located within the *arg*C structural gene.

Catabolism of arginine occurs in a pathway involving 6 enzymes. The genes for which are also located in two clusters, *roc*ABC and *roc*DEF. AhrC, in an arginine-dependent fashion interacts with operators within the promoter regions of *roc*A and *roc*D. These sites contain only one ARG box which is adjacent to the transcription start site.

It has been shown that affinity for the catabolic operators is at least 10-fold less than for the biosynthetic gene promoters[1]. This is consistent with the theory of co-operative binding of AhrC to the tandem repeats of ARG boxes upstream of the *arg*C and *arg*G.

Homologues of AhrC

Regulators of arginine metabolism have also been identified in *Escherichia coli, Bacillus stearothermophilus* and *Salmonella typhimurium* where they are called ArgR. These proteins, share homology with AhrC and possess the ability to repress the biosynthetic gene clusters. However, they have not been experimentally implicated in activation although high sequence and structural homology suggest they may carry out this function. The best characterised of the AhrC homologues is ArgR from *E.coli* (ArgREc). AhrC and ArgREc share 27% identity and can cross function *in vivo* to some extent. AhrC can carry out repression in *E.coli* but the reverse does not happen as ArgREc cannot repress the genes in *B.subtilis*. Both proteins are oligomeric and require arginine for the high-affinity binding to operators. The AhrC monomer has a molecular weight of 16.7 kDa and can be proteolytically cleaved to yield two domains of similar

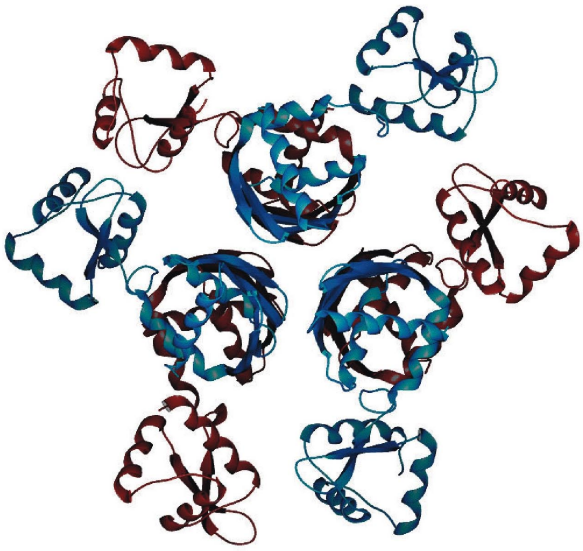


Figure 1: View down the three-fold of AhrC. The top trimer is coloured blue and the bottom trimer is red. The core domains form a compact 32 symmetrical hexamer and the DNA binding domains lie approximately in pairs around the periphery.

size. The N-terminal domain is responsible for DNA binding and two of these domains form a pair to bind one ARG box. The C-terminal domain termed the 'core' domain oligomerises and binds arginine. Structural studies on ArgREc core domain revealed a compact hexamer displaying 32 symmetry is formed upon oligomerisation. Binding sites for arginine were reported to be located near to the trimer-trimer interface.

Structure of AhrC

The crystal structure of AhrC has been solved in Leeds, to a resolution of 2.7 Å. The structure was solved using molecular replacement (AmoRe) with the ArgREc core hexamer as trial model. Information about crystal packing from electron microscopy was also incorporated.[2] To obtain an interpretable electron density map, combined phases from a heavy atom soaked crystal were used. The structure has been refined to an Rfactor of 22% (FreeR 27%). The intact repressor/activator protein comprises the core domains forming a compact hexamer with the DNA binding domains around the periphery of the core [Figure 1]. These domains do not obey the strict 32 symmtery.[3] Upon comparison with the core hexamer from *E.coli*, it can be seen that there is a rotation of one trimer with respect to the other when arginine binds, forming a tightened interface. This may in turn allow the DNA binding domains to move closer producing a more favourable DNA binding site.

Investigating arginine and DNA binding

Attempts are underway to grow crystals of holo-AhrC (arginine bound) to investigate the structural changes with regards to the position of the DNA binding domains. Ultimately it is hoped to visualise the intact repressor in its active form bound to its DNA binding sites. This will give a valuable insight into the mechanism of repression, via the formation of the DNA repression loop as seen in the *lac* repressor.

References

[1]Miller, C. M., Baumberg, S. & Stockley, P. G. (1997) Operator interactions by the *Bacillus subtilis* arginine repressor/activator, AhrC; a novel positioning and DNA-mediated assembly of a transcriptional activator at catabolic sites. *Mol. Microbiol.* **26,** 37-48

[2]Glykos, N. M., Holzenburg, A. & Phillips, S. E. V. (1995) Low-resolution structural characterization of the arginine repressor/activator from *Bacillus subtilis*: a combined X-ray crystallographic and electron microscopical approach. *Acta Cryst*. **D54,** 215-225

[3]Dennis, C. A., Glykos, N. M., Parsons, M. R. & Phillips, S. E. V. The structure of AhrC, the arginine repressor/activator protein from *Bacillus subtilis*. Submitted to *Acta Cryst. D.*

Funding

We acknowledge the support of BBSRC.

Structural studies of processing in galactose oxidase

Susan Firbank, Peter Knowles, Mike McPherson, Simon Phillips.

Introduction

Galactose oxidase (GO) is a monomeric copper containing enzyme that catalyses the oxidation of primary alcohols to their corresponding aldehydes. The enzyme has a wide range of substrates, but is strictly stereo-specific; whilst D-galactose is a good substrate, galactose oxidase shows no activity with L-galactose or D-glucose. In addition to the copper, the two electron reaction requires a second redox active centre, which is provided by a radical situated at a tyrosine residue (Tyr 272). The crystal structure of the mature enzyme revealed the presence of a novel thioether bond - a covalent link between Tyr 272 and a cysteine (Cys 228) at the active site (Ito *et. al*. (1991), *Nature*, **350**, 87-90). To investigate how this thioether bond forms, current work in Leeds is investigating the processing of this enzyme from its precursor.

Processing intermediates of GO

Galactose oxidase is thought to undergo several steps during processing. Cleavage of the signal sequence gives rise to a form that contains a seventeen amino-acid N-terminal pro-sequence, in addition to the mature sequence. This pro-form does not yet have the thioether bond at the active site. Generation of mature enzyme requires at least two further processing events - cleavage of the pro-sequence and formation of the thioether bond. Both of these events are autocatalytic, only requiring the addition of copper and molecular oxygen[1]. To try and elucidate these processes, the precursor of galactose oxidase (containing the pro-sequence and no thioether bond) has been purified and crystallised.

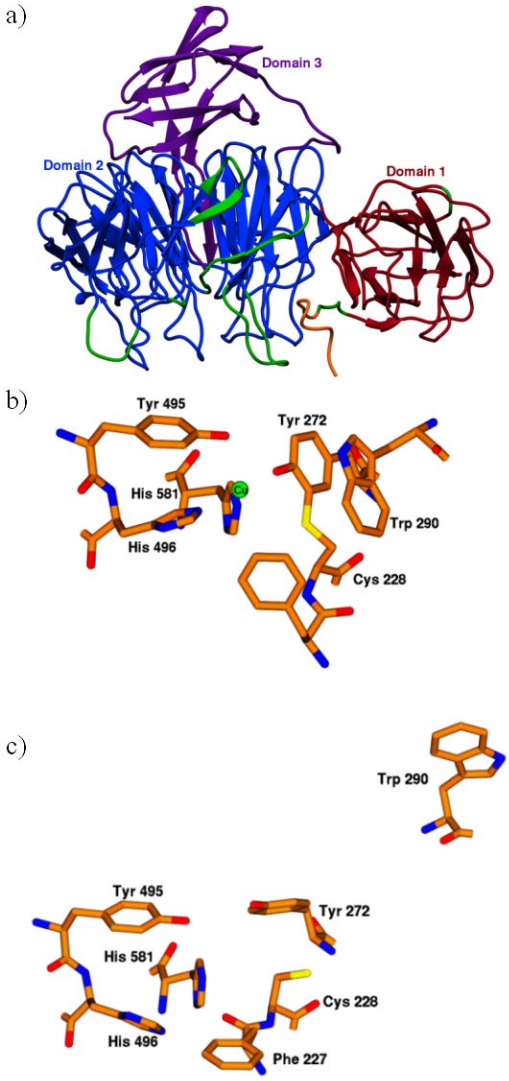


Figure1

a) Galactose oxidase precursor. Regions that differ significantly between the precursor and mature are coloured green. The pro-sequence is coloured yellow.

b) Representation of the active site in mature galactose oxidase.

c) Representation of the active site residues as found in the precursor.

Crystal structure of the precursor

X-ray diffraction data were collected on a crystal of the precursor, and the structure solved by molecular replacement using the mature structure as a model. The structure has been refined to a resolution of 1.4Å, and shows several differences between the precursor and mature form. Five regions of mainchain differ significantly between the two forms. The position of four of these regions can be attributed to the presence of the pro-sequence, whilst the position of the fifth appears to be due to differences at the active site. The most noticeable changes at the active site are the arrangement of the two residues that form the thioether bond, and also of the tryptophan that stacks over the thioether bond in the mature form. There are also slight differences in the positions of the other copper ligands. It is hoped that this structure will provide a starting point for modelling and understanding thioether bond formation.

Current and future work

Work is currently underway to obtain structures of intermediates between the precursor and mature forms. Since both pro-sequence cleavage and thioether bond formation require only copper and oxygen, anaerobic conditions can be used to try and trap intermediates. Copper has been soaked into a crystal of the precursor under such oxygen free conditions, and this structure is currently being refined to 2.4Å. We are also in the process of trying to purify and crystallise a form of the protein in which the pro-sequence has been cleaved, but the thioether bond has not yet formed. These structures, along with the solution work carried out by our collaborators, should give us more of an insight into the starting point for thioether bond formation.

Collaborators

M.S. Rogers and D.M. Dooley. Department of Biochemistry and Chemistry, Montana State University, Bozeman, Montana 59717.

Reference

Rogers, M.S., Baron, A.J., McPherson, M.J., Knowles, P.F., and Dooley, D. (2000) Galactose oxidase pro-sequence cleavage and cofactor assembly are self processing events. *J. Am. Chem. Soc*. **122,** 990-991

Funding

We acknowledge the support of BBSRC.

Control of peripheral and integral membrane protein localisation in eukaryotes

Vas Ponnambalam & Emma Stanley

Background

The maintenance of intracellular compartments such as the nucleus, endoplasmic reticulum, Golgi apparatus, mitochondria, lysosomes and plasma membrane is essential for the partitioning of biochemical reactions at separate locations within a eukaryote cell. How are the dynamics of each compartment controlled to maintain membrane boundaries? We have chosen model proteins that localise to compartments such as the Golgi apparatus and plasma membrane to study such processes in membrane systems.

Regulation of protein localisation by interaction with phospholipids

Cellular phosphoinositides are key regulators many cellular processes including intracellular signalling, gene expression and membrane traffic. A major area of interest is how lipid-binding proteins, lipid kinases and phospholipids co-operate to regulate protein selection within a transport vesicle to trigger membrane curvature and vesicle biogenesis before vesicle fusion with a different target membrane. A series of biophysical and structural biology projects are under way to decipher the nature of protein-protein and protein-lipid interactions to understand the three-dimensional interplay between molecules involved in such processes.

We have chosen two well-characterised protein domains to study the localisation of peripheral membrane proteins by interaction with membrane-bound phospholipids. These pleckstrin homology (PH) domains are 100-20 residue modules that are found in many eukaryote proteins (1) including phospholipase Cδ1 (PLCδ1) and oxysterol-binding protein (OSBP). These proteins localise to the plasma membrane and Golgi apparatus respectively via PH domains. We have used a green fluorescent protein (GFP) fusion with PH/PLCδ1 to show its ability to mediate plasma membrane localisation (Fig. 1). The PH/OSBP domain can similarly mediate Golgi localisation; in both cases localisation is critically dependent on PH domain binding to different phospholipids. We are developing biophysical assays and structural biology projects to study protein-phospholipid interactions. Such systems will be used to develop and test hypotheses on how protein recruitment to membrane bilayers is regulated by interaction with different phospholipids. PH domains are also found in many enzymes implicated in human disease; these studies will be valuable in designing drugs or screening compounds that may have therapeutic purpose.

Control of constitutive protein secretion

The majority of proteins displayed on the surface of a eukaryote cell are delivered by the secretory pathway from the endoplasmic reticulum via the Golgi apparatus. Soluble secretory proteins, enzymes, growth factor receptors, cell adhesion molecules and immune complexes are all delivered to the plasma membrane by this route. Many of the cellular components involved in the early part of the secretory pathway have been elucidated e.g. NSF, SNAPs, GM130, p115, GRASP65, giantin, COPI and COPII coat proteins. However, the key final step in membrane traffic from the Golgi apparatus to the cell surface has been difficult to study and the exact components involved in this step are unclear.

We have used a human membrane glycoprotein (TGN46) that is a resident of this final Golgi compartment, i.e. the *trans*-Golgi network (TGN), to probe this final step of constitutive protein secretion. This protein is expressed in all mammalian cells and localises to the TGN in each cell. However, it cycles to the cell surface by this step in constitutive

secretion, returning to the TGN via endosomes. The function of this protein is unclear although recent studies indicate that it may act as a chaperone or cargo receptor that packages secreted proteins for export to the cell surface (2). Intriguingly, this protein also appears to recruit a complex of cytoplasmic proteins including a GTPase, phosphoprotein and lipid kinase that may act to trigger the formation of novel transport intermediates that deliver secretory proteins to the cell surface (3). We have also shown that constitutive transport intermediates derived from the Golgi apparatus contain different secreted glycoproteins *en route* to the cell surface (4).

These findings point to a step in membrane traffic leaving the TGN for the cell surface; however, it cannot be discounted that there are parallel steps that transport different classes of secretory proteins. We are using TGN46 and other TGN residents as probes to identify and characterise these final steps in constitutive protein secretion. In addition to cell biological approaches including laser and confocal microscopy, projects to elucidate the three-dimensional structures of these different components are an integral part of a multidisciplinary approach to studying this event in eukaryote organisms.

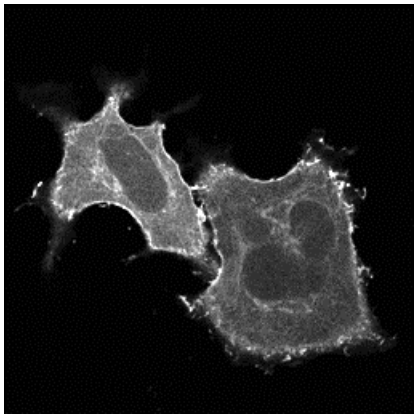


Figure 1. HeLa cells expressing a lipid-binding protein that is recruited to the plasma membrane on binding a specific phospholipid at the cytosolic face of the membrane bilayer. Note that plasma membrane ruffles in both cells are clearly defined by the localisation of this protein.

Collaborators
Prof. Sheena Radford (Leeds University)
Prof. Peter Downes (Dundee University)
Prof. Vivek Malhotra (Univ. California San Diego, USA)
Dr. Francis Barr (Max-Planck Institute, Munich, Germany)
Dr. John Walker (Leeds University)

References
[1] Lemmon MA, Ferguson KM (2000) Signal-dependent membrane targeting by pleckstrin homology (PH) domains. *Biochem J* **350:** 1-18.
[2] Wang J, Howell, KE (2000) The luminal domain of TGN38 interacts with integrin beta 1 and is involved in its trafficking. *Traffic* **1:** 713-723
[3] Ponnambalam S, Clough S, Downes CP, Lucocq JM, McLauchlan HJ, Towler MC (1999) Lipid kinases and trans-Golgi network membrane dynamics. *Biochem. Soc. Trans*., **4**: 670-673.
[4] McLauchlan HJ, James J, Lucocq JM, Ponnambalam S (2001) *Cell Biol Intl*, In press.

Funding
We wish to acknowledge the support of the MRC for this work.

Biophysical studies of β_2-microglobulin amyloid formation

David Smith, Victoria McParland, Susan Jones, Neil Kad, Sheena Radford

Introduction

Amyloid disease involves the association of protein or peptide monomers into long, non-covalently associated protein fibrils. These fibrils can be formed from a wide range of normally soluble proteins and peptide fragments *in vitro*, and about 20 different human amyloid diseases are currently known. Amyloid fibrils share a similar cross-β structure, involving repeating β-strands, which lie perpendicular to the fibril long axis.

Work in our laboratory focuses on the human disease, haemodialysis-related amyloidosis which involves the aggregation of wild-type human β_2-microglobulin (β_2m) into amyloid fibrils. β_2m forms the light chain of the cell surface class I MHC complex. The protein is shed continuously from the cell surface and is carried in the serum to the kidneys where it is catabolised and excreted. Upon kidney dysfunction β_2m is no longer cleared from the serum in this manner and, as a consequence, serum levels increase by up to 20-fold, resulting in the formation and deposition of β_2m amyloid in the joints. Like other amyloid diseases, the mechanism of β_2m amyloidosis is currently unknown. Our laboratory is currently using structural and kinetic methods to determine the conformation of amyloidogenic intermediates, the mechanism of fibril assembly, and the structure of the amyloid fibril itself. This work is essential in order to provide a molecular understanding of amyloidosis and as a prelude to deriving therapies against this, and other, amyloid diseases.

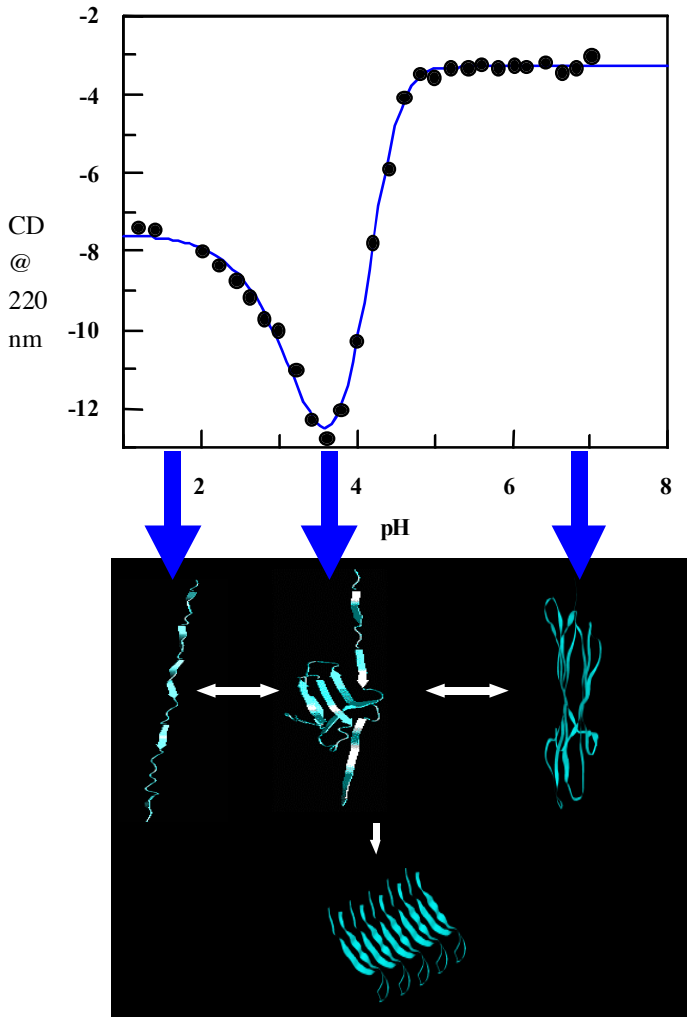


Figure 1. The effect of pH on the secondary structure of $\beta 2_m$ monitored by CD. This is also represented as a conformational change at around pH 4 leading to the population of the amyloid precursor.

Factors affecting amyloid formation

β_2m fibrils have been visualised in macrophage lysosomes suggesting that this compartment could be involved in amyloidosis *in vivo*. In accord with this hypothesis, we have shown that a reduced pH environment is crucial for the formation of β_2m amyloid fibrils *in vitro*. At neutral pH, β_2m is native, stable, and fibril formation is not seen even after incubation for months. By contrast, fibril formation is rapid below pH 5.0. Use of the amyloid-specific dye, thioflavin-T, has allowed us to determine the initial rate of fibril formation over a range of pH values. These results showed that fibril formation is rapid at pH 3.6, whereas at lower pH fewer nucleation sites develop and fibril formation is slowed. Our studies *in vitro* have also shown

that ionic strength plays a key role in fibril formation. At pH 3.6 the rate of fibrillogensis is increased markedly as the ionic strength is increased. Using negative stain electron microscopy, we have also shown that the conditions under which fibrils are grown affect their morphology. At pH 3.6 in high salt, the fibrils formed are short and curvilinear, whilst at lower pH the fibrils extend to much greater lengths.

The conformation of the monomeric amyloid precursor of β_2m

At low ionic strengths the rate of β_2m amyloidosis is very slow. Making use of this observation, we have found conditions under which the conformational properties of the monomeric amyloid precursor can be determined. Using circular dichroism, ANS binding, hydrogen exchange and ^{1}H NMR, we have shown that the amyloid precursor is partially folded in that it retains substantial β-sheet structure, lacks fixed tertiary interactions and is weakly protected from hydrogen exchange. Detailed titration of the protein using far UV CD has shown that one or more groups with an apparent pKa of 4.7 are involved in the formation of the amyloid precursor state (figure 1). Current work is aimed at extending these studies by the use of multidimensional heteronuclear NMR methods and site-directed mutagenesis to derive residue-specific information about the conformational properties of the amyloid precursor state. In addition, FTIR is being used to determine information about the conformational changes occurring during fibril assembly.

Examining fibrillar intermediates

In parallel with the studies described above, we are using atomic force microscopy and mass spectrometry to examine the structure of oligomeric species populated during fibril assembly. This will allow us to discover whether β_2m fibrillogenesis follows a nucleation dependent pathway or assembles through a ladder of intermediates. Studies of other amyloid diseases have suggested that small oligomeric amyloid precursors could be the key pathogenic agents in disease and hence it is critical to examine the occurrence and structure of these species in the search for new therapeutic agents.

Collaborators

Neil Thomson, Department of Physics and Astronomy and Astbury Centre for Structural Biology, University of Leeds.
Arnout Kalverda and Steve Homans, Astbury Centre for Structural Biology, University of Leeds.
Alison Ashcroft, Astbury Centre for Structural Biology, University of Leeds.
Sandy Davison, St. James' Hospital, Leeds.
Margie Sunde, Department of Biochemistry, University of Cambridge.
Mick Hunter and Ant Brown, British Biotech Pharmaceuticals Ltd., Oxford.

References

Aggeli, A., Bell, M., Boden. N., Harding, R., McLeish, T.C.B., Nykova, I., Radford, S.E. and Semenov, A. (2000) Exploiting protein-like self-assembly to engineer nanostructured materials. (2000), *The Biochemist* **22**, 10-15.

McParland V.J., Kad, N.M., Kalverda, A.P., Brown, A., Kirwin-Jones, P., Hunter, M.G., Sunde, M., Radford, S.E. (2000) Partially unfolded states of β_2-microglobulin and amyloid formation *in vitro*. *Biochemistry* **39**, 8735-8746.

Kad, N.M. and Radford, S.E. (2000) Partial unfolding as a precursor to amyloidosis: A discussion of the occurrence, role and implications. *Frontiers in Molecular Biology*. Ed. P. Lund, Oxford University Press. In press.

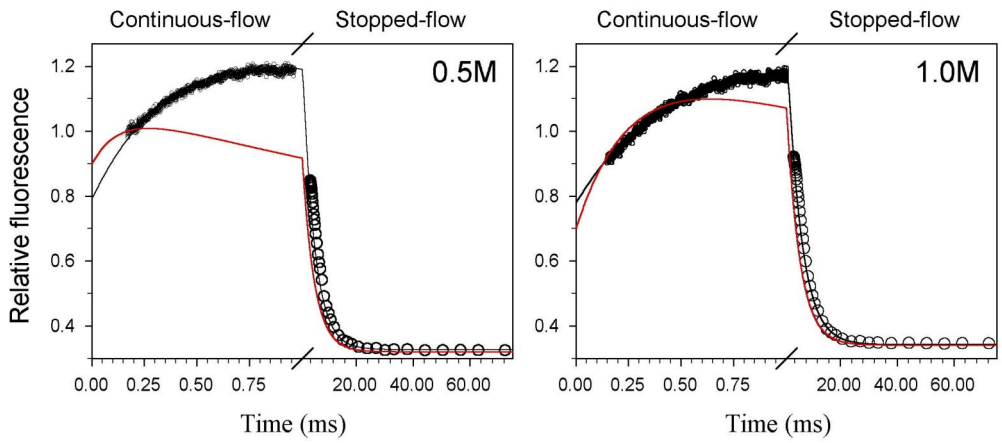
Funding
We thank British Biotech Pharmaceuticals Ltd, BBSRC, EPSRC and the Wellcome Trust for funding.

Searching for the native structure

Andrew Capaldi, Neil Ferguson, Claire Friel, Stan Gorski, Graham Spence, Sheena Radford

Introduction

Most single domain proteins can fold spontaneously *in vitro* from a highly disordered unfolded state into a biologically active native state within a few seconds. Therefore, the chemical properties of protein sequences are evolved to allow a particular structure to be rapidly selected from among the astronomical number available to them. The goal of our research is to understand these properties and how they control the sequence of events involved in protein folding. In order to address these issues we have been analyzing the folding and unfolding kinetics of four homologous small all-helical proteins (Im2, Im7, Im8 and Im9). Using stopped-flow fluorescence and CD we have shown previously that all of the Im proteins fold very rapidly (300-2000s^{-1} at 10°C) and that this process is limited by a transition state that is greater than 90% compact. Interestingly, we have found that Im7 folds through a stable intermediate while the other proteins do not.



Fluorescence changes during the folding of Im7. Data on the microsecond time-scale were acquired using a novel microcapillary continuous-flow rapid mixing device equipped with a UV sensitive CCD camera (built by Shastry and Roder). These data were then normalised and combined with stopped-flow data on a millisecond timescale to cover the entire time-course of Im7 folding. The combined data at a wide range of urea concentrations (two representative concentrations are shown as labelled) were then fitted simultaneously to a variety of three-state kinetic models. The black lines show the time-course predicted from the best fit to an on-pathway model (see text). The red lines show the time-course predicted from the best fit to the off-pathway model.

Role of intermediates in folding

We are performing a wide variety of experiments to determine whether the intermediate formed during the folding of Im7 is simply a misfolded species that forms faster than the native state or whether it is an obligate species populated during the search process. The first step towards answering this question was to determine whether the intermediate

ensemble is on (U⇔I⇔N) or off the folding pathway (I⇔U⇔N). This has been a long-standing question in protein folding because virtually all available kinetic and equilibrium data can be described equally well by either mechanism. By acquiring data on the tens of microsecond time-scale using a new microcapillary mixing device and modelling the resulting data together with the results from stopped-flow fluorescence (see figure) we were able to show that the intermediate ensemble formed during the folding of Im7 is on the folding pathway. Using this kinetic model we are now able to fit folding/unfolding data for this protein and determine the values of all the individual microscopic rate constants in the transition. Currently we are using this approach to accurately determine the effect of mutations on the folding process and hence which side-chains are involved in the different stages of folding. A similar approach is being taken to analyse the folding of Im9. In this way we will be able to compare the folding of these two homologues at the level of the individual residue. In addition, we are using both equilibrium and kinetic hydrogen exchange methods to determine which hydrogen bonds are formed in the intermediate ensemble.

Role of sequence in folding

Other studies under way have focused on determining why Im7 folds through an intermediate and other immunity proteins do not. One possibility is that all of the immunity proteins populate intermediate ensembles but that they are not stable enough to become significantly populated. This hypothesis is being tested by analysing the folding process under stabilising conditions. We are also investigating which sequence changes cause the different folding behaviour of the immunity proteins. Here we have been analysing both Im2/Im9 and Im7/Im9 chimeras. Our study of the Im2/Im9 chimeras has revealed that there is a surprising plasticity to the folding process and that sequence changes at only a few positions can result in the formation of an intermediate or in a gross change in the nature of the rate limiting transition state ensemble.

Collaborators

Ramachandra Shastry and Heinrich Roder, Fox Chase Cancer Center, Philidelphia PA
Geoff Moore and Colin Kleanthous, University of East Anglia

References

Capaldi, A. P., Shastry, M. C. R., Kleanthous, C., Roder, H., Radford, S. E. (2001) Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nature Structural Biology*, **8:** 68-72.

Ferguson, N., Li, W., Capaldi, A. P., Kleanthous, C., Radford, S. E. (2001) Using chimeric immunity proteins to explore the energy landscape for alpha-helical protein folding. *J. Mol. Biol.* **307:** 393-405

Funding

We acknowledge support from the BBSRC, the MRC, and the Wellcome Trust.

Single molecule techniques in protein folding

Anthony Blake, Chris Gell, David Brockwell, John Clarkson, Godfrey Beddard, John Trinick, Sheena Radford, Alastair Smith.

Introduction

Recent advances in instrumentation have allowed single molecules to be both manipulated in a controlled fashion and their properties to be quantified. Laser tweezers and the atomic force microscope (AFM) have been used previously to unfold proteins by the application of force by the separation of their N and C termini. Upon returning to their original position the protein refolds to the native state. Mechanical unfolding therefore allows proteins' response to mechanical stress to be studied at a single molecule level. This process has been used extensively to study various constructs of the giant muscle protein titin and other proteins which have been selected for the ability to withstand shear forces. This process should also be useful in the field of protein folding *i.e.* how an amino-acid chain selects a single native state over the millions of conformations available to it in space. The study on the single molecule level should enable details to be elucidated which are usually masked in ensemble studies, show rare events and give an idea of the range of structures that make up the denatured, native and intermediate (if any) states.

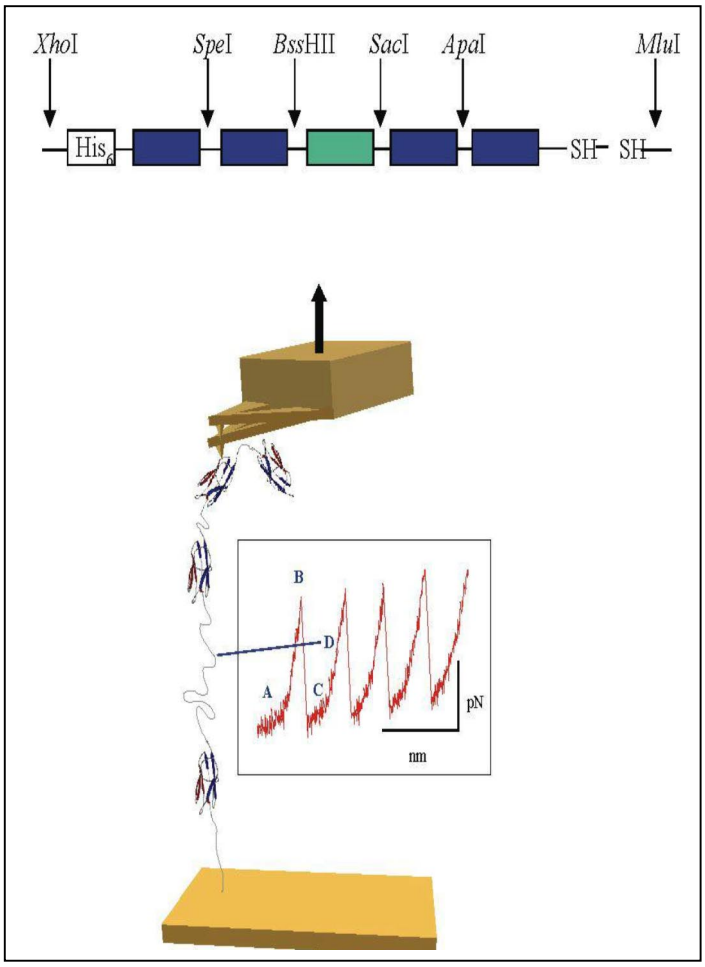


Figure 1
Above: Cartoon showing the design of the concatamer, at the DNA level. Arrows show the positions of unique restriction sites which define the cassettes. The hexa-histidine tag allows one step purification. Two C-terminal cysteines allows for attachment of proteins to a gold surface.
Below: Cartoon showing stepwise single molecule unfolding of an I27 concatamer. Proteins are non-specifically adsorbed onto the cantilever (top); upon retraction from the gold surface (bottom) the cantilever bends (point A) and the force is measured until one domain unfolds (point B) and the sudden increase in length reduces the measured force (point C). Further retraction of the cantilever causes the unfolded domain to straighten causing the force to gradually increase (point D) up to a point where the next domain unfolds.

Using the FRET technique, single molecule fluorescence can be used as a 'molecular ruler' allowing the separation between two extrinsic fluorophores to be used as a reaction co-ordinate which can monitor the folding of a protein molecule in real time, and deviations from the native state.

Our aim is to study protein folding by using both mechanical unfolding by AFM and single molecule fluorescence and then ultimately to join the experiments.

Mechanical unfolding
A ‘concatamer’ consisting of five copies of a single domain from titin (I27, a β-sandwich protein belonging to the Ig family) has been engineered (see figure 1). The cassette strategy used allows each copy to be easily substituted with a mutated I27 or even a completely unrelated protein. This modular approach therefore allows us to ask a range of different questions about proteins' response to mechanical stress. Firstly we have generated a concatamer consisting entirely of destabilised domains. By measuring the unfolding force dependence on cantilever retraction rate, information can be obtained on the unfolding kinetics and the unfolding transition state placement. Comparison of this data with the same parameters derived from traditional chemical denaturation studies will show whether the two techniques give information on the same unfolding pathway. The second phase of this study is to insert proteins into the concatamer which have not evolved to withstand mechanical stress. Repeating the same process as described above should reveal whether all proteins behave in a similar manner to I27.

The second area of research involves pulling elastomeric proteins such as elastin, byssus, dragline and the PEVK region from titin. These proteins store energy when deformed without rupture then recoil back to their original state. The usefulness of this approach is being assessed by initially studying the PEVK region from cardiac titin. This natively unfolded protein has been cloned into the concatamer and mechanical unfolding experiments in which the ionic strength of the buffer is changed are underway.

Finally, by attaching different protein domains at each end of the concatamer, which are differentially post-translationally modified in a unique manner, we hope to develop a surface attachment strategy that allows specific, directional tight binding of proteins to surfaces.

Funding
We thank the University of Leeds and the BBSRC for funding this research.

Fast folding kinetics of apomyoglobin initiated by a nanosecond temperature jump

George Dimitriadis, Adam Drysdale, Maria Healy, Alastair Smith, Sheena Radford

Introduction

In the past decade, experiments on protein folding have managed to probe the submillisecond regime, thus showing that some proteins fold, not directly, but through intermediate forms. In our laboratory we have been studying the fast folding kinetics of apomyoglobin (myoglobin without the heme) using tryptophan fluorescence as a probe of conformation following a nanosecond temperature jump.

Experimental

The apparatus used in the experiment can be seen in figure 1. The two counter propagating IR beams at 1.5 μm are produced by Raman shifting the output of a Q switched YAG laser at 1.064 μm in methane. These beams are steered through the centre of a 500 μm pathlength cuvette and uniform heating of the aqueous solution is achieved within the period of the IR pulse (~5 ns). Temperature jumps of ~20°C can be achieved. The UV beam that excites Trp fluorescence is produced by frequency doubling the output of a cavity dumped synchronously pumped dye laser using Rhodamine 6G. The resulting train of UV pulses at 280 nm are approximately 10 ps FWHM and occur at a controllable repetition rate with a maximum value of 38 MHz. Detection of the protein fluorescence is performed either with a fast photo-multiplier tube (Hamamatsu R5600) for time resolved analysis or a slower PMT (Hamamatsu, R928) for slower intensity measurements. These signals are digitised at a maximum rate of 2 Gs/s by an oscilloscope (LeCroy LC548AXL). The sample temperature is controlled by a water cooled peltier device.

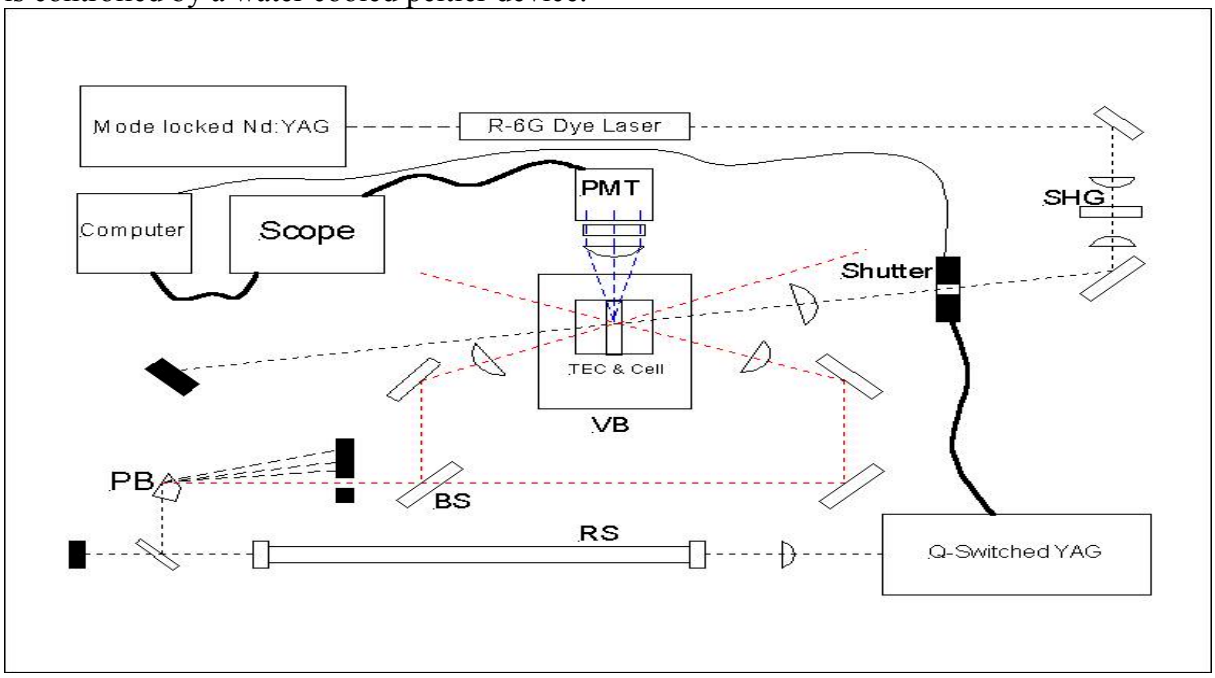


Figure 1 Diagrammatic showing the T-jump apparatus

Results

The apparatus has been tested using a 3mg/ml apomyoglobin solution in 10mM sodium acetate at pH 5.5. Under these conditions CD measurements indicate that the protein starts

to cold denature at 3°C and has a 50% denaturation at -8°C. Preliminary data have resulted in the calculation of a fast exponential phase after the T-jump with a lifetime of 18.9 μs. These data agree with a similar study (Gruebele M., Sabelko J., & Ervin J. (1998) *Accounts of Chemical Research* **31***, 699-707*). These preliminary data suggest the existence of a fast folding intermediate that seems to have the characteristics of a molten globule. Such an intermediate form can play a crucial role in the creation of the final native form of the protein.

Discussion

The data that the T-jump apparatus has up to now produced show that it is an experiment, which can probe successfully the first couple of milliseconds of the protein folding with a resolution of a few tens of nanoseconds. Such information can prove to be extremely useful in forming a better picture of protein folding. T-jump experiments have already been used to show that at least some proteins, like apomyoglobin, fold through complex pathways that involve one, or sometimes more, intermediates. Such data have changed the long held view of proteins folding in a switch off – switch on fashion, changing from the unfolded to the folded state when the conditions allow. Now it is clear that a significant number of proteins can follow complex kinetics with important steps lying in the sub-millisecond regime. The T-jump experiment can become a versatile tool for exploring the folding pathways of proteins in that time regime. Both time resolved experiments that follow kinetics and equilibrium perturbation experiments that allow the recognition of different folding intermediates under different conditions can be performed.

Future work

Our preliminary experiments have revealed a number of difficulties with the technique and the immediate goals are to develop a more robust T-jump experiment. Over the next year, these improvements will be evaluated using apomyoglobin and in the longer term we intend to apply the instrument to equilibrium perturbation experiments on a number of protein systems under study in the Astbury Centre.

Collaborators

Sheena Radford

Funding

We acknowledge the support of BBSRC

Single molecule spectroscopy to probe folding of individual proteins

Chris Gell, David Brockwell, Godfrey Beddard, Sheena Radford & Alastair Smith

Introduction

Spectroscopic measurements of concentrated bulk samples provide only an ensemble-averaged measure of a particular molecular property which may contain a contribution from a variety of different conformers and local environments within the sample. In order to study the heterogeneity of systems to reveal rare species such as protein folding intermediates, it is necessary to remove ensemble effects by performing experiments on individual molecules. We have constructed and characterised an ultra-sensitive inverted fluorescence confocal microscope to study protein folding and biomolecular interactions at the single molecule level. Our initial approach has been to determine whether the dynamics of a single protein can be followed by measurement of the diffusion coefficient which is a reflection of the overall size of the protein. Thus we have performed extensive studies of the diffusion in free solution of small dye molecules as model systems.

Experimental methods and data analysis

Figure 1 is a schematic of the apparatus. Briefly, the system comprises an argon ion laser (Coherent Innova, Coherent Lasers, UK) as the excitation source, a 140x, 1.3 NA oil immersion objective (Vickers UK), a dichroic mirror to steer the excitation beam into the objective and a single photon counting avalanche photodiode (EG&G SPD 200, EG&G UK).

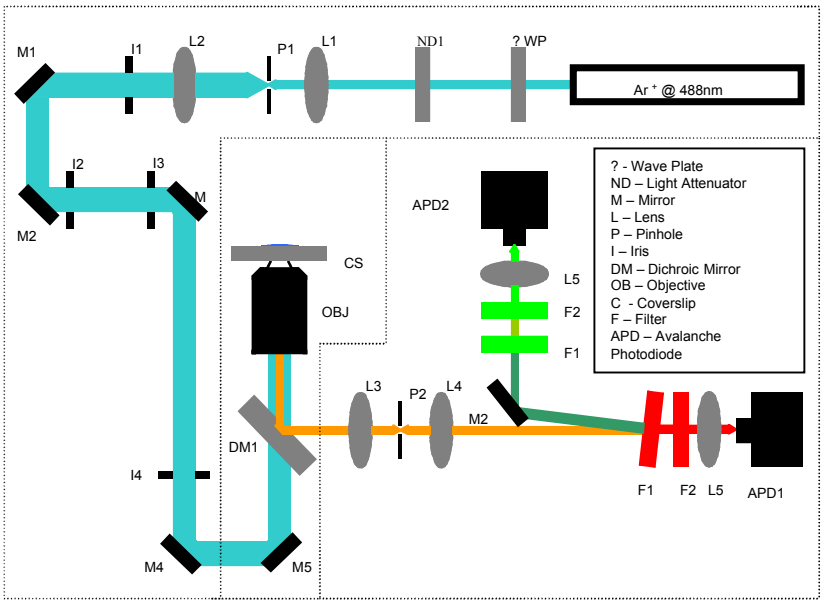


Figure 1 Experimental arrangement

Figure 2 illustrates a typical single molecule data set. An ultra-dilute sample of the organic dye BODIPY-Fluorescein diffuses through a volume defined by the tightly focused laser beam. A sample concentration of 0.1 nM or less and a small measurement volume (~0.1 fl) ensures that only single molecule events are likely to be detected. The bursts of fluorescence photons that occur when a molecule diffuses through the beam are either analysed in real time using a hardware autocorrelator (ALV, Germany) or are collected using a multichannel scaler (MCD2 FAST, ComTec, UK) and the autocorrelation function is calculated off-line

in software using Igor Pro (Wavemetrics, Lake Oswego, OR, USA). The routine used to produce the autocorrelation curves is available at:
http://www.mpi.leeds.ac.uk/faculty/smith/sms/html.

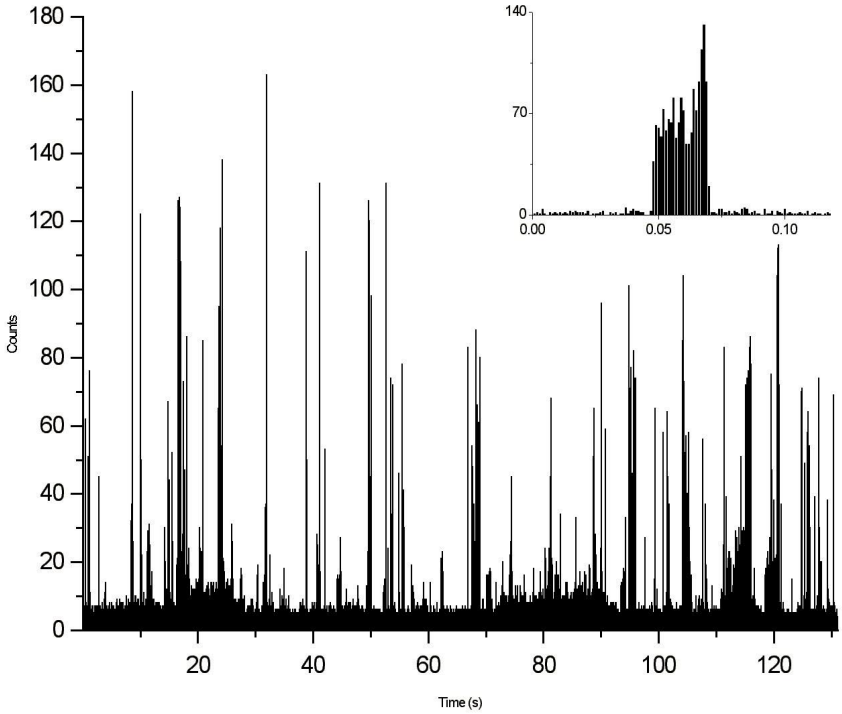


Figure 2 Single molecule photon burst events detected as individual dye molecules diffuse through the focused laser beam.

Autocorrelation analysis probes the fluctuating part of a signal, in this case the photon bursts, rather than the average. It is a powerful tool that has been used for many years in dynamic light scattering experiments and in fluorescence correlation spectroscopy. The temporal decay of the calculated autocorrelation function for a data set describes the persistence of the fluctuating events and a model has been proposed by Aragon and Pecora (S. R. Aragon and R. Pecora, (1976) *J. Chem. Phys*. 64, 1791) that allows the diffusion coefficient of the molecule to be determined from this temporal decay if the geometry of the focused laser beam is known.

Figure 3 gives three examples of autocorrelation curves for single molecule data sets for the diffusion of BODIPY Fluorescein in 0, 5 and 20 % glycerol. Fits to the autocorrelation curves of an analytical model that takes into account the physical situation producing the decay in the autocorrelation function results in a single number – the characteristic diffusion time, τ_d, of the molecules. This value is defined as the time each molecule takes to diffuse a distance equal to the length of the short axis of the sample volume (the sample volume being modelled as a cylinder). A clear increase in the diffusion time for BODIPY-Fluorescein can be seen with increasing solvent glycerol content, the increase fits the expected values well.

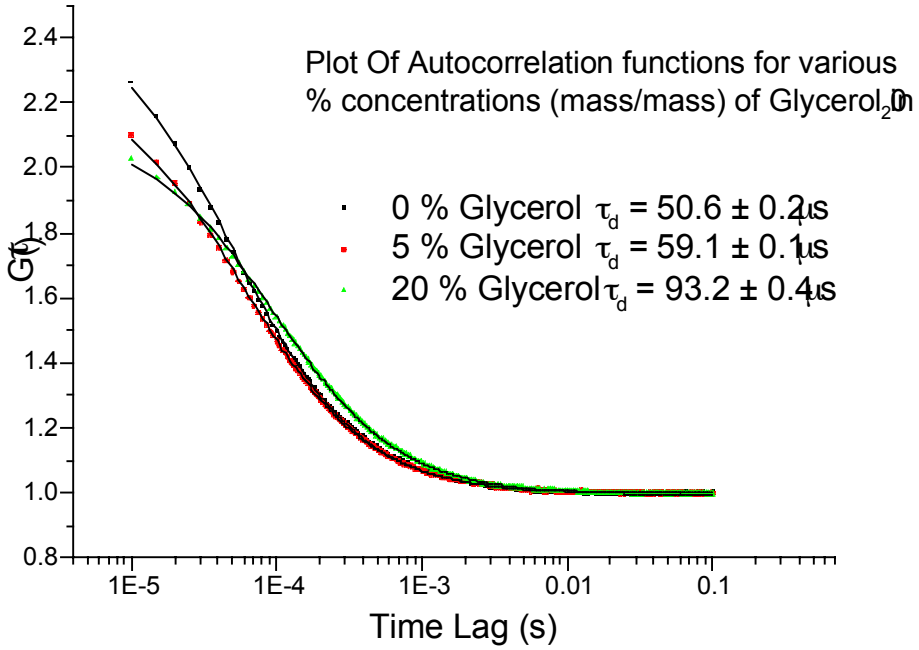


Figure 3 Autocorrelation functions calculated from the photon burst data of single BODIPY-Fluorescein molecules diffusing through the focused volume in three different concentrations of glycerol in water. Fits to the data of an analytical model (solid lines) give the characteristic diffusion time for that curve.

There are conflicting reports of the values of τ_d for rhodamine 6G (the only dye system that is reported in the literature so far) that vary over two orders o magnitude. We have shown that these discrepancies are due to the erroneous use of long sampling times to increase the signal to background and the inclusion of the zero lag time data point in software generated autocorrelation curves (Gell et al, submitted to *Appl. Phys. Lett.*).

Work in progress

Our next goal is to demonstrate whether the conformational changes of an individual protein can be observed as a change in the measured residence time t_d in our experiment. There have been two reports of fluorescence resonance energy transfer (FRET) of individual protein molecules used to monitor conformational changes (Deniz, A.A., Laurence, T.A., Beligere, G.S., Dahan, M., Martin, A., Chemla, D.S., Dawson, P.E., Schultz, P.G. & Weiss, S. (2000) Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA*. 97: 5179-5184; Talaga, D.S., Lau,W.L., Roder, H., Tang, J., Jia, Y., DeGrado, W.F. & Hochstrasser, R.M. (2000) Dynamics and folding of single two-stranded coiled-coil peptides studied by fluorescent energy transfer confocal microscopy. *Proc. Natl. Acad. Sci. USA*. 97: 13021-13026). However, double labelling with extrinsic dyes is not straightforward in general and often has very low yield. (Although one can argue that only one molecule is actually needed!) It would therefore be of more use if a single extrinsic dye label could be used as a probe of conformation via diffusion rate. Presently, it is not possible to perform these experiments using intrinsic tryptophan fluorescence because of the low photostability of tryptophan. However, it is a long term target to realise methods of performing single molecule spectroscopy using intrinsic fluorescence as the probe of protein conformation.

Funding

We acknowledge the support of the Wellcome Trust for this research.

Reflection-selection of RNA aptamers

Chris Adams, Roma Rambaran, Andrea Coates, David Bunka, James Strachan
and Peter Stockley

RNA aptamers are being selected against a number of small peptide targets composed of D-amino acid chains. The peptides are known to be able to form amyloid fibres in vitro, e.g. the Aβ peptide involved in Alzheimer's disease. Tight binding RNA aptamers can then be converted into molecules that bind the natural L-amino acid ligands by conversion into the corresponding enantiomer based on L-ribose sugars, reflection-selection. In order to speed up the selection process we have installed a Biomek robot allowing automated selection of nucleic acid ligands following the protocols described by Ellington and colleagues. (see also http://www.bmb.leeds.ac.uk/enabling_technology/).

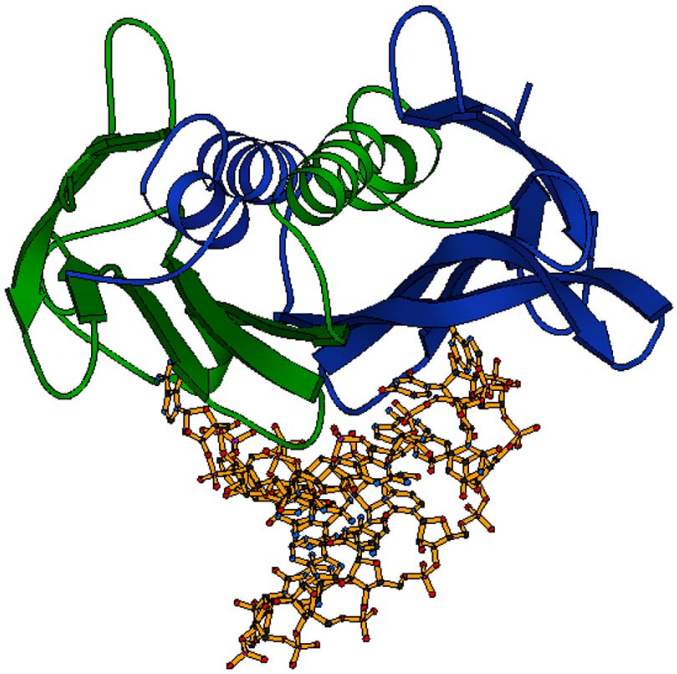
Collaborators
Andy Ellington (Texas)

Funding
We acknowledge the support of the Wellcome Trust and the MRC

RNA-protein recognition/virus assembly

Hugo Lago, Tim Moss, Andrew M. Parrott, Ben Wittaker, Jenny Baker, Andy Baron, Alison Ashcroft and Peter Stockley

The MS2 translational repression complex, that also serves as an assembly initiation complex, continues to provide unique insights into the details of sequence-specific RNA-protein recognition. In the past year, we have shown that a series of RNA aptamers, that mimic the natural operator binding but have distinct tertiary structures, have similar solution behaviour, as judged by the fluorescence emission spectra of RNAs modified with 2-aminopurine. That is, aptamers seem to be mimicking solution dynamic behaviour even when not liganded by the coat protein.



Ribbon diagram of a coat protein dimer bound to operator RNA. The RNA can be seen as a crescent, bound across the protein dimer.

Fluorescence spectroscopy, this time using stopped-flow kinetics, has also been used to probe the early events occurring when a coat protein dimer interacts with the RNA operator. Unexpectedly, it appears that a fraction of the coat protein molecules are unable to bind RNA even though the interaction is essentially diffusion driven, until they have undergone a conformational change. This is the first evidence linking RNA-binding with protein conformation and may provide insights into the mechanism of phage capsid self-assembly. To probe these secondary events we have been using Q-tof mass spectrometry to characterise higher order protein-RNA complexes. These studies are still on–going but it is clear that the Q-tof will be a major tool in analysing such assembly events.

Structure determination of an RNA chemically variant complex, in association with our longstanding collaborators in Sweden, provides strong evidence that understanding this important class of interactions will require extensive knowledge of both liganded and unliganded structures, since removal of a single hydrogen bonding group from a single base in the RNA results in major rearrangement of the RNA-protein interface. Electrostatic potential calculations of the effect of the chemical change suggest that the result is driven by the generation of unfavourable contacts in the unliganded RNA, and implies that caution will be needed in future in interpreting the results of such chemical variation experiments.

Collaborators
Nicola Stonehouse (LINK)
Simon Phillips (LINK)
Stuart Warriner (LINK)
David Peabody (New Mexico)
Andy Ellington (Texas)
Darrell Davis (Arizona)
Leo Beigelmann (Colorado)
Lars Liljas (Uppsala)

Publications
Parrott, A.M., Lago, H., Adams, C.J., Ashcroft, A.E., Stonehouse, N.J. and Stockley, P.G. (2000). "RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour." *Nuc. Acids Res.* **28**, 489-497.

Lago, H., Parrott, A.M., Moss, T., Stonehouse, N.J. and Stockley, P.G. (2001). "Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: Identification of a protein conformer unable to bind RNA." *J.Mol.Biol* **305**, 1131-1144.

Grahn, E., Stonehouse, N.J., Adams, C.J., Fridborg, K., Beigelman, L., Matulic-Adamic, J., Warriner, S.L., Stockley, P.G. and Liljas, L. (2000) "Deletion of a single hydrogen bonding atom from the MS2 RNA operator leads to dramatic rearrangements at the RNA-coat protein interface." *NAR* **28,** 4611-4616.

Funding
We acknowledge the support of the BBSRC and Leverhulme.

Transcriptional control

Sipra Deb, Ferenc Marincs, Phil Bardelang, Paul Beale, Jenny Baker, Kenny McDowall, Alastair Smith

Prokaryotes

We have shown that the *Bacillus subtilis* homologue of sigma54, the major variant sigma factor in *E.coli*, has considerable cross-functioning with the *E.coli* protein, being able to form functional holoenzyme with the core polymerase subunits and activate from sigma54 promoters in response to *E.coli* activator proteins. The *B.subtilis* protein, sigmaL, is also able to undergo isomerisation leading to DNA melting in response to activator when bound to a fork junction mimic. A detailed analysis of the thermodynamics of promoter-sigma interactions has been carried out for the first time using the technique of surface plasmon resonance. This has confirmed and quantitated the role of thymidine methyl groups in sigma factor-DNA affinity. We are now using fluorescence techniques to study the dynamics of sigma-DNA interaction; interaction with the activator and subsequent isomerisation.

An *E.coli* gene array covering all the ORFs is being constructed and will be used to investigate the *in vivo* consequences of detailed transcription factor functions identified and studied *in vitro*. This approach will be used to investigate co-operative DNA binding, sequence-specific recognition etc. in model systems such as MetJ, ArgR and TrpR. We will also study the effects of altering RNA degradation pathways using modified components such as RNAseE.

Eukaryotes

We have developed a novel SPR-based assay for studying the kinetics of RNA polymerase movement through tethered mononucleosomes. This has been used to show that a) the presence of histone core octamers reduces polymerase transit rates by roughly a factor of two, and b) that acetylation of core histones reduced the level of this general repression making the template more like free DNA. Reconstitution experiments with individual histone subunits is being used to investigate detailed effects of histone/DNA structure and modification on the rate of polymerase passage.

Funding

We acknowledge the support of BBSRC and the Wellcome Trust

Molecular interactions in the assembly of bacteriophage ϕ29

Stephanie Fonseca, Jonathan Wood and Nicola J. Stonehouse

Introduction

During assembly, a virus must recognise and package its genome with complete fidelity. The *Bacillus subtilis* bacteriophage ϕ29 shares a common assembly strategy with other dsDNA viruses including adenovirus and herpesvirus, and is a model system for understanding the DNA packaging process in this group of viruses. The DNA is pumped into a preformed icosahedral shell or procapsid, ATP hydrolysis providing the necessary energy. The ϕ29 packaging motor is an RNA-protein complex and consists of the connector particle through which the DNA is translocated, an essential RNA component termed packaging-RNA or pRNA, and an ATPase. pRNA, which binds the connector, has been shown to self-assemble into a higher order multimeric structure.

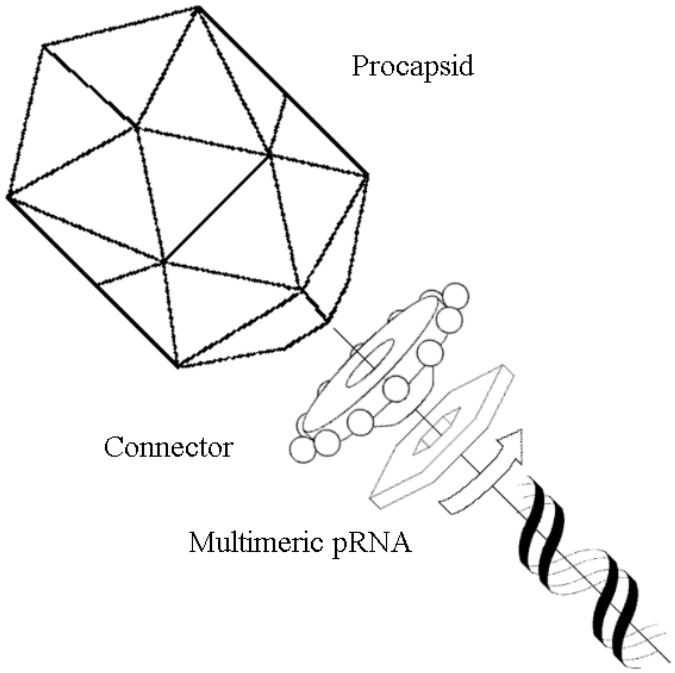


Figure 1

Schematic diagram showing the components of the ϕ29 DNA packaging motor

The self assembly of pRNA

Intermolecular base pairs between two loop regions in the secondary structure of 120 nucleotide (nt) pRNA form a multimeric RNA assembly in the presence of magnesium ions. This ability to form multimers is essential for DNA packaging, although different studies have suggested both active pentameric and hexameric pRNA rings. We are performing equilibrium sedimentation experiments in the analytical ultracentrifuge (AUC) with both the 120nt pRNA and a truncated 71nt molecule. Knowing the monomeric mass of pRNA, and using a buffer where the data can be analysed in terms of a simple self-association model, the multimeric states of pRNA can be examined and measures of the association constants

obtained. The effects of magnesium ion concentration are being investigated, together with the effects of short DNA sequences designed to compete in binding to the interacting loops.

Investigation of the structure and function of the connector

The connector is composed of a ring of 12 or 13 monomers of the protein gp10 (35.8kDa) with a central channel. The gene encoding the gp10 protein was inserted into a pET expression vector in order to produce protein with a His-tag for purification by metal affinity chromatography. Following purification, the protein was examined by Electron Microscopy (TEM), which showed that in the presence of Mg^{2+} and high salt the connectors aggregated to form structures termed 'rosettes'. Creation of a protease site was thus created at the gene level to remove the His tag after purification.

We have shown that non-specific RNA is an inherent part of the connector structure and *in vivo* may be involved in the association of gp10 monomers into a ring and bringing some stability to the complex. We are currently using TEM, AUC and fluorescence spectroscopy to investigate the interaction of the connector with other ϕ29 DNA packaging components.

Collaborators

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

Funding

We wish to acknowledge support from the MRC and the University of Leeds.

Replication and maintenance of Staphylococcal plasmids

Jamie Caryl, Fotis Papadopoulos, Matt Smith and Chris Thomas

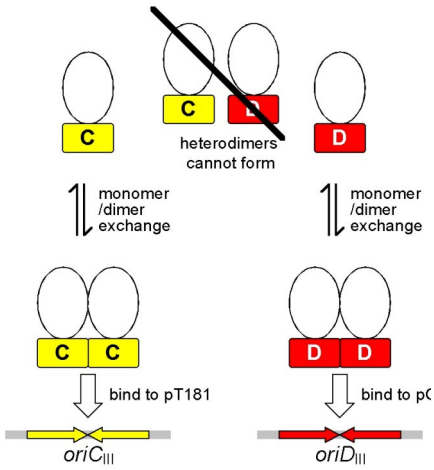
Introduction

Plasmids are small pieces of DNA resident within a bacterial cell that are both separate from the chromosome, and capable of replicating under their own control. In certain cases, plasmids are the reason why a bacterial infection may no longer respond to treatment with antibiotics: plasmids can carry genes that confer resistance on their bacterial host. Plasmids may also have the ability to transfer from their current host to a plasmid-free recipient strain, thus spreading antibiotic resistance throughout a population.

Our studies of the staphylococcal plasmid pC221 are aimed at understanding how the replication and mobilisation processes are initiated by plasmid-specified genes, and which cellular enzymes are also required in the process. To do this we employ a range of structure/function studies following the interaction of isolated enzymes with their respective target DNA. In close collaboration with Prof. Simon Phillips, we are also working on the co-crystallisation of these proteins with their targets in order to determine their three-dimensional structures.

Replication: RepD

RepD is the replicative initiator protein specified by pC221. It triggers replication by a "rolling circle" process, in which RepD specifically cleaves the plasmid at a defined origin of replication, forming a covalent protein-DNA bond in the process. Our understanding of the biochemistry of this process is through analysis of this topoisomerase-like reaction using a variety of specific mutants of RepD. These mutations help to identify the residues which constitute the active site of RepD, and their conservation in the plasmid database indicates that many previously uncharacterised reading frames are in fact related to RepD.



This year we have also begun to define the residues critical for maintaining RepD in the dimeric state. Unexpectedly, we found that the monomeric subunits of RepD are easily exchanged in solution, posing questions over their state in the cell, their plasmid specificity and the means of regulation of replication. However, we have found that dimerisation is indeed only possible between cognate monomers, preserving independent copy number of related plasmids.

Replication: PcrA

Following initiation of replication by RepD, a means of separating the two strands of plasmid DNA is required for replication to be able to proceed. Having previously established that RepD stimulates the processivity of the PcrA helicase of *Bacillus stearothermophilus*, we have now reproduced this data with the cognate helicase of *Staphylococcus aureus*. During the course of these experiments we have amended the previously published sequence of the helicase, and find that it falls within an operon containing the likely staphylococcal DNA ligase. Despite rigorous study we are unable to

demonstrate a physical interaction between PcrA and RepD, and are consequently still investigating the mechanism of the observed enhancement.

Mobilisation: MobA, MobC and *ori*T

Although not self-transmissible, pC221 can still be mobilised to recipient strains in the presence of larger conjugative plasmids. Of the previously identified genes of pC221, MobA remains the likely candidate for the nicking enzyme which initiates the mobilisation process. Yet purified MobA shows no detectable nicking activity alone *in vitro*.

We have now established that an additional reading frame, encoding a small MobC product, is also required, and that MobA/C together show plasmid-specific nicking activity – reflecting the plasmid-specific initiation of replication demonstrated by RepD. Furthermore, we have identified the *nic* site at the origin of transfer, *ori*T, in both pC221 and a close relative, pC223. Of all the previously characterised *nic* sites, that of pC221/pC223 most closely resembles the site in the border regions of T-DNA, the tumorigenic DNA transmitted by Ti plasmids of *Agrobacterium* species to their plant hosts. Our current efforts are directed towards study of the interactions between MobA, MobC and the *nic* site.

References

This work has been presented at the UK Plasmid Biology Workshop, Birmingham, UK, and the International Symposium of Plasmid Biology 2000, Prague, Czech Republic (to be reported in the journal Plasmid during 2001).

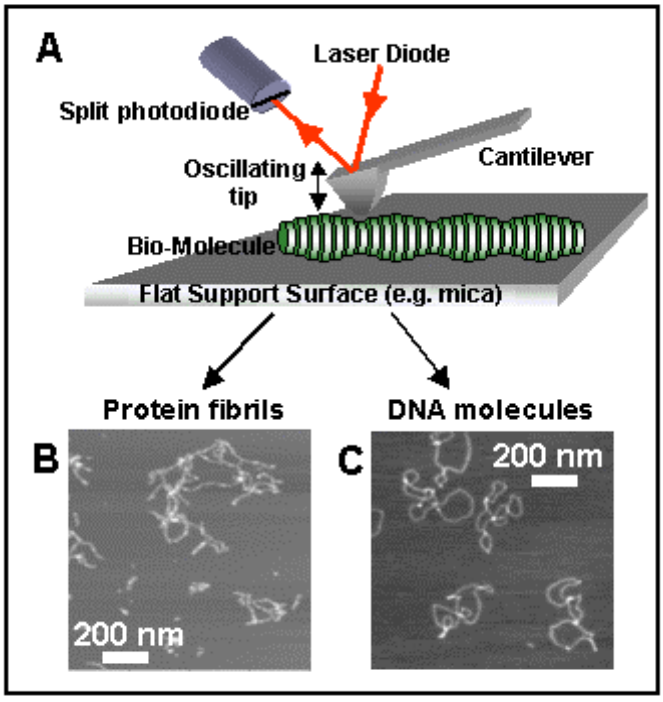
Funding

We acknowledge the support of BBSRC and Wellcome Trust

Atomic force microscopy of protein-protein and DNA-protein interactions

Neil H. Thomson

The atomic force microscope (AFM)
This is a high resolution microscopy that can be operated in vacuum, air and fluid. It works through the use of a flexible microfabricated cantilever with a very sharp tip at the end. The tip is scanned over a surface and the deflections of the cantilever are monitored using a laser beam reflected off its end onto a split photodiode detector. Line-by-line a profile of the surface is built up and stored as an image by a computer. The structure and interactions of molecules can be imaged when they are adsorbed to or bound on top of sufficiently flat surfaces, such as mica (a layered mineral crystal). Part A of the figure shows schematically how the AFM works. It is well suited to studying biological molecules in native environments because it can be operated under aqueous solutions. Summarised below are two main projects using AFM to investigate protein assembly and the processing of DNA by a protein, which controls DNA topology, known as DNA gyrase.



A) Principal of the atomic force microscope (AFM),
B) Protofibrils of β_2-microglobulin assembled in vitro at pH 1.6
C) pBR322 plasmid DNA supercoiled from reaction in the cell with DNA gyrase

Amyloid fibril formation
With Neil Kad and Sheena Radford.

Many proteins and peptides under certain conditions exist in conformational states that are amenable to aggregation. This aggregation process is complex, but leads to the formation of long, fibrillar protein assemblies with a common core secondary structure of crossed β-sheets no matter what the sequence of the protein. Known generally as amyloidosis, uncontrolled protein self-assembly is implicated in many degenerating diseases, such as Alzheimer's and Parkinson's. The low resolution structure of amyloid fibrils and their assembly is being investigated using AFM on a protein called β_2-microglobulin, which builds up in the joints of patients on long-term renal dialysis. Part B of the figure shows some typical protofibrils imaged in the AFM. These are thought to be intermediates in the formation of mature amyloid fibrils.

Mechanism of action of DNA gyrase with Jonathan Heddle and Tony Maxwell.
DNA gyrase is a motor protein in a class known as topoisomerases, which are responsible for controlling the topological properties of DNA (e.g. amount of supercoiling). DNA gyrase is unique amongst this class, because it can introduce supercoils as well as remove

them. To wind or unwind DNA it must break double-stranded DNA and pass another segment of the DNA through itself, before resealing the break. AFM is being used to investigate how this bacterial protein interacts with DNA to further understand its mechanism of action. It should give additional information about how gyrase contacts DNA and the ultimate goal is to observe gyrase supercoiling DNA in the microscope in real-time. Part C of the figure shows supercoiled DNA molecules extracted from the bacterial cell after processing by DNA gyrase.

Collaborators
Sheena Radford, Biochemistry and Molecular Biology, University of Leeds.
Tony Maxwell, John Innes Centre, Norwich.

Funding
NHT is an EPSRC advanced research fellow

Muscle group - Titin project

Larissa Tskhovrebova, Alex Liversage, Peter Knight, Lesley Wilson and John Trinick

The role of titin in muscle elasticity is being explored. Individual titin molecules were imaged by fluorescence light microscopy, also by atomic force microscopy (AFM) and electron microscopy (EM), both in compact and extended conformations. The microscopy results are consistent with a molecular elastic mechanism in titin which involves at least three phases. During the first phase the molecule straightens from its normal coiled conformation. In the second phase the small "PEVK" polypeptide unfolds. Finally, the β-structure immunoglobulin and fibronectin domains that make up the bulk of the molecule unfold.

The question of relating the single molecule force and imaging studies to the properties of titin *in situ* was examined for the first time. Although titin molecules probably operate independently through much of the I-band, near the end of the thick filament they form a hexameric aggregate, end-filaments, and near the Z-disc they bind to thin filaments. Thus the single molecule data cannot be extrapolated directly to the role of titin *in vivo*.

The question of symmetry mismatch between different regions of muscle sarcomeres was explored and the implications of this for muscle structure were explored, particularly in respect of titin. A novel model of titin anchorage in the Z-disc was proposed.

The flexibility of the titin molecule is quantitatively characterised by its persistence length, which was measured by a new type of analysis of the micrographs. The 13 nm value obtained is compatible with light scattering data but is 3-fold lower than estimates from whole sarcomere mechanics.

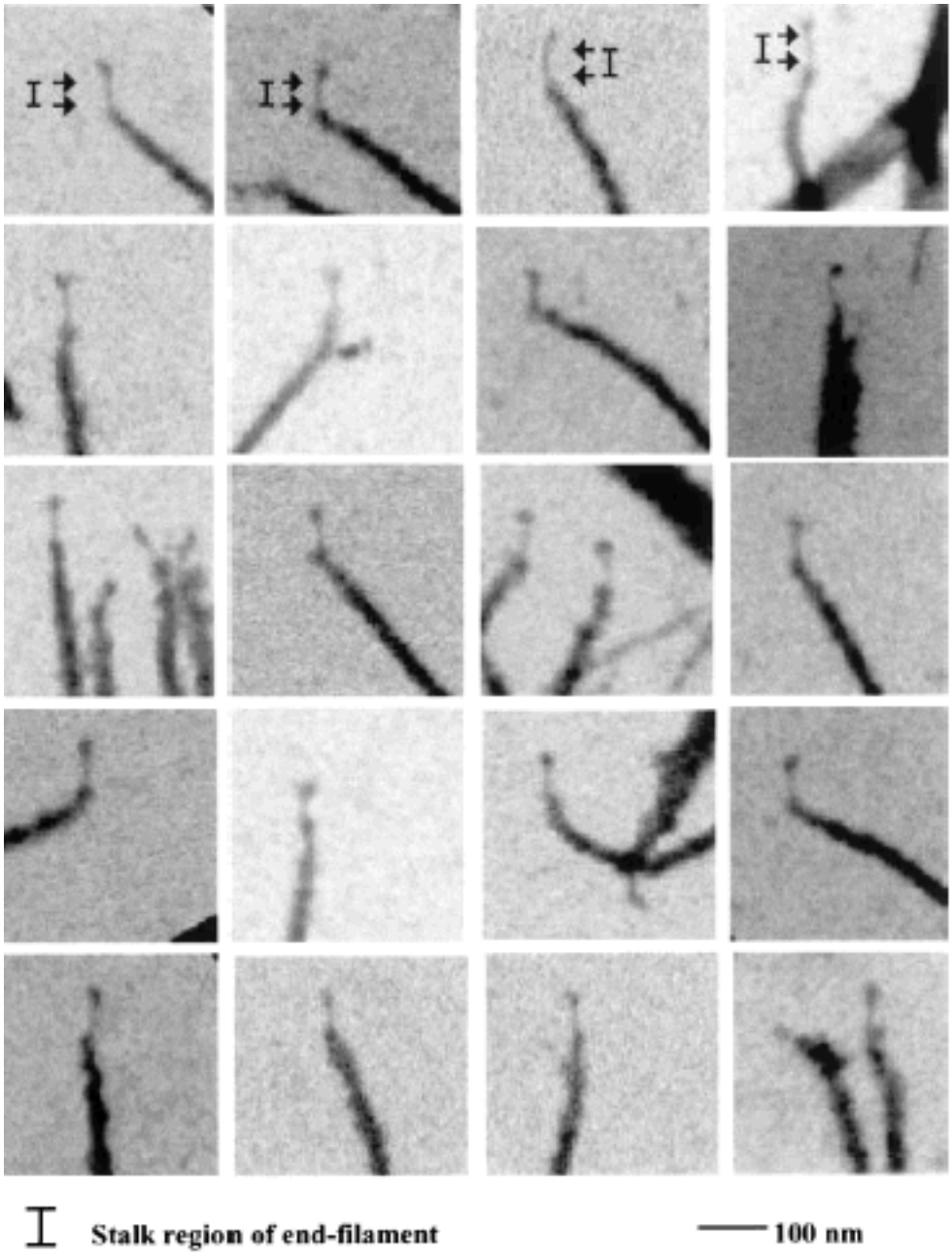
Collaborators:
Dr David Holmes (Wellcome Extracellular Matrix Centre, Manchester University).

Publications:
Tskhovrebova, L. and Trinick, J. (2000). Extensibility in the titin molecule and its relation to muscle elasticity. *Elastic filaments of the cell., Eds H Granzier and G Pollack, published by Kluwer.* 163-173

Tskhovrebova, L., Han, W. H., and Trinick, J. (2000). Studies of mechanical unfolding of single titin molecules using AFM AC mode, *Biophysical Journal,* **78,** 2627.

Funding
We acknowledge the support of the Wellcome Trust and the British Heart Foundation for this work.

Figure: STEM images of end-filaments used to make mass measurements to determine the amount of titin in muscle



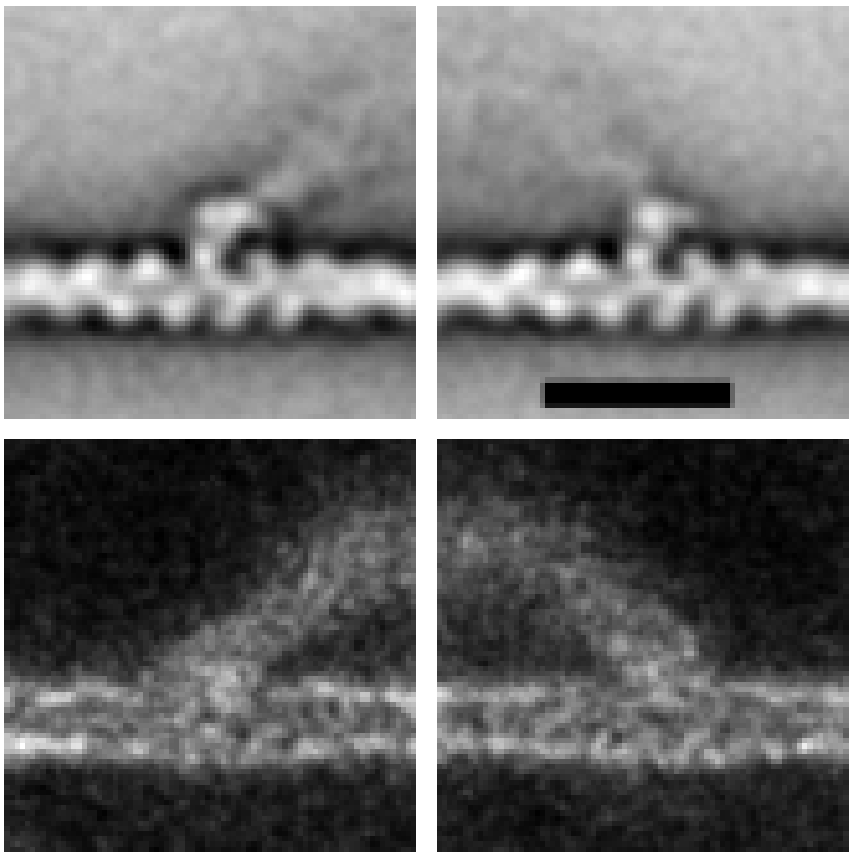
Muscle group - Myosin V project

Peter Knight, Matt Walker, Stan Burgess and John Trinick

For the past ~50 years the central problem of muscle contraction has been to understand the molecular mechanism of force production. In 1968 Hugh Huxley put forward the tilting cross-bridge hypothesis which proposes that force originates from a change in angle of the heads of the myosin molecules whilst attached to actin, fuelled by ATP hydrolysis. Still unproven, this hypothesis is widely believed, although now seems more likely that the major conformational change is within the myosin head rather than in the angle made with actin. The single most important reason why it has been difficult to test it is that, although the tightly bound states towards the of the 'power-stroke' are easy to study, the weaker states near the start of the stroke tend to dissociate at the low protein concentrations feasible *in vitro*.

We have been studying, by electron microscopy, the structures of a non-muscle isoform of myosin, myosin V, complexed with actin. Single myosin V molecules carry cargoes in a variety of cell types, including nerve, and have evolved to walk 'processively' without detachment. The weakly bound states of its power-stroke are as a consequence less easily dissociated. The heads of myosin V are also twice as long as the more commonly studied myosin II from muscle, making them easier to study by electron microscopy. Indeed, it was proposed that their large size would allow them to span the repeat of the actin filament helix and thereby walk with a straight rather than a helical trajectory. We obtained the first detailed micrographs of myosin V attached to actin by both heads, showing that the molecule can indeed span the actin helix and walk straight. Significantly, the leading head has a new attached conformation on actin. This is similar in shape to the crystal structure of myosin II heads in ADP-aluminium fluoride, which was thought to mimic the weak ADP.Pi state. Our data therefore provide strong support for the tilting bridge hypothesis. In order to improve the signal-to-noise in our data, the micrographs were subjected to a new variation of single particle image processing that treats windowed filament segments as the particles.

Figure: Single particle image processing of leading and trailing heads of myosin V attached to actin by both heads. The direction of walking is to the right. Note the bent pre-power shape of the lead heads. The variance image is shown below. Scale bar 36 nm.



Collaborators

Drs James Sellers, Fei Wang and John Hammer III (National Institutes of Health, Bethesda) and Dr Howard White (Eastern Virginia Medical School).

Publications

Walker, M. L., Burgess, S. A., Sellers, J. R., Wang, F., III, J. A. H., Trinick, J., and Knight, P. J. (2000). Two-headed Binding of a Processive Myosin to F-actin, *Nature, 405*, 804-807

Burgess, S. A., Knight, P. J., Walker, M. L., Schmitz, S., Sparrow, J. C., and Trinick, J. (2000). Real-space 3-D reconstruction of frozen-hydrated arthrin and actin filaments at 2 nm resolution, *Biophys. J., 78*, LA47

Funding:

We acknowledge the support of the BBSRC and NIH (USA) for this research.

Bioinformatics of macromolecular sequence and structure

Nikolaos Darzentas, Nicola Gold, Vidhya Krishnan, Andrew Nightingale, Howard Parish, Steven Pickering, Michael Sadowski, Amy Williams, David Westhead

Introduction

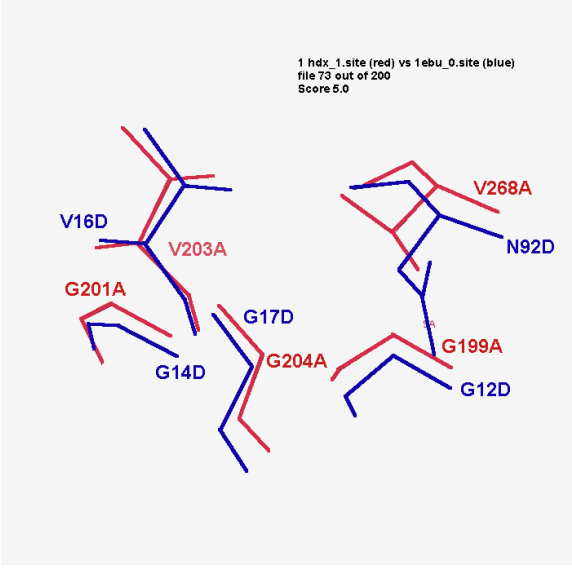
The group works predominantly in the area of biological sequence and structure analysis. The principal themes of our research are

- protein structure analysis,
- prediction of protein structure and function, and
- analysis of genome sequence data, including single nucleotide polymorphisms and conserved parts of UTRs (UnTranslated Regions).

High-throughput genome sequencing has resulted in complete genome sequences of many organisms including eubacteria, archaebacteria, simple eukaryotes and man. The focus of post-genome research is to understand these sequences in terms of biological function. Bioinformatics methods to predict function from sequence are now making key contributions to this effort, but the need for improved methods remains. At the same time, data mining tools can play a key role in the discovery of new elements of functional significance in this huge data set.

Within the protein structure theme we are interested in methods to predict protein structure from sequence and protein function from sequence and structure. Recently interest in high throughput protein structure determination and structural genomics (experimental initiatives with the broad aim of structure determination for large numbers of the proteins encoded in the sequenced genomes) has brought the problem of prediction of function from structure to the fore. Our BALSAMIC project aims to create methods and software tools with this aim. At its centre is a database of protein active sites and ligand binding sites, created by automatic methods from the database of known protein structures. The research focuses on the creation of algorithmic tools to search and cluster this database using site similarity as the main criterion. Thus prediction of function for a candidate site in a newly determined protein structure might be informed by a search of our database for a similar site whose function and/or ligand binding properties are already known. Some examples are shown in figures 1 and 2.

Figure 1. By using methods from graph theory, similarity in the residue arrangements in active sites and ligand binding sites in distantly related proteins with can be detected. Pairs of residues are matched by treating each site as a graph with edges labelled with the inter-atomic distances. The maximal common sub-graph can be found by forming the product graph and then using the algorithm of Bron and Kerbosch to enumerate its maximal cliques. The example shows a site match: the first site is coloured red, and the matched site is blue.



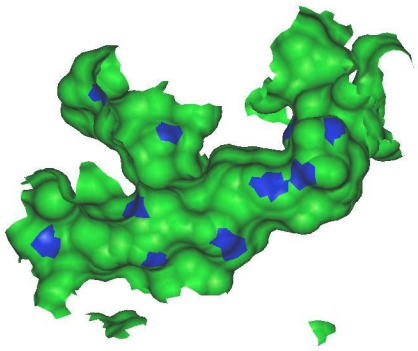
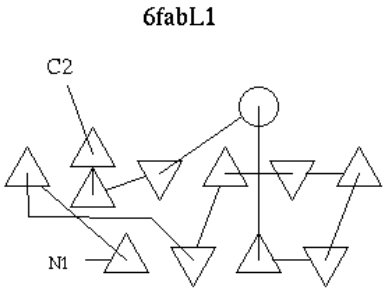


Figure 2. By modifying the graph theory algorithm slightly, common surface regions described by their geometrical and physico-chemical properties can be matched. This is important because many protein functions like catalysis and molecular recognition occur on protein surfaces.

Results (figures 1 and 2) indicate that we can successfully match a family of related ligand binding sites, and identify their common residues and surface regions. It is possible to detect similarity in sites, to describe the subtle differences in specificity that exist in proteins taken from the same family, and to detect site similarity when there is no similarity in overall structure or fold. These results indicate that our methods will be a useful complement for function prediction methods based on similarity in sequence or overall structure.

Data mining aims to discover patterns in data that were previously unknown and might have functional significance. We apply these methods to protein structures using a simplified topological description (figure 3) of the structure that enables the construction of fast algorithms, and we are also applying them to UTR sequence data with the aim of the discovery of new functional elements of nucleic acid sequences.

Figure 3. A simplified topological description of the variable domain taken from an antibody light chain. The triangles represent β strands and the circles represent helices. Strands directed out of the page are drawn as "up" triangles and strands directed into the page are drawn as "down" triangles. The fold is a β sandwich made from two anti-parallel sheets.



Collaborators
Drs. N.D Efford, A.J. Bulpitt and S. Bullock, School of Computing, Informatics Research Institute, University of Leeds.

Dr. D. Gilbert, Department of Computer Science, City University.

References

Westhead, D.R., Slidel, T.W.F., Flores, T.P.J. and Thornton, J.M. (1999). Protein structural topology: automated analysis and diagrammatic representation. *Protein Science.* **8**: 897-904.

Gilbert, D., Westhead, D., Nagano, N. and Thornton, J. (1999). "Motif-based searching in TOPS protein topology databases", *Bioinformatics* **15**: 317-326.

Funding:
We wish to acknowledge the support of the BBSRC, The Royal Society, ASTRAZeneca and GlaxoSmithKline for this work.

Astbury Seminars 2000

Monday 10th January, 2000
Prof. Michael G Rossmann (Biochemistry & Molecular Biology, Purdue University)
"Interaction of picornaviruses with their receptors"

Thursday 3rd February
Dr Sipra Deb (School of Biology, University of Leeds)
"Dynamics of Eukaryotic Transcription as monitored by Surface Plasmon Resonance"

Tuesday 18th April
Prof J Paul Grigera (Universidad Nacional de La Plata, Angentina)
"Hydration of Macromolecular Crystals: Agreement between NMR and Diffraction data"

Thursday 4th May
Mini-symposium to commemorate opening of new NMR suite by Prof Robert Kaptein (Bijvoet Centre for Biomolecular Research, Utrecht)

Thursday 18th May
Professor Alan Cooper (Department of Chemistry, University of Glasgow)
"Microcalorimetry of Protein Folding and Interactions"

Wednesday 28th June
Professor Gerald Stubbs (Vanderbilt University,Nashville)
"Tobacco Mosaic Virus at 102"

Thursday 6th July
Dr Clive Bagshaw (Department of Biochemistry, University of Leicester)
"Myosin: single molecules and single turnovers."

Friday 21st July
Prof Joachim Frank (Structural and Cell Biology, Wadsworth Centre, Albany, NY)
"The Ribosome - structure and functional binding studies using cryo-electron microscopy"

Thursday, 7th September
Dr Geoff Howlett (University of Melbourne)
"Phospholipid and macromolecular crowders control amyloid formation by human apoC-II"

Thursday, 5th October, 2000
Dr. Peter Knight (University of Leeds)
"Structural Changes that drive myosin molecules along actin filaments"

Thursday, 2nd November, 2000
Prof. Anthony L. Fink (Chemistry & Biochemistry, UCSC)
"Pesticides induce synuclein fibrillation: a possible molecular basis for Parkinson's disease"

Monday, 4th December
Prof. Ian Wilson (The Scripps Research Institute, La Jolla, California.)
"Molecular basis of MHC recognition by the T cell receptor."

Friday, 8th December, 2000
Prof. Jeremy H. Lakey (Biochemistry, University of Newcastle)
"Using bacterial outer membrane proteins in sensor applications"

We would like to acknowledge our corporate sponsors for 2000:

Belle Technology
Bruker Analytical
Jasco (UK) Ltd
Philips
X-ray Systems

http://www.astbury.leeds.ac.uk