Vance, Steven J. (2012) *The relationship between structure and function in natural surfactant proteins.*
PhD thesis.

http://theses.gla.ac.uk/3142/

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Glasgow Theses Service
http://theses.gla.ac.uk/
theses@gla.ac.uk

University of Glasgow

The Relationship between Structure and Function in Natural Surfactant Proteins

Steven J. Vance

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

School of Chemistry
College of Science and Engineering
University of Glasgow

Abstract

Surfactant activity is a property more commonly associated with small molecules than biological macromolecules. However, the significant advantages of improved biodegradability and biocompatibility that could be presented by natural surfactant proteins has elevated interest in a group of only a few proteins where intrinsic surfactant activity appears to be the primary function. Two examples of this group, ranaspumin-2 (Rsn-2) from the foam nests of the tungara frog and latherin, the surface active protein component of horse sweat, appear to be different from other surfactant proteins in the form of their activity. However, the exact molecular basis of this activity is poorly understood. This thesis describes work to rationalise surface activity and related properties in these unusual proteins.

The properties of Rsn-2 and latherin including surface activity, interaction with lipid membranes and behaviour in solution were investigated to provide further insight into the characteristics that distinguish the surfactant proteins from both conventional surfactants and other proteins. A second protein component of the foam nests, ranaspumin-1 (Rsn-1), of previously unknown function was also found to be highly surface active and is proposed to function in a similar manner to Rsn-2.

A model whereby Rsn-2 functions via a clamshell-like opening was tested through a combination of specialised NMR techniques and site-directed mutagenesis. The results identified features associated with surfactant activity, all of which were consistent with the model. The potential of Rsn-2 as a recombinant fusion partner for the production of functional surfactants or foams was proven by construction of a fluorescent conjugate.

Solution state NMR was used to determine the structure of latherin. Information on the dynamic processes taking place in the molecule were derived by analysis of NMR relaxation data. The structure revealed is a super roll fold, similar to a single domain of the BPI-like proteins. A model is proposed whereby latherin recognises the air-water interface via three dynamic, hydrophobic loops at one end of its long cylindrical structure and then unfolds to expose its hydrophobic core at the air-water interface.

Table of contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisors, Alan Cooper, Brian Smith and Malcolm Kennedy. Alan, for being a continual source of wisdom and invaluable advice, as well as providing me with the opportunity of this PhD. Thank you to Malcolm, for his insight, encouragement and most important of all, the endless supply of anecdotes, appropriate for every occasion. Brian, without whose guidance and knowledge I would have been truly lost. His teaching and patience have been integral to my growing understanding of both NMR and biology. All three deserve an extra thank you for the unenviable job of proof-reading large sections of this thesis. I have been truly privileged to work with such talented scientists. I cannot thank them enough.

Many people have contributed expertise to the work in this thesis. Thank you to Dr Sharon Kelly for her help with all things CD and for being an absolute delight to work with. Dr Guilherme Brand, whose guidance was invaluable to the ITC and synthetic vesicle work. Thanks to Dr Volodymyr Nechyporuk-Zloy for contributing his imaging skills. Sally Antill, for her work in producing and characterising the Rsn-2 mutants. To Dr Joanna Jakus and Dr Nicola Picken, for teaching me everything I know about the dark arts of molecular biology, protein expression and purification. A very big thank you to Kate Griffiths, for her advice, assistance, optimism and on several occasions preventing the lab from descending into total chaos. To Margaret Nutley for all her help with the many issues I encountered throughout my time in the group.

Thanks to all the regular office members, Ross, Khuram, Adrian and Musadiq who have been a pleasure to work alongside. All the project and visiting students that have passed through for making the lab such a fun and interesting place to work. Also to all the friends I have made throughout the department, in particular all the members of the Alchemists.

I'd also like to thank my family, in particular my parents Stewart and Ethel who have supported me throughout my extended stay in full time education. I am eternally grateful. To Ross, Jack and Katie for their encouragement and ensuring that my feet remained firmly on the ground. Thank you to Mary-Jane for supplying me with books such on bizarre subjects.

And last but not least, to Jennifer, for her kindness, patience and support.

Author’s declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Steven Vance

January 2012

Definitions/Abbreviations

AA	Amino Acid
ANS	1-Anilinonaphthalene-8-Sulphonic acid
ARIA	Ambiguous Restraints for Iterative Assignment
ATP	Adenosine Triphosphate
B_0	External Magnetic Field
BASE	Human Breast Cancer and Salivary Gland Expression
BCIP	5-Bromo-4-Chloro-3-Indolyl Phosphate
B_{eff}	Effective Magnetic Field
BPI	Bactericidal/Permeability-Increasing Protein
BSA	Bovine Serum Albumin
CATH	Class Architecture Topology Homologous Superfamily
CCPNMR	Collaborative Computing Project for NMR
CD	Circular Dichroism
cDNA	Complementary DNA
CETP	Cholesteryl Ester Transfer Protein
CIAP	Calf Intestine Alkaline Phosphatase
CNS	Crystallography and NMR System
CPMG	Carr-Purcell-Meiboom-Gill
CRD	Carbohydrate Recognition Domain
CSA	Chemical Shift Anisotropy
Da	Daltons
DANGLE	Dihedral Angles from Global Likelihood Estimates
dATP	Deoxyadenosine Triphosphate
DFT	Digest Flow Through
DMPC	Dimyristoylphosphatidylcholine
DMPG	Dimyristoylphosphatidylglycerol
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
DOPE	Discrete Optimised Potential Energy
DSC	Differential Scanning Calorimetry
DTNB	5-5’-dithio-bis-2-nitrobenzoic Acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
EK	Enterokinase
FID	Free Induction Decay
FITC	Fluorescein Isothiocyanate
FMF	Fast Model Free
FMN	Flavin Mononucleotide
FT	Fourier Transform
GFP	Green Fluorescent Protein
GNB	Gram-Negative Bacteria
HDL	High Density Lipoprotein
HRV	Human Rhinovirus
HSQC	Heteronuclear Single Quantum Coherence
INEPT	Insensitive Nuclei Enhanced by Polarisation Transfer
IPAP	In-phase Anti-phase

IPTG	Isopropyl β-D-1-Thiogalactopyranoside
IRRAS	Infrared Reflection Absorption Spectroscopy
ITC	Isothermal Titration Calorimetry
JHBP	Juvenile Hormone Binding Protein
LBP	Lipopolysaccharide Binding Protein
LDL	Low Density Lipoprotein
LOV	Light, Oxygen or Voltage Sensing
LPS	Lipopolysaccharide
MALDI-TOF	Matrix Assisted Laser Desorption/Ionization-Time of Flight
MaxEnt	Maximum Entropy
MCS	Multiple Cloning Site
MW	Molecular Weight
MWCO	Molecular Weight Cut-Off
NBT	Nitro Blue Tetrazolium
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PLTP	Phospholipid Transfer Protein
PLUNC	Palate, Lung and Nasal Epithelium Clone
PPM	Parts Per Million
RDC	Residual Dipolar Coupling
RF	Radio Frequency
RMSD	Root Mean Squared Deviation
RNA	Ribonucleic Acid
$\mathbf{R_s}$	Stokes Radius
Rsn	Ranaspumin
SCOP	Structural Classification of Proteins
SDS	Sodium Dodecyl Sulphate
SEC	Size Exclusion Chromatography
SP	Surfactant Protein
SSE	Sum Square Error
$\boldsymbol{\tau_c}$	Rotational Correlation Time
$\boldsymbol{\tau_m}$	Molecular Rotational Correlation Time
$\boldsymbol{\tau_{mix}}$	Mixing Time
$\mathbf{T_m}$	Melting Temperature
TNB	2-Nitro-5-Thiobenzoic Acid
TOCSY	Total Correlation Spectroscopy
TOP1	Takeout Protein 1
TOPO	Topoisomerase
Trx	Thioredoxin
UV	Ultraviolet
VLDL	Very Low Density Lipoprotein
WT	Wild-Type
$\boldsymbol{\chi_m}$	Magnetic Susceptibility

1 Introduction

Proteins are an immensely broad range of molecules, evolved to fulfil an equally broad number of roles throughout nature. The unique properties of proteins hold promise for potential applications. One such property is surfactant activity, more normally associated with small molecules than macromolecules. This thesis is concerned with studies on the structure and function of some examples from this relatively rare class of natural surfactant proteins: ranaspumin-2, from the foam nests of the túngara frog and latherin, the surface active protein component of horse sweat. Not only are these proteins intrinsically interesting themselves, but the biodegradability and biocompatibility that these protein surfactants offer, provides potentially significant advantages over conventional surfactants within the appropriate systems. However, for this potential to be fulfilled the molecular basis for the function must be understood. Surfactant activity is dependent upon some degree of amphiphilicity in the molecular structure that is not necessarily revealed in the amino acid sequence of these proteins and therefore only by first understanding the relationship between the structure and function of these surfactant proteins can we gain the insight required to begin to fully exploit their activity to our benefit. The lack of techniques for investigating molecular structure to high resolution at interfaces requires an indirect approach to this problem.

Various proteins such as the hydrophobins and the pulmonary surfactants have been associated with surfactant activity. More recently ranaspumin-2 and latherin have been identified as further examples of a group of only a few of proteins where intrinsic surfactant activity appears to be the primary function [1,2]. Of this group only the hydrophobins have been investigated in great detail and therefore provide the only precedent for this work. Both proteins have already differentiated themselves from the hydrophobins and therefore represent at least one new class of surfactant proteins. The aim of this work is to build on existing knowledge of molecular structure and behaviour of these proteins to produce new insight into the mode of action of this unique group of proteins.

1.1 Surfactant Proteins

1.1.1 Surfactants

Water, “the universal solvent” or the “matrix of life” has numerous unusual characteristics that differentiate it from other liquids [3]. These “anomalous” properties include: expansion upon freezing, high heat capacity, high boiling and melting points, and are all a consequence of the high affinity water molecules have for one another through hydrogen bonding [4]. Another of these properties is its high surface tension (74.2 mN m^{-1}) compared to other liquids, for example ethanol (22.3 mN m^{-1}) and acetone (23.7 mN m^{-1}) [5]. Again this property is a result of water’s high self-attraction from hydrogen bonding. Surface tension can be considered a measure of a liquid’s affinity for an immiscible phase [6]. Molecules within the bulk solvent are completely surrounded by hydrogen bonding partners whereas molecules at an interface with another material are limited in the number of interactions they can form and can therefore be considered an energetically unfavourable state. The formation of a new interface increases the number of molecules in this unfavourable state resulting in an increase in the free energy of the system as a whole. Liquid systems are therefore thermodynamically driven to minimize surface area producing a partition/interface with a minimised area in order to reduce the number of molecules at the energetically unfavourable surface environment. This force, opposing increases in surface area, is described as surface tension. The requirement of water to minimise surface energy also drives the exclusion of species that cannot satisfy its solvent-solvent interactions into a separate phase. This action is termed the hydrophobic effect.

The surface tension of water can be influenced by surface active agents (surfactants), molecules that have a characteristic amphiphilic chemical structure, containing features that have both high (hydrophilic) and low (hydrophobic) affinity for the solvent. Surfactants tend to be small organic molecules, usually consisting of a hydrophilic, polar head group and an extended hydrophobic, hydrocarbon chain (Figure 1.1). These conventional surfactants are commonly subdivided depending upon the nature of their head group as non-ionic, cationic, anionic or zwitterionic.

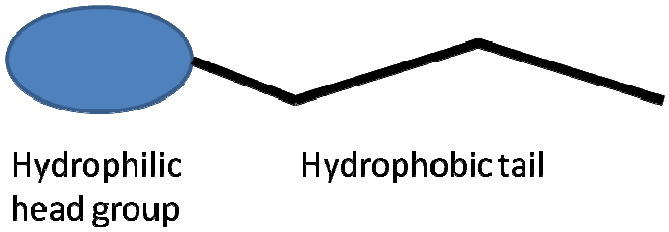


Figure 1.1- The typical molecular structure of a surfactant.

In an aqueous solution the hydrophobic effect drives the exclusion of the hydrophobic component of the surfactant. However the creation of a separate phase altogether is prevented by the presence of the hydrophilic group. This drives the adsorption of surfactant molecules to the interface or, when higher concentrations of surfactant are available, to form supramolecular structures such as micelles or bilayers. At the interface with air (or other non-polar surface) the surfactant is orientated such that the hydrophobic group is directed towards the hydrophobic phase, shielded from the water by the hydrophilic group which is directed towards the polar solvent. Now rather than a direct interaction between the two immiscible phases, each is now able to form favourable interactions with their corresponding fragment of the surfactant molecule. As a result the surface energy and hence surface tension of the system is significantly reduced.

This reduction in surface tension allows the dispersion of species that would normally be immiscible in a particular solvent. The resultant mixtures include foams (gas dispersed in liquid) and emulsions (liquid dispersed in liquid). The qualities a system must possess in order to form a foam or emulsion include: the ability to overcome the forces that drive it to minimise its surface/interfacial area and the appropriate rheological and surface properties to inhibit the coalescence of dispersed species [5]. As a consequence it is not possible to create foams or emulsions of a pure liquid as they lack the ability to fulfil these criteria. As discussed earlier, the addition of surface active species reduces the thermodynamic consequences of interface formation and also alters the surface properties thus creating a system more favourable for foam or emulsion formation. However despite the effect of surface active molecules these dispersions remain inherently unstable and often succumb to various kinetic or environmental effects.

The ability of surfactants to influence surface energy finds use in almost every industry, hence the surfactant market was worth over $23 billion in 2009 and this number is expected to continue to grow rapidly [7,8]. The most obvious application is in detergency: “the action of surfactants that causes or aids in the removal of foreign material from solid

surfaces by adsorbing at interfaces and reducing the energy needed to effect the removal" [9]. Detergents are so integral to sanitisation in body, clothes and homecare that surfactant usage is a criterion used by many analysts to determine the development of a society [10]. The food industry is also highly dependent upon the activity of surfactants with uses ranging from extraction of unwanted hydrophobic entities such as cholesterol, to the foaming or emulsification agents found in a huge number of products [11]. Some well-known examples include: the foamy head on a pint of beer stabilised by amphiphilic polypeptides and identified as a sign of quality, protein from egg white being whipped into foam to form meringue and the dispersion of egg yolk in oil by phospholipid emulsifiers to produce mayonnaise [11]. Surfactants have also found roles within medicine either directly as treatments in conditions such as respiratory disease or in the field of drug delivery [11]. These are just a few examples of the multitude of diverse applications with high practical importance and value that surfactant activity is critical to.

As well as an ability to reduce surface tension conventional surfactants also have the ability to interact with biological molecules. The hydrophobic effect drives the exclusion of surfactants into non-polar environments, which can include hydrophobic domains within globular proteins or lipid membranes. The presence of the amphiphilic surfactant disrupts stabilising hydrophobic interactions altering the physical properties of the structures. This ability to denature biological molecules is a useful technique and is commonly utilised in many areas of biochemistry, for example SDS-PAGE, cell lysis and in the solubilisation of membrane proteins [12]. On the other hand this interaction with biological molecules can also be highly undesirable in systems where native biological molecules are required. Small molecule surfactants can therefore be considered incompatible with biomolecules and are therefore not commonly encountered in biological systems, except in special instances (bile salts, for example, for dispersal of fats in the mammalian digestive system) [13].

Another disadvantage to the rapidly growing usage of surfactants in both domestic and industrial applications is concern with regards to the potential environmental implications [14]. Surfactants are routinely discarded into wastewaters where they may bio-accumulate and where they, or their degradation products, may be environmentally hazardous [14,15]. This increase in usage, greater awareness and tightening environmental regulations have provided scope for the development of a new range of efficient, biodegradable, environmentally friendly surfactants [16].

1.1.2 Biological Surfactants

New surfactants based upon biological molecules are emerging. Factors such as biodegradability, low toxicity, biocompatibility, digestibility and cheap production have all been cited as potential advantages these biological surfactants offer over chemical surfactants. These unique properties provide the potential for these molecules to replace conventional surfactants in a wide range of applications.

The term biosurfactant usually refers to small amphiphiles with low molecular weight and at least one moiety displaying biological behaviour [17]. Classes of biosurfactants include: DNA surfactants, lipopeptide surfactants and peptide surfactants. They consist of either an alkyl chain or hydrophobic peptide sequence as the hydrophobic moiety, attached to a hydrophilic biological molecule such as a peptide or DNA strand. A typical example is the acyl peptide surfactant, surfactin [18]. Surfactin is produced by the Gram-positive bacteria *Bacillus subtilis* and as well as being highly surface active (Surfactin is capable of reducing surface tension of aqueous solutions to 27 mN m^{-1} [19]) it has also been identified as having antibacterial, antitumoural and hypocholesteromic properties and inhibition towards fibrin clot formation [20]. Surfactin consists of a loop of seven amino acids, with its N and C–termini both linked to the same fatty acid chain. The chain usually consists of 13-15 carbons and provides the hydrophobic character of the molecule. The amphiphilic nature and therefore potential surface activity of surfactin and many of these biosurfactants is apparent from their structure. Their activity and properties are, in many ways similar to that of conventional surfactants.

There are also examples of protein surfactants. In these cases the source of activity is rather more obscure than that observed in small molecules. Proteins are large, complex molecules that depend upon a precise orientation of residues within their three dimensional structure for activity. This active/native protein structure, normally the most thermodynamically favoured state under physiological conditions, is defined by the range of properties and interactions of the individual amino acids that comprise the polypeptide sequence [21]. These interactions include: hydrogen bonding, electrostatic interactions, dispersion/van der Waals interactions, disulphide bridges and hydrophobic forces, that all counter the significant reduction in entropy encountered in the restriction of flexibility of the unfolded state [22]. In particular, the exclusion and congregation of aliphatic side chains into a hydrophobic core by the hydrophobic effect has been attributed as the dominant force within protein folding.

It is therefore clear that like surfactants, globular proteins contain both hydrophilic and hydrophobic characteristics, i.e. are potentially amphiphilic. However, unlike conventional surfactants, in their native form the hydrophobic entity of the protein is shielded from the solvent. Thus the amphiphilicity is not available for surface adsorption. However, the forces that stabilise the protein can be overcome to cause protein denaturation: where the amino acid sequence remains intact but the secondary and tertiary structures are compromised. Denaturation can be caused by various external forces that disrupt the stabilising forces mentioned previously, exposing hydrophobicity, resulting in protein aggregation and/or surface adsorption. In some cases, adsorption to the interface as a consequence of some degree of surface hydrophobicity in the native fold is enough to cause denaturation alone [23]. At the air water interface the proteins subsequently undergo a change in conformation, presumably such that the hydrophobic regions reorient towards the gaseous phase. This adsorption produces a reduction in surface energy which can give rise to foams and emulsions and is the basis for many techniques utilised within the food industry [24].

Despite this inherent potential of proteins to form surface active structures, there are relatively few cases of proteins in nature with the specific role of surfactant activity. There are however, some notable exceptions, the most studied of which, are the pulmonary surfactant proteins.

1.1.2.1 Pulmonary Surfactant Proteins

The pulmonary surfactant proteins are the best known group of proteins associated with surfactant activity. Pulmonary surfactant is a complex mixture of phospholipid (90%) and protein (10%) that covers the alveolar epithelium, reducing surface tension to prevent collapse upon exhalation [25]. The protein component consists of four main surfactant proteins (SP), which can be subdivided into two groups: the large hydrophilic SP-A and SP-D and the small hydrophobic SP-B and SP-C.

SP-A and SP-D are members of a family of collagen containing, C-type lectins termed ‘collectins’ and both form large oligomeric structures from numerous copies of one or two chains [26]. Oligomers of both proteins consist of N-terminal, disulphide linked, elongated triple helical collagen regions and a globular, C-terminal, carbohydrate recognition domain (CRD). These domains allow the proteins to bind multiple ligands including sugars, Ca^{2+}

and phospholipids. Via these interactions SP-A and SP-D are able to bind to the surface of pathogens to play key roles within respiratory immunity [27].

The membrane associated SP-B & C are known to have a more direct effect upon respiratory physiology. The importance of SP-B in particular is demonstrated by the onset of lethal, respiratory failure in SP-B deficient hosts [28]. Although SP-C is not critical for survival, deficiency is associated with severe respiratory disease [29]

SP-B is found as a disulphide linked dimer of two highly hydrophobic, saposin-like, 79 amino acid peptides [30]. Although no structure has yet been reported, modelling to the structure of another saposin-like homologue, NK-lysin, predicts a monomer containing five amphipathic helices, the polar faces of which are believed to interact with lipid chains [31,32].

SP-C is a 35 amino acid, valine rich (69%) protein with no significant sequence similarities. The structure has been determined by NMR in mixtures of organic solvents [33,34]. The majority of the structure consists of a rigid alpha helix, which spans the phospholipid bilayer [34,35]. The N-terminus however is unstructured and contains prolines and palmitoylated cysteines. This dynamic region is believed to be the source of the protein's ability to interact with and disrupt lipid structures [36].

The abilities of both SP-B & SP-C to promote the transfer of phospholipids between structures has been demonstrated *in vitro* [37]. This activity has been associated with numerous roles within surfactant homeostasis including: the transfer of lipid from the storage and transport structures, lamellar bodies (LB), into the air liquid interface and the rapid exchange between compressed and expanded states experienced during the respiratory cycle [33]. Research has also identified SP-B involvement in surfactant biogenesis, demonstrated by a lack of LB in SP-B deficient hosts [38,39]. Furthermore, SP-B is also required for the full processing of SP-C, perhaps going some way towards explaining the increased severity in SP-B deficiency over SP-C, despite the apparent overlap in activity [40].

It is clear that the pulmonary surfactant proteins are integral to the low surface tension required for effective respiratory function. However rather than the proteins having a direct action upon surface tension as a result of some intrinsic surface exposed amphiphilicity, they rely on an ability to order conventional surfactant molecules. The pulmonary

surfactant proteins can therefore be considered mediators rather than surfactants. There are however proteins such as the hydrophobins which are able to act as surfactants without the requirement of additional amphiphilic molecules i.e. the activity is an intrinsic property of the proteins themselves.

1.1.2.2 Hydrophobins

Hydrophobins are a group of proteins produced by filamentous fungi. The ability of these proteins to induce reductions in the surface tension of water to levels below 30 mN/m has led to the family being described as “the most surface active proteins known” [41]. These natural surfactants are proposed to fulfil numerous roles within fungal growth and reproduction. The secreted protein self-assembles into amphiphilic layers reducing the energy required for aerial hyphae or fruiting bodies to breach the air-water interface or aiding the attachment of hyphae onto hydrophobic surfaces [42,43].

Hydrophobins are reasonably small proteins of approximately 10kDa in size and are distinguishable by an unusual conserved motif: eight cysteine residues form a striking arrangement where the 2^{nd} & 3^{rd} and the 6^{th} & 7^{th} cysteines are found directly adjacent within the primary sequence, creating a single, pair, single, single, pair, single pattern [44]. Members of the family can be separated into two groups, type I and type II, characterised by their distribution of hydrophilic and hydrophobic amino acids, also observed in differences in solubility of resultant aggregates [45].

The first reported structure was a crystal structure of HFBII, a type II hydrophobin from *Trichoderma reesei* [46]. The protein forms a novel fold, consisting of two β-hairpins, which form a central β-barrel structure and one short α-helix (Figure 1.2). The cysteines all form disulphide bridges that are distributed between every feature of regular secondary structure creating a highly stable and compact structure. This also explains the proteins stability despite the lack of a hydrophobic core, characteristic of globular proteins. Instead the majority of hydrophobic residues are concentrated to the loops at one end of the protein giving rise to a hydrophobic surface patch. The result is a molecule that displays the amphiphilicity required of a surfactant and hence explains the source of the observed activity.

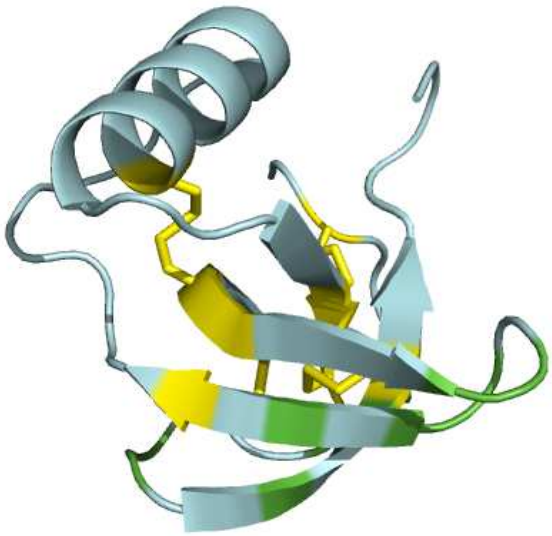


Figure 1.2- Hydrophobin monomer. Cartoon representation of the crystal structure of the HFBII monomer. The four disulphide bonds are coloured yellow and aliphatic residues that contribute to the hydrophobic patch are coloured in green. Diagram was created from PDB file 1R2M.pdb using PyMOL software [46,47].

Homologous to small molecule surfactants, the hydrophobins shield their hydrophobic patches from the solvent by forming various aggregates such as dimers and tetramers, dependent upon protein concentration [48]. In the crystal structure of HFBII, the asymmetric unit contains a dimer with the hydrophobic patch on each molecule directed towards each other (Figure 1.3) [46].

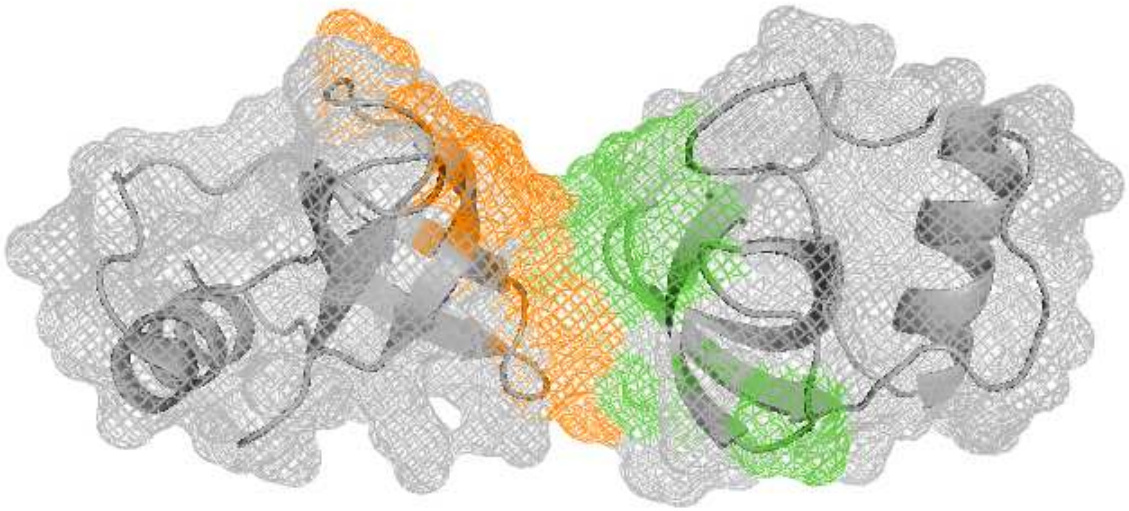


Figure 1.3- Hydrophobin dimer. Crystal structure of the HFBII dimer showing both cartoon and surface representations. The hydrophobic surface patches are coloured green and orange for each monomer. Diagram was created from PDB file 1R2M.pdb using PyMOL software [46,47].

It was observed that the contact face only consists of approximately half of the hydrophobic patch, leaving the remaining hydrophobic surface exposed to the solvent. This has led to the proposition that the dimer state can be considered a high energy state and that larger hydrophobic environments such as the air-water interface, could provide a situation where the entire patch could escape solvation, producing a low energy state [46]. This energy difference could be what drives the molecule's adsorption into interfaces rather than remaining as dimers in solution.

The potential of the unusual properties of the hydrophobins in numerous roles including biomedical applications, fusion partners for protein purification and anti-fouling agents has been realised and has led to numerous patents [49]. The proteins that form the focus of this thesis ranaspumin-1, ranaspumin-2 and latherin also possess high levels of surfactant activity. These proteins have differentiated themselves from the hydrophobins in their activity and therefore present alternatives to the hydrophobins where the molecular basis of the observed activity is poorly understood. However before introducing members of this group, a major impedance in the study of these proteins must be considered: the inability to extract high resolution data with regards to protein structure from the air-water interface.

1.1.3 Techniques for the Study of Molecular Structure at the Interface

Conventional high resolution techniques such as X-ray crystallography and solution state NMR are not applicable at the interface. Traditional methodology such as surface tension measurements and Langmuir trough methodology can be employed to characterise a species' behaviour at the interface, such as surface adsorption isotherms, but do not usually provide insight into the form or orientation of the contributing molecules.

Neutron reflection is a technique where a beam of monochromatic neutrons is reflected off the surface of interest [50]. Neutrons are utilised as their wavelengths are two to three magnitudes shorter than that of light allowing thinner films to be probed. The intensity of both the incoming and exited beams are monitored to produce a ratio of the two, termed neutron reflectivity [51]. The neutron reflectivity can then be plotted against the wave vector, a function of beam wavelength and scattering angle, to produce a reflectivity profile of the surface. This reflectivity profile is characteristic of the composition or density of the surface layers. Unlike X-rays, neutron scattering involves interaction with the atomic nucleus and is therefore sensitive to isotopes. Thus, by carrying out neutron reflectivity in various H_2O/D_2O mixtures the resultant profiles can be fitted to theoretical models to produce estimates of density and thickness of layers of surfactant perpendicular to the interface [51]. This technique is valuable for probing the form of these layers but provides little insight into the structure of the constituent surfactant molecules.

Infrared reflection absorption spectroscopy (IRRAS) provides more direct information related to the structure of the molecules at the interface. IRRAS is a variant of conventional infrared spectroscopy where the infrared light is reflected off the target surface at a low incidence angle before detection [52]. Particular molecular conformations are identified based upon characteristic absorption wavelengths. With regard to proteins, IRRAS can be utilised to probe secondary structure content and its relative orientation to the interface [52]. Again the technique is unable to produce high resolution data with regards to the three dimensional structure of the molecules in this interfacial region.

1.2 Ranaspumin-1 & Ranaspumin-2

The foam nests produced by the Tungara frog provide a hospitable environment in which embryos develop into tadpoles. The nests are able to survive for several days without succumbing to dehydration, predation or microbial attack. This compatibility of surfactant activity and cell development is both remarkable and unique as conventional detergents are highly efficient at cell disruption. It is the source of these various remarkable features of the foam that creates the focus of this investigation.

1.2.1 Engystomops pustulosus

Engystomops pustulosus (Tungara frog, previously designated *Physalaemus pustulosus*) (Figure 1.4) is one of 650 species belonging to the Leptodactylidae family common to areas of South and Central America and the Caribbean [53]. They are small (30-50mm), brownish coloured, ground dwelling frogs residing in leaf-litter and grassed areas or marshes [54]. The frog's skin is pustular (covered in small lesions), hence the name pustulosus.

Figure 1.4- An adult *Engystomops pustulosus*. Image courtesy of Alan Cooper and Malcolm Kennedy.

Tungara frogs breed in shallow, slow moving, or stagnant bodies of water, ranging anywhere in size from a water-filled footprint to a large flooded field [54]. During breeding the female deposits several eggs as the male clasps her from the top. He then, whilst depositing sperm, beats the fertilised egg containing mixture with his legs, whipping

it up into foam. The foam nests are left unattended for the remainder of the development from eggs to tadpoles. At the point when tadpoles are ready to hatch the foam dissolves allowing the tadpoles to escape into surrounding water.

1.2.2 Foam nests

The foam nests of *Engystomops pustulosus* attracted the interest of researchers due to the apparent plethora of unusual properties demonstrated by this natural material. The ability of the nests to survive for the duration of embryogenesis (approximately 3 days, although the nests have been observed to endure for up to 10 days in the absence of developing eggs) in the hostile environment they are deposited is a significant feat in itself [55]. The foams are mechanically stable and do not succumb to physical manipulation but are however sufficiently elastic to adapt to various conformations as required by their environment. The nests resist evaporation in tropical heat, with no source of hydration, nor are the apparently defenceless, potentially rich sources of nutrition, disturbed by predation. And despite the fact that the nests are routinely produced in pools of stagnant water, teeming with microbial life, they appear resistant to bacterial infection. Furthermore the compatibility of the nests and the reproductive cells suggest that this anti-microbial activity is not a consequence of membrane disruption caused by the foam producing, surface active agents within the nest, as may be expected [55]. The sum of these remarkable properties is an extraordinary material with numerous potential applications.

The foams have been studied and characterised. They have a density of approximately 0.1g/cm^3, the wet phase of which consists of a mixture of protein and carbohydrates, both at concentrations of 1-2mg/mL. The bulk nest mixture has been shown to be highly surface active as would be expected from a material capable of producing stable foams. However, there is no detectable level of fats or lipids normally associated with surfactant activity, indicating a protein or carbohydrate source, which may go some way to explaining the compatibility with the nests' biological components [56].

The protein components of the foam have been studied extensively and six main proteins, ranging in mass from 10 to 35kDa, have been determined. They have been named the "ranaspumins" (Latin: *rana*, frog; *spuma*, froth, foam) and numbered 1 to 6 by their increasing apparent molecular weight by SDS-PAGE [55]. The ranaspumins can be separated into three categories based upon their sequences and proposed function: ranaspumin-1, ranaspumin-2 & ranaspumins 3-6.

1.2.3 The Ranaspumins

1.2.3.1 Ranaspumin-1

Ranaspumin-1 (Rsn-1) has a mass of 11.4kDa and an estimated acidic pI of 5.06. Rsn-1's amino acid sequence (Figure 1.5), contains 109 residues including two cysteine residues, which may indicate the presence of a disulphide bridge. Other areas of note are the glycine rich N-terminus which may indicate flexibility and the high proportion of proline residues throughout the sequence [55].

```
  1. GGGNIGGGAK
 11. LGPEKPATPG
 21. IQDLLKSLLS
 31. VLNLSPPAIP
 41. EDAEAVSYRD
 51. AKNGKFRLIK
 61. IHLGGELYCH
 71. VKQIAGPILA
 81. LPIVSDVVEV
 91. TGKECGKTED
101. DPLEDFPIP
```

Figure 1.5- Rsn-1 amino acid sequence [55].

Similarity searches have identified intermediate similarities between Rsn-1 and members of the cystatin super family of proteins (Figure 1.6). The cystatins inhibit the activity of cysteine protease enzymes and are common throughout nature. This led to the initial suggestion that Rsn-1 acting as a protease inhibitor could be responsible for the foam nests' resilience against microbial attack [55]. However, the well-defined conserved motifs associated with the cystatins' activity are all absent from the Rsn-1 sequence [57]. This observation is re-enforced by the lack of any anti-cysteine protease activity from recombinant protein [55]. This suggests that Rsn-1 may have a similar fold to the cystatins but an alternative function.

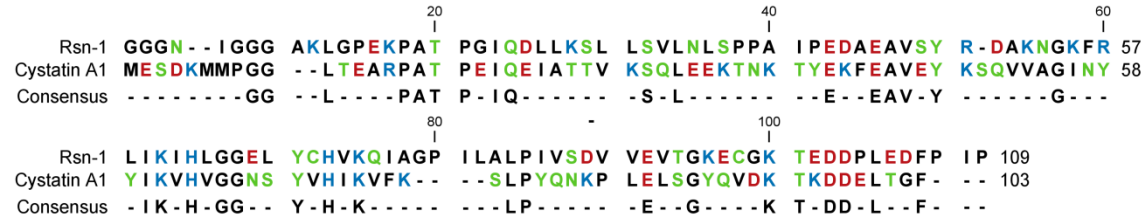


Figure 1.6- Sequence alignment of Rsn-1 and cystatin A1. The Rsn-1 amino acid sequence was aligned with the sequence of its closest homologue, pig cystatin A1 (NP-999024). Hydrophobic residues are shown in black, hydrophilic in green, basic in blue and acidic in red. The alignment was created using the CLC Bio Workbench 4.0.2 software [58]**.**

The role of Rsn-1 within the foam mixture therefore remains unresolved. Its homology and likely similar fold to that of the cystatins may indicate a function alongside that of surfactant Rsn-2 which also has, despite a lack of sequence homology, a cystatin-like fold. Secondary structure predictions (PsiPred [59]) based upon amino acid sequence indicate a similar distribution of regular secondary structure to that observed in Rsn-2. It is therefore possible that Rsn-1 may present another surface active component of the bulk foam acting via a similar, if not identical, mechanism to that of Rsn-2. Rsn-1's greater degree of sequence similarity to members of the cystatin family, over Rsn-2 may indicate its position as an evolutionary intermediate between the two. The structure and characteristics of Rsn-1 are therefore of interest, not only to establish the proteins role within the foam nest, but if Rsn-1 is confirmed as a cystatin-like surface active agent, the solution structure would provide a second model for the rationalisation of activity in Rsn-2. Perhaps allowing the identification of common motifs required for the surfactant activity of both proteins.

1.2.3.2 Ranaspumin-2

Ranaspumin-2 (Rsn-2) is an 11kDa, monomeric protein and was identified as being of particular interest in early work as a result of its unusual amino acid sequence which is unlike any known protein (Figure 1.7).

```
 1. LILDGDLLKD
11. KLKLPVIDNL
21. FGKELLDKFQ
31. DDVKDKYGVD
41. TKDLKILKTS
51. EDKRFYYVSV
61. DAGDGEKCKF
71. KIRKDVDVPK
81. MVGRKCRKDD
91. DDDDGY
```

Figure 1.7- Rsn-2 amino acid sequence [55].

There are numerous remarkable features within the Rsn-2 sequence. It contains an unusually high level of both lysine (18.8% vs. the 5.9% SwissProt average) and aspartic acid (21.9% vs. the 5.5% SwissProt average). As a result the protein will contain an extraordinarily high level of charged residues, however due to the opposing charges nullifying each other Rsn-2 has only a slightly acidic pI of 5.16. Furthermore, six of the twenty one aspartic acids are found consecutively at the C-terminus in a motif not present in any other known protein. This motif also contributes to an unusual amphiphilic amino acid distribution, with a slightly hydrophobic N-terminus and hydrophilic C-terminus (Figure 1.8).

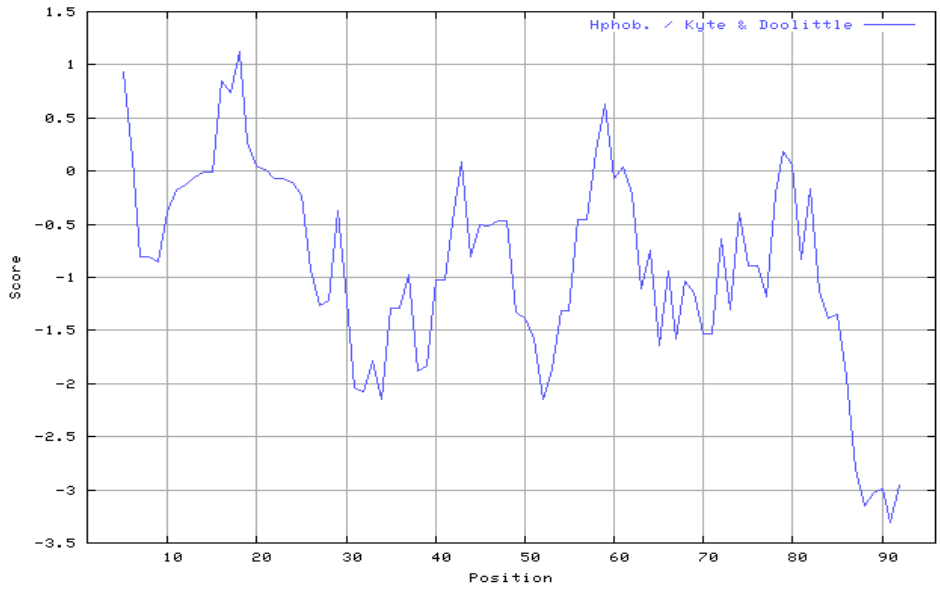


Figure 1.8- Hydropathicity profile of Rsn-2. Primary sequence of Rsn-2 plotted against Kyte and Doolittle hydropathy index values [60]. Chart created using Protscale programme [61].

This suggested amphiphilicity together with the lack of homology fuelled the proposition that Rsn-2 could be the surface active component of the foam nests. Recombinant Rsn-2 was assayed against control proteins to investigate for evidence of surface activity. Rsn-2 was observed to produce a significantly greater reduction in surface tension, at substantially lower concentrations (< 10µg/mL) than any of the control proteins [55]. Recombinant Rsn-2 also appears to be active at concentrations equivalent to an approximate 10x dilution of the bulk nest mixture, suggesting that Rsn-2 alone is significantly more surface active than the complete foam mixture. Rsn-2 was also found capable of producing foam under agitation, however foams produced from Rsn-2 exclusively collapsed after a few minutes, compared to days for the bulk mixture, indicating that some other component is required to stabilise the natural foam.

The interesting ability of the bulk foam to interact with, without disrupting biological molecules has been identified in Rsn-2 also [62]. The potential of this unique property for a surfactant molecule is illustrated by work carried out by Wendell et al. in the field of artificial photosynthesis [63]. The group have created a system which utilises adenosine triphosphate (ATP) produced by light harvesting lipid vesicles to convert CO_2 into sugars via the Calvin-Benson-Bassham cycle. In bulk solution the efficiency of the system is limited by availability of CO_2 and light. By transferring the system to an Rsn-2 based foam, a substantially larger liquid to CO_2 surface ares was created providing greater availability of substrate. The foam also allowed further penetration of light into the system. The Rsn-2 foam system was shown to increase the yield of sugar by 300% vs. the same

system in bulk liquid. It was also demonstrated that this would not be viable utilising conventional surfactants which disrupted the lipid vesicles.

There is no evidence of any interaction of Rsn-2 with lipids or carbohydrates, nor is its activity limited to high concentrations as is required with denatured proteins [62]. Rsn-2 is similar to the hydrophobins in that the observed surfactant activity is an intrinsic property of the protein. However, there is no sequence homology between the two groups, nor does Rsn-2 associate together to form multimers in solution, which is critical for the hydrophobins to remain soluble as discussed earlier. Therefore Rsn-2 must employ some other mechanism and hence represents a potentially new class of surfactant protein [62].

Apart from the amphiphilicity within the Rsn-2 sequence there is no indication of what this mechanism is. In the case of hydrophobins the activity can be rationalised from the protein's structure. Therefore in an attempt to understand the source of Rsn-2's surfactant activity the structure of protein was determined by NMR (Figure 1.9) [62].

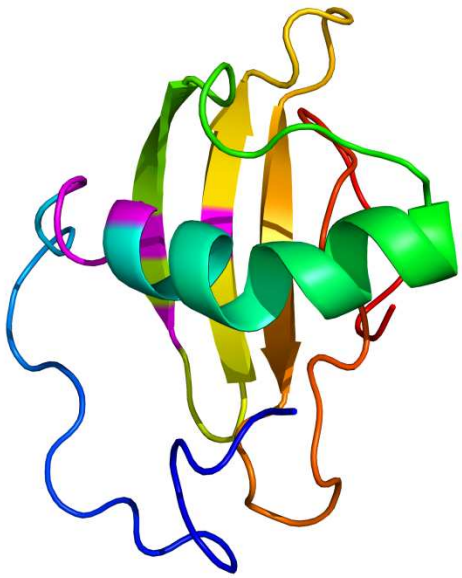


Figure 1.9- Solution structure of Rsn-2. Cartoon diagram showing Rsn-2 coloured blue to red from the N to C- termini. The residues identified as possessing alternate conformations from NMR data. Image created from PDB file 2WGO.pdb using PyMOL software [47,62].

Rsn-2, despite the lack of sequence similarity, has a cystatin-like fold, a kinked α-helix over-laying a four stranded β-sheet. Strands three and four of the sheet are linked by a disulphide bond and the hydrophobic N-terminus and hydrophilic C-terminus are largely unstructured. On inspection of the surface there is nothing to suggest the amphiphilicity

associated with a surfactant, the protein being enclosed in polar residues as would be expected in any globular protein (Figure 1.10).

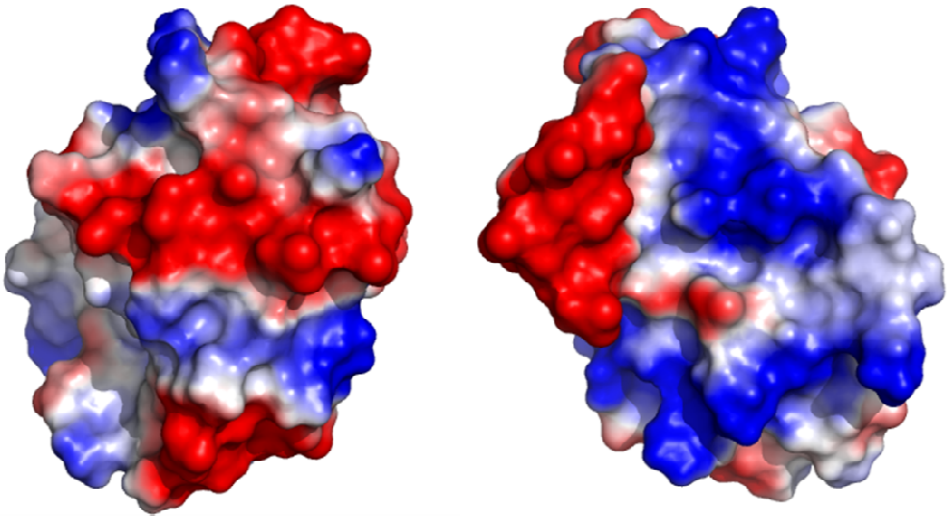


Figure 1.10- Rsn-2 surface vacuum electrostatics. Alternate faces of the Rsn-2 surface showing positive atoms blue, negative atoms red and uncharged atoms white. Image created from PDB file 2WGO.pdb using PyMOL software [47,62].

The solution structure however only describes the structure of Rsn-2 in solution. What is critical in order to understand its observed activity is its structure at the air water interface. Neutron reflectivity was used to investigate the form of the protein sheet at the interface. The data produced a two-layer model around the air-water interface [62]. At the interface, extending into the air is a densely packed layer (>90%) of protein at a constant thickness of 8-10 Å, regardless of concentration. Below, a 20-40 Å layer of loosely packed protein at levels of 10-30% dependent upon concentration. It was proposed that the upper layer corresponds to the majority of the protein with hydrophobic and hydrophilic residues projecting into the air and aqueous environments respectively and that the lower layer consisted of additional hydrophilic regions extending further into the water. This would require almost all of the protein to fit into a 10 Å plane. However the solution structure designates that Rsn-2 is greater than 20 Å in its smallest dimension. This raises further inconsistencies with regards to the relationship between activity and the solution structure, as it is impossible for a globular Rsn-2 structure to be compatible with the experimental data on interface dimensions.

As mentioned earlier, it is possible for proteins to denature on adsorption into the air-water interface and become surface active. This was deemed not to be the case for Rsn-2 for several reasons. Rsn-2 achieves far greater effects upon surface tension at considerably

lower concentrations than observed for any denatured conventional protein [55]. Rsn-2 has also been shown by differential scanning calorimetry (DSC) to be highly stable in solution (T_m= 80 ± 1°C) [62]. More directly, infra red reflection absorbance spectroscopy (IRRAS) has identified that there is no change in the proportions of secondary structure between the solution and surface structures.

As a consequence of these results, it was concluded that Rsn-2 must undergo some conformational exchange in order to fulfil both its observed role and the experimental data. A model was proposed whereby the α-helix unfolds away from the β-sheet, exposing the proteins hydrophobic core to the air and aligning the hydrophilic exterior in one plane, directed into the water (Figure 1.11). This was described as a "clam-shell" like opening of the protein and would give rise to the amphiphilicity required for the surfactant activity, maintain the secondary structure and agree with the neutron reflectivity data [62]. Furthermore this model gave an improved correlation to the IRRAS data as opposed to the "closed" solution structure. Also detected were a series of peaks within the NMR spectra corresponding to alternate conformations of residues 21-Gly, 22-Lys, 48-Thr, 50-Glu and 57-Val (Figure 1.9). It was suggested that this secondary form for these amino acids that constitute the lips of the Rsn-2 structure (the area that would be most affected by the proposed clam-shell opening) may represent some precursor to the hypothesised open, active form of Rsn-2. It was also proposed that the two termini play a significant role: the hydrophobic N-terminus acts a trigger on interaction with the air, pulling Rsn-2 to the surface, whereas the charged, poly-aspartic acid C-terminus, acts as an anchor remaining in the water.

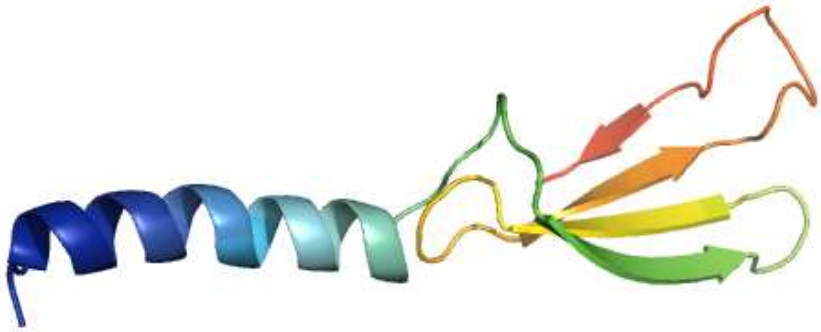


Figure 1.11- Open Rsn-2 conformation. Cartoon diagram showing the hypothesised open Rsn-2 structure. Coloured blue to red from the N to C- termini. Image created from PDB file 2WGO.pdb using PyMOL software [47,62].

However, due to the lack of high resolution structural information available at the air-water interface this potential secondary conformation of Rsn-2 is difficult to prove and thus remains a hypothesis.

1.2.3.3 Ranaspumins-3 to 6

Ranaspumins-3, 4, 5 & 6 range in mass from 18-26kDa and possess sequences with similarities to the sugar binding lectin family of proteins. Rsn-6 shows greatest similarity to the galactose binding, C-type lectins, common throughout nature. Rsn-3, 4 & 5 show greatest alignment to the fucose specific, F-type or 'fucolectins'. This proposed functionality has been confirmed for Rsn-4, however in contrast to the family to which it belongs, the activity is not inhibited by fucose but rather by galactose. This observation is not entirely surprising as, despite the degree of conservation between Rsn-3, 4 & 5 and the fucolectins, there are differences in the amino acids postulated to comprise the carbohydrate binding site suggesting differing specificity [55].

These lectin-like proteins are proposed to play a number of defensive roles within the foam nest [55]. Firstly, against microbial attack, where lectins can bind to invading cells which contain target sugars on their surface, preventing dissipation and hence colonisation of the nest. Secondly, as a deterrent against predation of the protein and carbohydrate rich foams. The agglutination activity of lectins has been shown to be highly disruptive to the gut tissues of both insect and vertebrates, potentially making the nests a less appealing source of nutrition [64].

It is also believed that the lectins may perform some structural function. In Rsn-3 and Rsn-5 hydrophobic signal peptides are retained after secretion [55]. This has led to the suggestion that the lectin proteins may utilise these hydrophobic termini to act as a link between the surfactant protein and carbohydrate components of the foam to create a mechanically stable foam matrix (Figure 1.12). As discussed earlier foams produced with Rsn-2 alone do not maintain the same stability as the bulk mixture and therefore the lectin-like ranaspumins and carbohydrates are critical to formation of a stable complex.

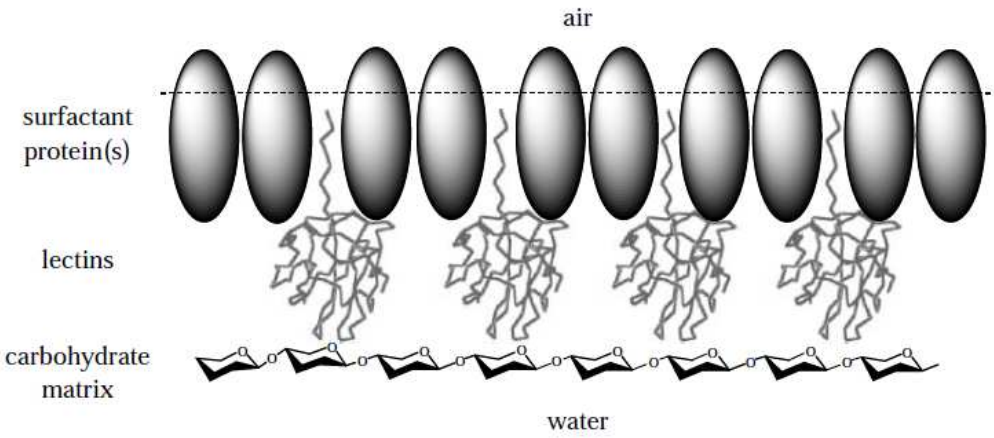


Figure 1.12- Cartoon of proposed role of Rsn-3, 4, 5 & 6.. Reproduced from [55].

1.3 Latherin

1.3.1 Equine Thermoregulation

The equine family consists of six species, the domestic horse (*Equus caballus*), Prewalski's horse (*Equus prewalski*), the African wild ass (*Equus asinus*), the Asian ass (*Equus hemionus*), Grevy's zebra (*Equus grevyi*), the common zebra (*Equus burchelli*) and the mountain zebra (*Equus zebra*). The family evolved as plain feeders, relying upon caution and an ability to outrun potential predators for survival. As a result equidae have developed inherent athletic capabilities, far superior to any other mammals. The energy required to fuel these exertions is derived from the conversion of stored chemical energy. This process is somewhat inefficient and it has been stated that approximately 80% of the converted energy is lost as heat [65]. This released heat energy has to be dissipated to prevent hyperthermia. Most mammals rely upon respiratory evaporation as the main form of heat loss [66]. Inspired air undergoes heat exchange upon the large surface area of the lungs and is exhaled along with water vapour. Panting may also be employed to increase respiratory rates and thus heat loss. Equines and humans are unusual in that, although respiratory evaporation is employed, they depend more upon evaporative cooling from sweating as the primary thermoregulatory process. It appears this process evolved in these endurance animals to overcome the reduced efficiency of respiratory evaporation during exercise due to the competing processes of maximal gaseous exchange and maximal evaporative heat loss [67]. The dependence upon sweating for thermoregulation is demonstrated by the condition anhidrosis: the inability to sweat. Anhidrosis not only significantly affects the ability of the horse to exercise but also causes significant discomfort [68].

The process of sweating in equines also has an apparent disadvantage. Unlike humans, equines are covered in a thick waterproof pelt which intuitively should inhibit the sweat vaporisation. This however does not appear to be the case: sweat readily transfers to the surface. This observation alongside the ability of the sweat to form foams/lathers led to the proposition that some surface active species is present in sweat which wets the hydrophobic hair aiding the spread of sweat and hence evaporation and cooling [69]. This theory was reinforced by results identifying the concentration dependence of surface tension for solutions containing horse sweat [70].

Equine sweat is alkali and unlike that of humans, is hypotonic relative to plasma [65]. Ions such as Na^+, K^+ and Cl^- are present at concentrations ranging from 50-250 mM, as well as

trace amounts of metals including calcium and magnesium [71]. Numerous other substances have been detected but are associated to secretory function. There is no evidence to suggest the presence of lipids, which is surprising given the sweats' ability to produce foam.

The major difference between the sweat of humans and that of the equine family is the high concentration of protein (5-10 g L^{-1}) present in the latter [71]. Early investigators assumed that serum albumin was the main protein constituent detected in horse sweat. However, subsequent work indentified the proteins detected as previously unreported [72]. Two protein bands were separated by size exclusion chromatography, the "heaviest", according to the elution volume, labelled H, and the "lightest", labelled L. The greater resolution achieved by Beeley *et al*. [70] allowed these bands to be purified further separating the L peak into a further two distinct peaks and identifying three major protein containing fractions. The surfactant activity of the fractions was assessed by capillary rise measurements, with the aim of identifying the source of the sweats' surfactant activity. No activity was observed for the two heaviest fractions. However the third displayed significantly reduced surface tension and, furthermore, produced foam readily. SDS-PAGE showed that the surface active fraction consisted of a single protein with an apparent mass of 20kDa containing at least one intramolecular disulphide bond. This protein was attributed as the source of the sweat's surfactant activity and was consequentially named latherin [70]. Latherin has since been reported in the sweat of numerous members of the equidae family in highly conserved forms suggesting this a feature common throughout all equidae species [73]. Latherin is not limited to the sweat, it is also found within the saliva of horses, indicating some dual function, possibly to aid the mastication of the dry food sources that equidae forage [71,73].

1.3.2 Latherin Sequence and Similarities

Latherin from the domestic horse (*E. caballus*) consists of 228 amino acids, the first 20 of which appear to be a secretory signal peptide that is removed posttranslationally (Figure 1.13) [73]. Cleaved latherin has a molecular weight of 22.6kDa (uncleaved, 24.2kDa) and is acidic with a pI of 4.11. The amino acid composition is unusually rich in aliphatic residues, in particular leucine which contributes almost 25% of the residues present, compared to the SwissProt average for all proteins within the database of 9.67%. The sequence contains one pair of cysteines which are believed to form an intramolecular disulphide linkage due to the differences in migration of reduced and non-reduced samples of latherin purified directly from horse sweat [70].

```
-19. mlkvsclfvl  lcgllvpssa
  1. QQIPPEVSSQ  ITDALTQGLL
 21. DGNFLSLLNA  INLEGLLNTI
 41. LDQVTGLLNI  LVGPLLGPSD
 61. AEIKLQDARL  LQLSLEFSPD
 81. SKGIDIWIPL  ELSVYLKLLI
101. LEPLTLYVRT  DIRVQLQLES
121. DEDGKYRLAF  GHCSLLPRAI
141. ELQSGNPLSL  PVNAVLGTIE
161. NALGNFITED  LGAELCPTLN
181. SLVSNLDLQL  VNNLINLILD
201. RANVDLSV
```

Figure 1.13- Horse latherin amino acid sequence [73]. The signal peptide is shown in lower case and underlined. Leucine residues are shown in bold red. Cysteines are shown in bold blue.

Latherin has been identified as a member of the palate, lung and nasal epithelium clone (PLUNC) proteins, a branch of the bactericidal/permeability increasing fold (BPIF) family of proteins. As well as the PLUNCs, the BPIF family includes BPI (bactericidal/permeability-increasing protein), LBP (lipopolysaccharide-binding protein), PLTP (phospholipid-transfer protein) and CETP (cholesteryl ester transfer protein). The PLUNCs are defined by sequence and structure similarity to the bactericidal permeability increasing-like (BPI-like) group of proteins, highly conserved genomic organisation and exon sizes and two conserved cysteine residues that are also conserved features of the BPI-like proteins [74].

1.3.2.1 BASE

Latherin's closest homologue by sequence is another member of the PLUNC family, human breast cancer and salivary gland expression (BASE) protein. In normal tissue BASE is exclusively expressed within the salivary glands but has also been detected in some breast cancer cell lines [75]. BASE has therefore been identified as a potential diagnostic marker for breast cancer. BASE has been identified as a defective gene in humans which due to a 1 bp deletion, produces a truncated 179 amino acid protein, rather than a 228 amino acid latherin-like protein [76,77]. Bingle *et al*. proposed that BASE represented a "dying gene" that, despite not producing a viable folded protein, was still expressed [78].

1.3.2.2 BPI, LBP, CETP & PLTP

The BPI-like proteins are a group of four lipid binding proteins: bactericidal/permeability-increasing protein (BPI), lipopolysaccharide binding protein (LBP), cholesteryl ester transfer protein (CETP) and phospholipid transfer protein (PLTP). Although, as stated above, latherin is more similar to the PLUNCs, the BPI-like proteins are of interest with regards to the scope of this project as crystal structures of BPI and CETP provide the only sources of high resolution structural information for any members of the superfamily, thus may provide some insight that could be useful in understanding the activity of latherin.

LBP and BPI are two similar proteins with related but opposing functions. Both consist of approximately 450 residues, distributed into two domains and share over 45% sequence identity [79,80,81]. They function as part of the mammalian innate immune response to lipopolysaccharide (LPS). LPS is a major a component of the outer wall of gram-negative bacteria (GNB) [82]. When detected within the mammalian blood system, LPS initiates a general inflammatory immune response that is generally beneficial, but in some cases an excessive systemic response can lead to the potentially lethal condition of endotoxic shock [83]. LBP, produced mainly in the liver, plays a stimulatory role, key to the host cells recognizing low levels of LPS [79]. BPI, which is released from the neutrophils, has the opposite, inhibitory effect, neutralising the immune response to LPS [84]. BPI has a secondary function and on interaction with LPS bound to the GNB cell wall, causes bacterial cell death [85]. This activity is highly selective, probably as a result of BPI's high affinity for LPS [82,86].

The crystal structure of BPI was determined by Beamer *et al*. (Figure 1.14) [80]. The two domains (residues: 1-199 & 200-456) fold in what has been described as a 'boomerang-shaped molecule', with pseudo-twofold symmetry. Each domain forms a barrel at either end of the protein, separated by a central β-sheet that contains structural features from both domains. Each barrel is constructed from three structural elements: a long α-helix, a short α-helix and a five-stranded anti-parallel β-sheet, forming a previously unreported fold. Despite the lack of substantial sequence similarity (13% sequence identity) between the two domains, they form remarkably similar structures. This can be demonstrated by a root mean squared deviation (RMSD), a measure of the average distance between two equivalent atoms in two or more superimposed structures, of only 0.9Å for the main chain atoms of the two domains [87]. The crystal structure also identified the presence of two phosphatidylcholine molecules bound within hydrophobic pockets on the concave surfaces of each domain.

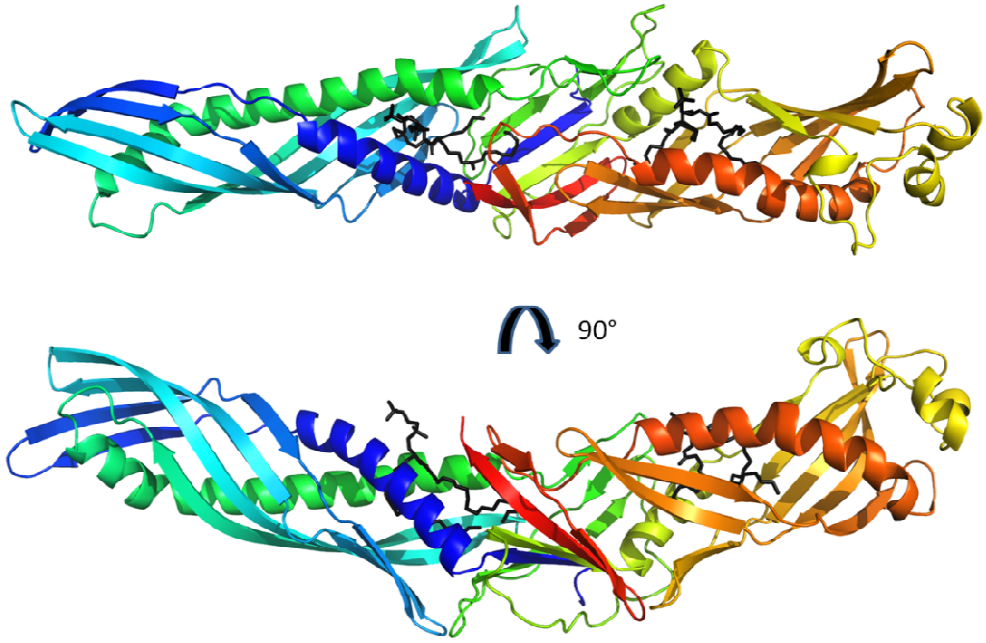


Figure 1.14- Crystal structure of BPI. Cartoon diagram showing BPI coloured blue to red from the N to C- termini. The phosphatidylcholine molecules are shown in black. Diagram was created from PDB file 1BP1.pdb using PyMOL software [47,80].

Although no structure is available, molecular modelling has predicted that LBP forms a very similar fold to that of BPI, as would be expected due to the high degree of sequence similarity [81,88].

In both LBP & BPI proteins there are distinct functional differences between the two domains. Both C-terminus domains are associated with delivery of LPS to specific host molecules to mediate subsequent response [81,89], whereas, the N-terminal domains bind GNB and aggregates of LPS. More specifically, in both proteins a cluster of basic amino acids situated upon β-hairpins at the extreme tip of this domain are believed to mediate initial electrostatic interactions with acidic sites on LPS [81,89,90,91]. Studies have confirmed the amino-terminal domain of BPI as the source of the antimicrobial cytotoxicity, to the extent that activity is retained in the absence of the entire C-terminal domain. There is also biophysical evidence indicating that on interaction with LPS-containing membranes, the amino terminal domain of BPI unfolds before initiating cell disruption [92]. Such is the potency and selectivity of the antibiotic and LPS-neutralising activity of BPI, that a recombinant N-terminal fragment (residues 1–191) is currently undergoing clinical trials [89,93].

Cholesteryl Ester Transfer Protein (CETP) and phospholipid transfer protein (PLTP) both play roles within lipoprotein metabolism. Lipoproteins are biomolecular assemblies consisting of both protein and lipid components that transport hydrophobic lipid molecules throughout the aqueous phase of the blood system [94]. They are organised in such a way that hydrophilic regions of the constituent molecules are projected outwards and hydrophobic inwards, analogous to a micelle structure. Lipoproteins are grouped by their density which decreases with increasing size: high density lipoprotein (HDL), low density lipoprotein (LDL) and very low density lipoproteins (VLDL). Elevated LDL/VLDL levels have been demonstrated to be associated with atherosclerosis; the blockage of arteries due to the accumulation of fatty materials such as cholesterol [95]. HDL however, has been identified as atheroprotective [96,97].

CETP is a 74 kDa, 476 amino-acid, glycoprotein that is secreted primarily from the liver [98,99]. It is the only reported member of the PLUNC/BPI superfamily that is not found on human chromosome 20, located instead upon chromosome 16 [100]. CETP catalyses the transfer of cholesteryl esters from HDL to LDL and VLDL lipoprotein [101]. Consequentially CETP has been identified as a potential drug target to prevent the development of atherosclerosis [102].

Phospholipid transfer protein (PLTP) also consists of 476 amino acids, however has a higher mass of 81 kDa as a result of six N-linked glycosylation sites [103]. PLTP acts as a mediator in plasma lipoprotein metabolism, promoting the transfer of phospholipids from

phospholipid vesicles or lipolysed VLDL to form new HDL [104]. PLTP has also been attributed the ability to remodel these HDLs [105,106]. PLTP's activity is not restricted to phospholipids as it is able to transfer other amphiphilic molecules including cholesterol, LPS and vitamin E [107,108,109].

The crystal structure of CETP was determined by Qiu *et al*. [110], confirming predictions of a similar fold to that of BPI (Figure 1.15). The two structures are almost superimposable, with the N-termini in particular matching up well (RMSD of 1.6Å for Cαs). The crystal structure of CETP also contained bound lipids: two cholesteryl oleate molecules and two dioleoylphosphatidylcholine molecules occupy a continuous tunnel that spans the length of the protein. The tunnel has two openings on the concave surface that are each plugged by the phospholipids, identical to the location of the phospholipids in the BPI structure [80,110]. The two cholesterol esters are completely buried within the hydrophobic core of the protein.

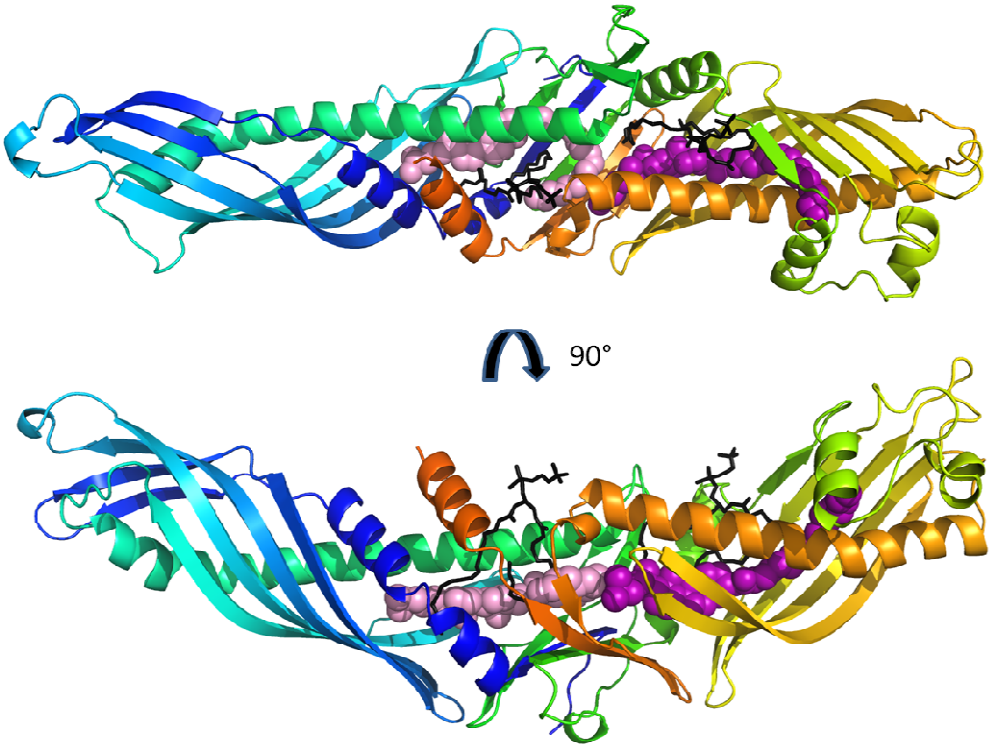


Figure 1.15- Crystal structure of CETP. Ribbon diagram showing CETP coloured blue to red from the N to C- termini. The bound phosphatidylcholine molecules are shown as black bonds and cholesteryl esters as pink and lilac space fills. Diagram was created from PDB file 2ODB.pdb using PYMOL software [47,110].

Although there is no reported structure of PLTP, the high degree of sequence identity with BPI (28%) allowed Huuskonen *et al.* to model PLTP's structure [111]. It was observed that residues surrounding the lipid-binding pockets in BPI were well conserved, suggesting a conservation of the structural features also. This was reinforced by mutating tryptophan groups into both proposed binding pockets, which produced a significant reduction in phospholipid transfer activity.

As well as these conserved features, there are some key differences between the CETP structure and the PLTP model and their close relations BPI and LPS. The most obvious structural difference is an extended C-terminus, common to both CETP and PLTP, which in CETP is found in the form of a distorted, amphipathic, alpha helix which interacts with the N-terminal domain. This helix is proposed to play a role in facilitating neutral lipid exchange and has previously been identified as critical for activity in both CETP and PLTP [110,112,113]. The angle between the two CETP domains is also significantly less obtuse than observed in BPI. This observation has been rationalised as a consequence of BPI's requirement to interact with bacterial cell surfaces, where further curvature could cause steric impedance, as opposed to the relatively small HDL [110].

There are also areas of concentrated variation on the loops at the distal tip of the N-terminus, not only between the two groups (BPI & LPB and CETP & PLTP) but also between CETP & PLTP. As discussed earlier, in BPI and LBP this area contains numerous positively charged residues, which have been identified as crucial for the protein's ability to bind negatively charged regions of LPS [81]. In the PLTP model a cluster of hydrophobic aromatic residues replaces the basic groups. Acting upon the hypothesis that this surface-exposed hydrophobic patch was involved in substrate interaction, Desrumaux *et al*. confirmed that these residues are critical for PLTP's ability to bind HDL [114]. In the CETP structure, this area is also occupied by hydrophobic residues that give rise to a hydrophobic patch at the extreme of N-terminal domain. The loops also contain a sequential pair of solvent-exposed tryptophans at positions 105 and 106. However, unlike in the other the members of the family, no direct relationship between this region in CETP and biological activity has been reported.

1.3.2.3 PLUNCs

The PLUNC family of proteins consists of at least nine members, all encoded by sequential genes on human chromosome 20 [100,115]. The PLUNCs are found throughout the respiratory tract of air-breathing vertebrae and, as yet, have no confirmed function [78,116]. However, the relationship to the BPI-like proteins has led to the suggestion that the PLUNCs may function as elements of the innate immune system, localised to a region openly exposed to microbial attack. Latherin, therefore, not only presents an opportunity to investigate the functionality of a unique surfactant protein, but also to determine the structure of a whole family of proteins, for which there is little structure and function information.

Despite the relatively low intra PLUNC family sequence identity of 16-28%, and pairwise identity ranging from 13-22% between the PLUNC family and BPI, all members have been predicted to form a BPI-like fold [115]. The family can be subdivided into two groups according to size, the ‘long’ (LPLUNC) and the ‘short’ (SPLUNC) PLUNCs. The LPLUNCS range in size from 458-484 amino acids, equivalent to both domains of BPI, the C-termini of which are believed to be structurally similar to the corresponding BPI region. The SPLUNCs, all of approximately 250 amino acids probably form a single domain and based upon sequences, are predicted to be more like the N-terminus of both the LPLUNCs and BPI than the C-terminus. The N-terminal domain also displays the greatest variability throughout the various PLUNCs. Once again, much of this variability is concentrated to the residues predicted to form β-hairpins at the exterior tip of the domain, which, if involved in biological activity, may indicate a broad range of selectivity.

One member of the PLUNC family shares the high leucine composition of latherin. Human PLUNC, also known as SPLUNC1, consists of 237 amino acids, 23.2% of which are leucines and there is also 28% sequence identity between the two proteins (Figure 1.16) [100].

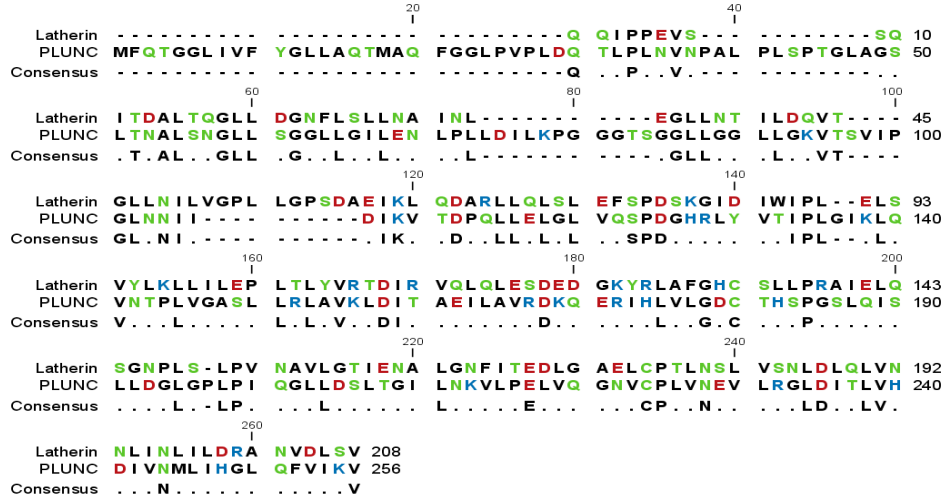


Figure 1.16- Multiple sequence alignment of latherin and PLUNC. The latherin amino acid sequence was aligned with the sequence of human PLUNC (NP_057667). Hydrophobic residues are shown in black, hydrophilic in green, basic in blue and acidic in red. The alignment was created using the CLC Bio Workbench 4.0.2 software [58].

Circular dichroism data also indicates a similar content of secondary structure, which is more similar to latherin than the N-terminus domain of BPI (Table 1.1).

Protein	α-helix	β-sheet	Other
Latherin	30 %	18 %	51 %
SPLUNC1	24-33 %	15 %	51-61 %
BPI (N-term)	27 %	46 %	27 %

Table 1.1- Comparison of secondary structure between latherin, PLUNC and the N-terminal domain of BPI. Latherin secondary structure was predicted from CD data [73]. PLUNC secondary structure was also predicted from CD data [100]. BPI secondary structure was calculated from the BPI crystal structure [80].

SPLUNC1 has been detected within the human respiratory system [117,118,119]. Like the rest of the PLUNC family, its exact function is unknown but a role in the innate immune response has been proposed. This hypothesis has been reinforced by a number of findings: SPLUNC1 has been shown to be bactericidal to a number of bacteria including *Pseudomonas aeruginosa* and *Mycoplasma pneumoniae* [120,121], inhibit the formation of *Pseudomonas aeruginosa* biofilms [100], bind LPS, and act as an anti-inflammatory agent [120,122].

Many of the fluids that coat the respiratory tract are known to be surface active [123]. However, the proteins responsible for this activity within the lungs, SP-B & SP-C, are not present in secretions of the upper airway, suggesting an alternate source of surfactant activity [100,124]. The sequence similarity to the known surfactant protein latherin, led Gakhar *et al*. [100] to investigate whether SPLUNC1 also exhibited such activity. This was

indeed found to be the case: SPLUNC1 was found to be highly surface active at levels below its physiological concentrations (10-250 µg/mL) in a number of assays. It was also shown that other members of the PLUNC/BPI superfamily: BPI and LBP, did not share this activity and therefore this was a not a universal property of the superfamily. It was therefore concluded that SPLUNC1 was at least one of the factors responsible for the surface activity observed in upper respiratory tract fluid.

1.3.2.4 Allergens

The latherin sequence has also been identified as the source of two horse dander allergen peptides Equ c 4 and Equ c 5 [73,125]. Furthermore, it was found that the IgE antibody of horse-allergy patients bound to recombinant latherin identifying the intact protein itself as an allergen.

1.3.3 Latherin Properties

Latherin as discussed earlier has been identified as the source of the surfactant activity of horse sweat. Significant reductions in the surface tension have been recorded for solutions at concentrations of 100ug/mL and reach as low as 56mN m^{-1} at 1mg/mL (the concentration found in sweat [71]) compared to 72mN m^{-1}, the value of pure water [73]. This degree of surfactant activity is comparable to that of conventional small molecule detergents and significantly greater than observed for standard proteins. Latherin, has also been found to be a highly effective wetting agent as would be expected of a detergent and required to fulfil its proposed function [70,73]. This particular property has also been shown to be reversible, which explains why the pelt does not become permanently wetted, removing its natural water proofing. The surfactant activity is demonstrated further by neutron reflectivity measurements which show latherin forms a 10Å thick layer at the surface. It was recognized that this is remarkably thin layer for a 22kDa protein to form.

The structure and properties of latherin have been probed using numerous biophysical techniques. Beeley *et al*. were able to show that latherin behaved in both sedimentation-velocity ultracentrifugation and size exclusion chromatography as a monomer, thus excluding the possibility that latherin formed multimers in solution like the hydrophobins and must therefore act by some other method [70]. Size exclusion chromatography was used to predict MW and Stokes radius by comparison with protein standards. A Stokes radius of 2.47nm and MW of 30000 were calculated. This high apparent MW (relative to the calculated MW from the amino acid composition) was rationalised by suggesting that

latherin was asymmetric resulting in increased progression through the column. A frictional ratio of 1.27 was calculated from the experimental data. DSC experiments gave a relatively high T_m of 105°C, characteristic of a highly thermostable protein in solution [73]. Circular dichroism (CD) indicated an estimated 30% α-helix, 18% β-sheet/extended structure, 22% β-turn and 29% other secondary structures. Secondary structure prediction programmes agreed reasonably well with the CD data, as well as forecasting a central region, rich in β-strand, flanked with predominantly α-helical, N & C-termini. The single tryptophan at position 87 within the amino acid sequence was identified as a useful probe for florescence spectroscopy. An unusually high emission at 350nm, characteristic of a polar/aqueous environment that was not shifted by denaturation of the protein led to the proposition that the tryptophan was solvent exposed. An alternative possibility was that the residue was in a particularly polar environment within the protein. The ability to quench the fluorescence would however suggest that the solvent exposed tryptophan is the more likely. The quenching raised further questions regarding the tryptophan's environment. Iodide, normally considered to be highly effective at quenching exposed tryptophans, produced the lowest effect. It has been suggested that this may indicate the presence of electrostatic effects from surrounding negative charges. Although it is difficult to predict the environment of Trp-87 from sequence data alone, there is an aspartic acid at position 85, which supports this hypothesis. The role of this seemingly surface exposed tryptophan is unknown, however the feature is not unique and has been described as a 'sticky finger' in a group of lipid binding proteins [126]. It is hypothesised in these proteins that the protruding hydrophobic group is involved in lipid transfer upon interaction with cell membranes or other proteins [126]. It has hence been proposed that the exposed tryptophan may act as a "sticky finger" facilitating latherin's affinity for non-polar environments.

2 Protein Production

2.1 Overview

The process of extracting proteins of interest from their natural source can be a painstaking and expensive process. Additionally, in instances where proteins are produced by animal sources, ethical issues arise. Often recombinant expression methods are far more favourable and allow high quantities of protein to be expressed routinely and economically in the laboratory. Recombinant methodology is even more critical for the study of proteins by NMR as the low natural abundance of NMR active isotopes such as ^{13}C and ^{15}N requires the incorporation of these isotopes through labelling techniques that would not be feasible in the majority of native systems. Sources of isotopes can be expensive and it is therefore desirable that reasonable yields of recombinant protein, with regards to isotope source, are achieved in order to make the experimentation financially viable.

All recombinant protein used throughout this work was produced following a standard procedure. In outline this can be broken down into three parts: molecular biology, protein expression, and protein purification. Any notable observations or deviation from the general protocol is described in detail for each target protein and in depth descriptions of all protocols and recipes can be found in Chapter 10.

2.1.1 Molecular Biology

DNA encoding proteins is transferrable between organisms. This permits genes to be transferred between species to create new recombinant DNA allowing processes such as amplification or expression of target genes to be carried out rapidly and routinely in a manner not feasible in natural hosts.

For all three proteins, plasmid constructs already existed in some form, removing the necessity for polymerase chain reaction (PCR) amplification from complementary DNA (cDNA). It was however, on occasion desirable to transfer the genes of interest into or between expression vectors. This was carried out using either routine subcloning methodology or PCR subcloning.

2.1.1.1 Routine Subcloning

Subcloning is the practice of transferring genes from one piece of vector DNA, the parent vector, to another, the destination vector. Restriction enzymes, enzymes that cleave the sugar-phosphate backbone at specific recognition sites determined by the nucleotide sequence, were used to extract the gene of interest (the insert) from the parent vector. The restriction enzymes cut the double stranded DNA in a staggered manner producing “sticky end” overhangs on either the 5’ or 3’ strand. The excised insert was then isolated from the digested parent vector by gel extraction to prevent reinsertion. The destination vector was digested using the same restriction enzymes to produce “sticky ends” complementary to those on the insert. This facilitates incorporation of the insert as well as ensuring correct orientation. The incorporated insert was then covalently bound within the destination vector using a ligase, an enzyme that creates new sugar-phosphate backbone between the insert and vector. The recombinant plasmid DNA was then ready to be transformed into the expression cells of choice.

2.1.1.2 PCR subcloning

PCR subcloning was employed in instances where suitable restriction sites were not available for routine subcloning. Primers were designed to be complementary to the coding and non-coding strands at the beginning and end of the target gene but also to include the required restriction sites for future subcloning. PCR was then carried out with a proof reading polymerase, producing an amplified DNA fragment containing the gene with flanking restriction sites. 5’ adenine (A) overhangs were then added using Taq polymerase allowing the simple TA cloning into a destination vector with complementary 3’ thymine (T) overhangs. In this work a TOPO TA vector was used. This is a plasmid where the phosphate group of both complementary 3’ T overhangs are covalently bound to an enzyme, topoisomerase I via a tyrosyl residue [127]. Topoisomerase is found in the *Vaccina* virus where it cleaves and rejoins DNA during replication. Upon annealing of the A-tailed insert to the T-overhangs the free 5’ hydroxyl group of the insert attacks the phospho-tryosyl bond releasing the topoisomerase I, producing new phosphate backbone and hence ligating the insert within the vector. This new vector was used to clone the gene of interest with appropriate restriction sites, before it was then subcloned into the appropriate destination vector.

2.1.2 Protein Expression

All protein expression carried out within this project utilised Novagen's pET expression system. Within the pET system, gene expression is under control of bacteriophage T7 promoter. This has advantages over native *E.coli* transcription as the T7 RNA polymerase is much more active, transcribing nucleotides at a rate five times faster than its *E.coli* counterpart [128]. Also, the T7 promoter is not recognised by *E.coli* RNA polymerase, making it highly selective. As a result almost all host cells' resources are transferred towards target gene expression producing unnaturally high levels of recombinant protein.

T7 RNA polymerase is not native to *E.coli* cells and therefore expression must be carried out in modified host cells with the T7 RNA polymerase gene supplied encoded in the chromosomal DNA under control of the isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible, lacUV5 promoter. The Lac repressor, encoded within the *E.coli* genome, inhibits the lacUV5 promoter preventing basal levels of expression. Expression is initiated upon the addition of IPTG which binds to the lac repressor releasing the promoters and allowing the transcription of T7 RNA polymerase and subsequently the target gene.

2.1.3 Protein Purification

The purification of the three proteins followed the same strategy based upon conventional techniques. The process can be separated into four key steps: cell lysis and insoluble cell debris removal, Ni^{2+}-affinity chromatography, fusion tag cleavage and removal, size-exclusion chromatography.

2.1.3.1 Cell Lysis and Insoluble Cell Debris Removal

Of the methods available, sonication was employed as the system for lysis of all samples of sufficient volume, as initial work indicated that this resulted in a greater proportion of soluble, target protein compared with detergent lysis. It is possible that the amphiphilic nature of the detergent environment causes the surfactant proteins to denature, or it may simply be the case that the mechanical sonication provided more thorough cell disruption, releasing more protein into the aqueous environment. Insoluble cell debris was removed by high speed centrifugation and careful decanting of the supernatant, leaving a solution containing the protein of interest along with the remaining soluble cell components.

2.1.3.2 Ni^{2+}-Affinity Chromatography

Of the various purification techniques available, Ni^{2+}-affinity chromatography is one of the most popular due to its simplicity and reliability [129]. It utilises the high affinity of histidine residues for Ni^{2+}, via electron donor groups on the imidazole side chain. A series of six consecutive histidines were encoded within the expression plasmid for expression at the N-terminus of the target protein to produce a polyhistidine fusion tag (His-tag). By using immobilised metal affinity chromatography (Ni^{2+} bound within a nitrilotriacetic acid (Ni^{2+}-NTA) column) the His-tagged protein was retained on the column and hence isolated from untagged proteins. The bound protein was then eluted by washing with a solution containing a high concentration of imidazole (300mM), which displaces the bound histidines. Any contaminant proteins that have a natural affinity for the Ni^{2+} ions were removed by washes with solutions containing low imidazole concentration (5-30mM) prior to elution of the target protein. Unspecific binding via electrostatic interactions was inhibited by the presence of NaCl in solutions.

2.1.3.3 Fusion Tag Cleavage and Removal

Although fusion tags are useful for the initial purification from the bulk cell debris, it is desirable that they be removed as there is a risk that they may interfere with subsequent work. The presence of a fusion tag is especially undesirable during NMR structure elucidation as the extra spectral features arising from the additional amino acids complicates the already difficult assignment and structure calculation processes. The expression vectors used within this work contained a protease recognition site between the fusion tags and the multiple cloning site (MCS). These protease sites were utilised for convenient tag cleavage.

Once the tag had been cleaved from the protein it was removed by performing Ni^{2+}-affinity chromatography once more. The cleaved tag bound to the resin and the protein passed straight through. This also acted as a convenient purification step as the majority of contaminants that bound and eluted alongside the protein of interest during the first Ni^{2+}-affinity purification were retained once again and hence removed for the target protein.

2.1.3.4 Size-exclusion Chromatography

The protein of interest was by this stage relatively pure, but to remove any remaining contaminants size-exclusion chromatography (SEC) was used. This is a chromatographic method where a sample is passed through an inert gel matrix containing a range of pore sizes. The molecules' progression through the column is retarded by their ability to enter the matrix: large molecules, that enter no or very few pores, are eluted first, small molecules, which enter many pores, are eluted last. This step was useful to remove contaminants such as small molecule buffer components or proteins of dissimilar molecular weight.

2.2 Rsn-1

High yields of pure Rsn-1 were required for planned NMR structure elucidation and biophysical characterisation work.

2.2.1 Existing Rsn-1 Construct

In previous work Dr. Mackenzie used an *Rsn-1*/pCRT7/BL21(DE3) construct to produce recombinant Rsn-1 for initial NMR experimentation & protease inhibition assays [130]. The levels of expression achieved with this construct were insufficient for the production of high concentration samples required for structure determination by NMR. A higher yielding construct was required in this work.

2.2.2 Rare Codon Hypothesis

The genetic sequence of *Rsn-1* was investigated for indications of the cause of the problems in expression. Eukaryotic genes often contain codons that are rarely used in *E.coli.* When the frequency of these codons is low the *E.coli* can normally cope and produce the desired protein. However, in cases where rare codon usage is high or where these codons appear sequentially, stalled translation can occur and cause truncated or failed protein expression [131]. The sequence (Figure 2.1) was inspected for the presence of rare codons. Of the 109 amino acid residues represented, 14 were identified as rare in *E.coli*. Also noted was the succession of three rare codons at positions 74, 76 & 77 in the sequence as it has been identified that expression problems become prevalent when rare codons are found in clusters, such as doublets and triplets [132]. It was therefore

hypothesised that the high level of rare codons in *Rsn-1* could be responsible for the low expression levels observed thus far.

```
1.   GGA GGT GGT AAC ATT GGA GGT GGT GCA AAG
11.  TTG GGA CCA GAA AAG CCG GCA ACC CCA GGG
21.  ATA CAA GAT CTC CTG AAG TCG CTT TTG TCT
31.  GTG CTG AAT CTC AGT CCT CCG GCT ATT CCT
41.  GAA GAT GCG GAG GCC GTC TCC TAC AGA GAT
51.  GCA AAG AAC GGC AAA TTT AGG TTA ATC AAG
61.  ATC CAT CTC GGA GGG GAG CTG TAC TGT CAT
71.  GTG AAG CAA ATA GCA GGA CCC ATC CTG GCT
81.  CTA CCT ATC GTC TCA GAT GTA GTG GAA GTT
91.  ACT GGA AAA GAA TGT GGG AAA ACC GAG GAC
101. GAC CCT CTC GAG GAT TTC CCC ATT CCC
```

Figure 2.1- DNA sequence of Rsn-1 gene. Codons defined as rare within *E.coli* are shown in bold and underlined [131].

To investigate this hypothesis the *Rsn-1*/pCRT7 vector was transformed into both BL21(DE3) (Novagen, #69450) and Rosetta(DE3) (Novagen, #70954) expression cells. The Rosetta(DE3) cells are derived from the BL21 strains but contain an additional plasmid that possess genes which supply the rare tRNAs [133]. By increasing the number of the limiting codons, *E.coli* can adapt to express genes containing these rare codons to increased levels [134]. The Rosetta cells were shown to improve the level of expression considerably compared to the original BL21(DE3) construct (Figure 2.2).

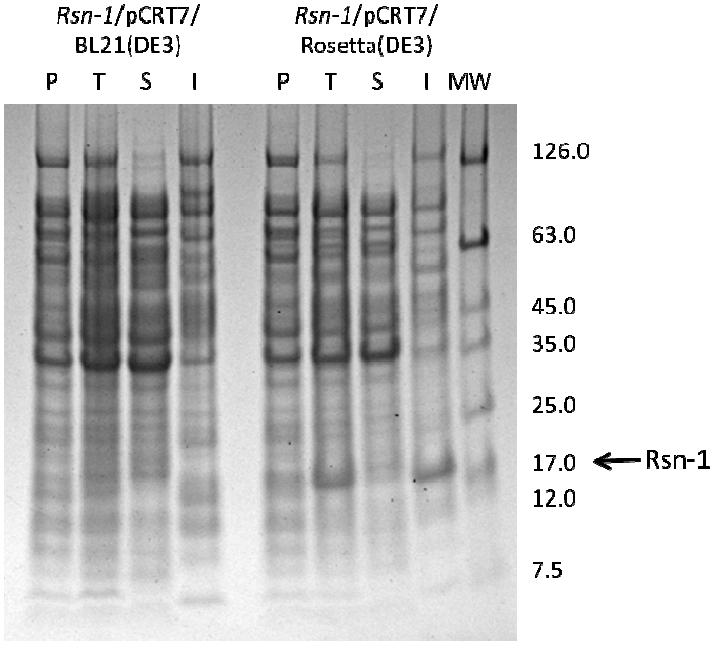


Figure 2.2- SDS-PAGE analysis of Rsn-1 construct comparison. The identity of each construct is shown above the gel. P= preinduction sample, T= total sample, S= soluble sample & I= insoluble sample. MW= molecular weight markers (NEB, #P7708). Cultures were induced with IPTG at a

final concentration of 1 mM and left to express at 37℃ for 4 h. The fusion tagged Rsn-1 expression band can be observed adjacent to the 17kDa MW marker. There is no indication Rsn-1 being over-expressed in the *Rsn-1*/pCRT7/BL21(DE3) construct, whereas the *Rsn-1*/pCRT7/Rosetta(DE3) construct does produce Rsn-1.

An alternative solution to the rare codon problem was to create a synthetic *Rsn-1* (*sRsn-1*) gene. This is an increasingly popular practice as it allows the user to create a gene/construct specific to their requirements. The Geneart Company was employed to produce the *sRsn-1* gene. The company's GeneOptimizer software was used to design the optimised gene [135]. The primary benefit was the optimisation of codon usage for expression where rare codons are replaced with more favourable alternatives for expression in *E.coli*. The gene was also optimised with regards to other factors such as RNA stability and GC content. NcoI and HindIII restriction sites were added flanking the gene, to allow simple future subcloning. The gene was supplied in cloning vector pGA-18 (Figure 2.3).

Figure 2.3- Vector map of *sRsn-1* plasmid. Cloning vector containing *sRsn-1* gene and requested NcoI and HindIII restriction sites.

2.2.3 *sRsn-1*/pET-30 Expression Construct

Expression plasmid pET-30 (Novagen, #69909) was selected as the destination vector as it had been successfully utilised in the past in the group (Figure 2.4). The *sRsn-1* gene was subcloned into the pET-30 plasmid using the NcoI & HindIII restriction sites.

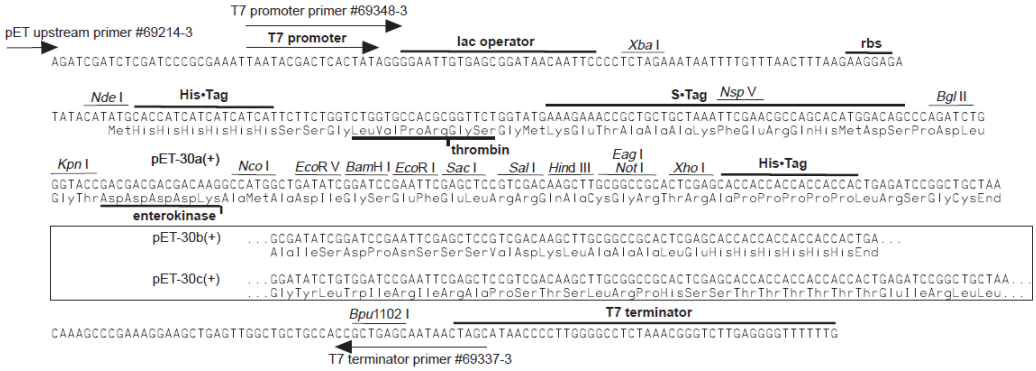


Figure 2.4- pET-30 cloning/expression region. Image reproduced from Novagen pET-30 vector map (Novagen, #TB095).

Expression of this construct produced Rsn-1 with cleavable N-terminal S•Tag and His-tag that increased the molecular weight of Rsn-1 from 11.5kDa to 16.3kDa (Figure 2.5). As discussed already the His-tag is utilised in the purification of the protein by Ni^{2+}-affinity chromatography. The S-Tag is an additional fusion tag encoded in the pET-30 plasmid. The S-Tag itself is a 15 amino acid peptide derived from pancreatic ribonuclease A. The tag can be utilised for purification or detection of recombinant protein and therefore offers an alternative to the His-tag. In this work the S-Tag was not utilised. Sequencing was carried out confirming the gene had been inserted into the vector as desired.

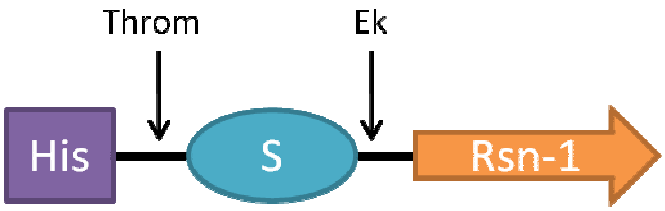


Figure 2.5- Diagram of *sRsn-1*/pET-30 expression product. Where S= S-Tag and His= His-tag represent N-terminal fusion tags. Throm= thrombin protease recognition sequence and Ek= enterokinase protease recognition sequence.

The expression of *sRsn-1*/pET-30/BL21(DE3) was compared to that of the *Rsn-1*/pCRT7/BL21(DE3) and *Rsn-1*/pCRT7/Rosetta(DE3). The construct containing the synthetic Rsn-1 gene gave substantially higher levels of expression and was hence selected as the superior construct and taken forward into expression optimisation.

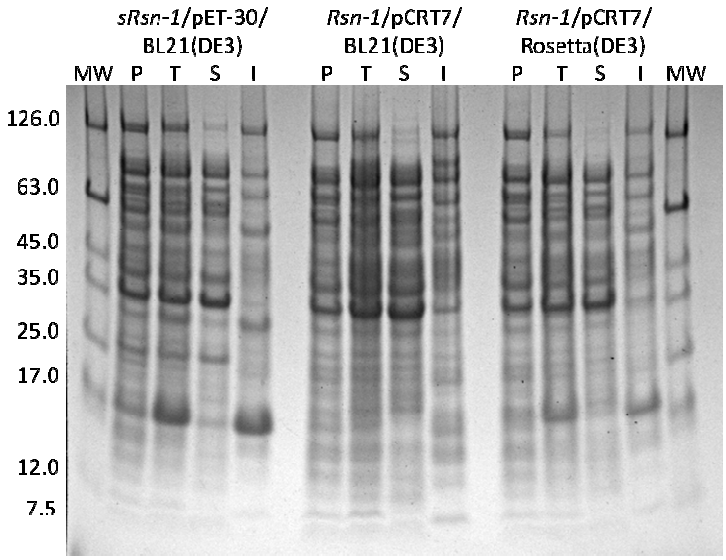


Figure 2.6- SDS-PAGE analysis of synthetic and native Rsn-1 construct comparison. Where the identity of each construct is shown above the gel. P= preinduction sample, T= total sample, S= soluble sample & I= insoluble sample. MW= molecular weight markers (NEB, #P7708). Cultures were induced with IPTG at a final concentration of 1 mM and left to express at 37℃ for 4 h. The fusion tagged Rsn-1 expression band can be observed adjacent to the 17kDa MW marker. The sRsn-1/pET-30/BL21(DE3) construct produces significantly greater amounts of Rsn-1 than either of the constructs encoding the native Rsn-1 gene.

2.2.4 *sRsn-1*/pET-30 Expression

Although the *sRsn-1*/pET-30 vector produced the greatest amount of Rsn-1 (Figure 2.6), the expressed protein was almost exclusively insoluble (Figure 2.7). Insoluble protein is a common consequence of a host cell's inability to handle the over-expression of heterologous proteins [136]. The misfolded protein aggregates and is incorporated into inclusion bodies. Although it is possible to purify the protein from inclusion bodies, the refolding process must be optimised for each protein which can often develop into a time-consuming and expensive process. Therefore, it is preferable to obtain soluble, correctly folded protein during expression.

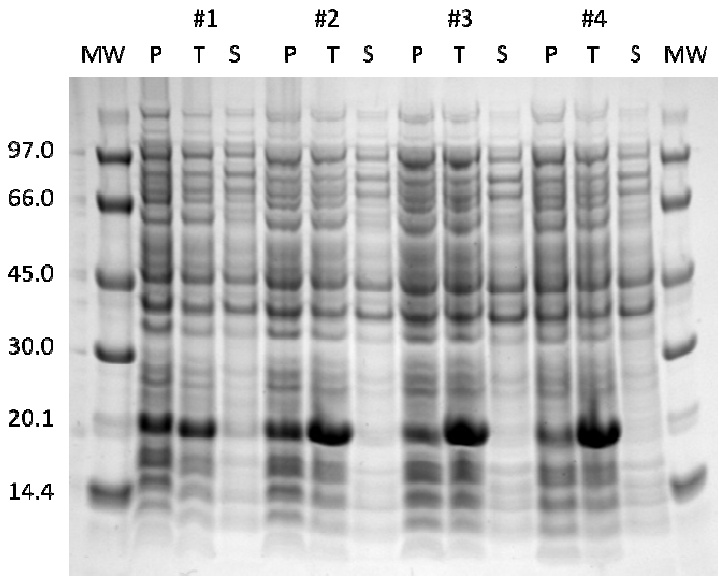


Figure 2.7- SDS-PAGE analysis of *sRsn-1*/pET-30/BL21(DE3) expression at 37℃. Where #1, 2, 3 & 4 refer to the four cultures assessed. P= preinduction sample, T= total sample & S= soluble sample. MW= molecular weight markers (GE Healthcare, #17-0446-01). Cultures were induced with IPTG at a final concentration of 1 mM and left to express at 37℃ for 4 h. The Rsn-1+tag expression band can be observed adjacent to the 20.1kDa MW marker. No soluble Rsn-1 was produced in any of the cultures.

Many techniques can be utilised to increase the levels of soluble protein produced. Fusion tags to improve solubility of the target protein, optimised codon usage, specialised host cells and co-expression of molecular chaperones are just a few examples of commonly utilised techniques. Another approach involves altering the expression conditions to reduce the rate of cellular metabolism. This is postulated to aid solubility by aligning the rates of transcription, translation and post-translational processing [137]. In an attempt to increase the amount of soluble protein produced, the optimisation of various expression conditions such as temperature, IPTG concentration, length of expression were carried out. The most significant modification was the reduction of expression temperature from 37 to 20°C, which produced a greater proportion of soluble Rsn-1 (Figure 2.8) than expressions carried out at higher temperatures.

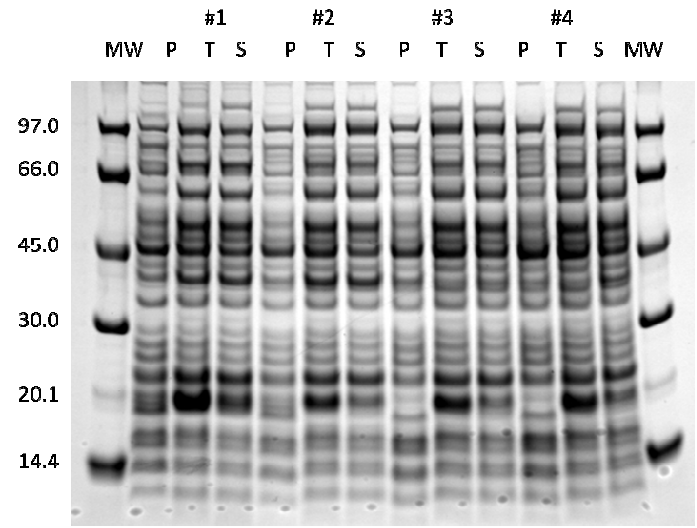


Figure 2.8- SDS-PAGE analysis of *sRsn-1*/pET-30/BL21(DE3) expression at 20℃. #1, 2, 3 & 4 refer to the four cultures assessed. P= preinduction sample, T= total sample & S= soluble sample. MW= molecular weight markers (GE Healthcare, #17-0446-01). Cultures were induced with IPTG at a final concentration of 1 mM and left to express at 20℃ for 18 h. The Rsn-1+tag expression band can be observed adjacent to the 20.1kDa MW marker. Some Rsn-1 is observed in the soluble fraction. Culture #1 in particular produced a reasonable amount of soluble protein and was hence used in subsequent expressions.

Although a large proportion of the Rsn-1 expressed remained in the insoluble fraction, it was proposed that the 100-1000 μg of soluble protein produced may be sufficient for NMR experimentation.

2.2.5 Confirming Identity of Rsn-1

Upon production of soluble Rsn-1, it was deemed wise to confirm that the protein that had been over-expressed was definitely Rsn-1, as well as folded, before further effort and expense was expended to produce an isotopically labelled sample. One common technique for identification or detection of proteins is Western Blotting. Proteins are initially separated by mass using SDS-PAGE, before transfer to a membrane for further analysis. The membrane is probed using antibodies, specific to the target protein, conjugated to a signal enzyme. As Rsn-1 specific antibodies were not available, anti-polyhistidine antibodies, commonly available from commercial suppliers were used. These antibodies are specific to the motif of six consecutive histidines, characteristic of His-tags, allowing for the detection of polyhistidine tagged recombinant proteins. As discussed in section 2.2.3, the recombinant plasmid DNA created for the expression of Rsn-1 was designed to produce Rsn-1 with both His and S fusion tags. Insertion of the Rsn-1 gene with the correct location and orientation into the pET-28 plasmid was confirmed by DNA sequencing,

therefore Rsn-1 should be the only protein in the *E.coli* cell lysate that contains a polyhistidine motif.

A test expression carried out using the sRsn-1/pET-30/BL21(DE3) construct was analysed by both SDS-PAGE and Western Blot. The SDS-PAGE indicated an over-expression band in both the total and insoluble fractions, at a molecular weight corresponding to that expected for Rsn-1 + fusion tag (16.3 kDa) (Figure 2.9). The Western Blot carried out with monoclonal anti-polyhistidine-alkaline phosphatase (Sigma, #A5588) identified the over-expressed protein as containing a polyhistidine motif (Figure 2.9). Together, these results indicate that the over-expressed protein is highly likely to be recombinant Rsn-1.

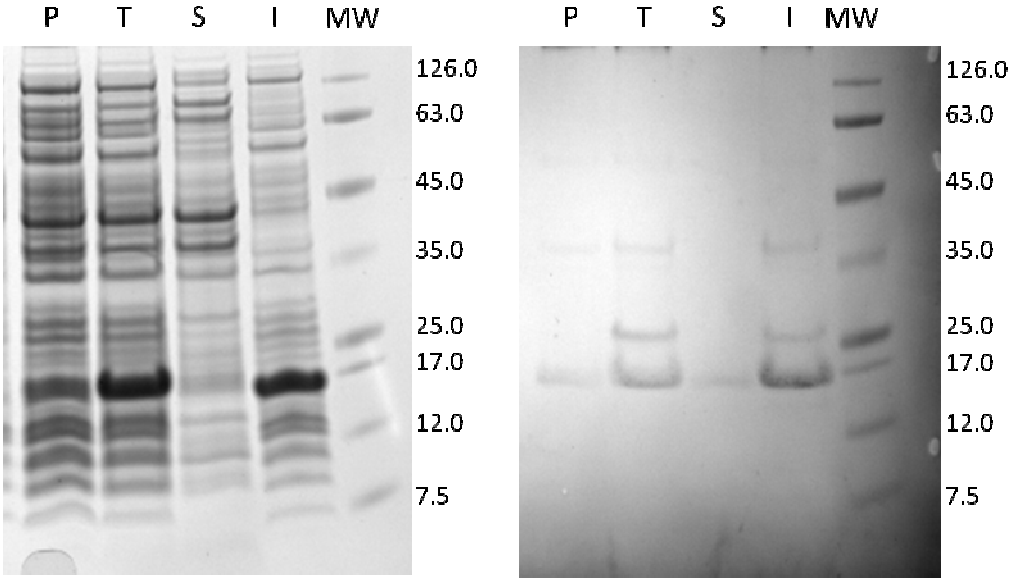


Figure 2.9- SDS-PAGE gel and Western Blot of Rsn-1 expression. The *sRsn-1*/pET-30/BL21(DE3) cultures were induced with IPTG at a final concentration of 1 mM and left to express at 37°C for 4 h. P= preinduction sample, T= total sample, S= soluble fraction & I = insoluble fraction. MW= molecular weight markers with values in kDa (NEB, #P7708). A clear over expression band is observed in the total sample and insoluble fraction of the gel at a MW corresponding to that of the tagged Rsn-1 (16.3 kDa). The western blot indicates that the anti-polyhistidine antibody binds to the over expressed protein and therefore that it contains the polyhistidine motif.

To confirm this possibility, the identity of the over-expressed protein was further analysed by peptide mass fingerprinting. In this technique the protein of interest is cleaved into smaller peptides using a protease. The peptide fragments are then analysed by mass spectrometry and the results compared to a database containing theoretical peptide fragments of proteins. A best match is then identified to predict the protein's identity. In this work peptide mass fingerprinting was carried out by Dr. Richard Burchmore (Infection, Immunity and Inflammation Life Sciences, University of Glasgow). A band of the over-expressed protein was cut from a SDS-PAGE gel and supplied to Dr. Burchmore.

Proteolysis was carried out using trypsin and the resultant peptides analysed by MALDI-TOF (matrix-assisted laser desorption/ionization-time of flight) mass spectrometry. No match to the peptide profile was detected in any database, nor did the observed peptide mass fragments match those predicted from the Rsn-1 sequence. The results were both unexpected and unexplainable.

An alternative approach for identifying the recombinant protein using 1D NMR was undertaken. Although 1D NMR alone cannot be used for protein identification, spectra recorded from a sample created by Dr. MacKenzie and confirmed as Rsn-1 by MALDI-TOF spectroscopy, was available for comparison [130]. A protein sample was produced from the *sRsn-1*/pET-30/BL21(DE3) construct and soluble, over-expressed protein was purified by a single Ni^{2+}-affinity chromatography step. To save time, the fusion tag was not cleaved as it should not affect the features of the 1D spectrum that were of interest. The tagged Rsn-1 was passed through size exclusion chromatography to remove any remaining impurities, buffer exchanged into 20mM NaPi, 150mM NaCl, pH 7.5, 5% D_2O and a 1D spectrum recorded.

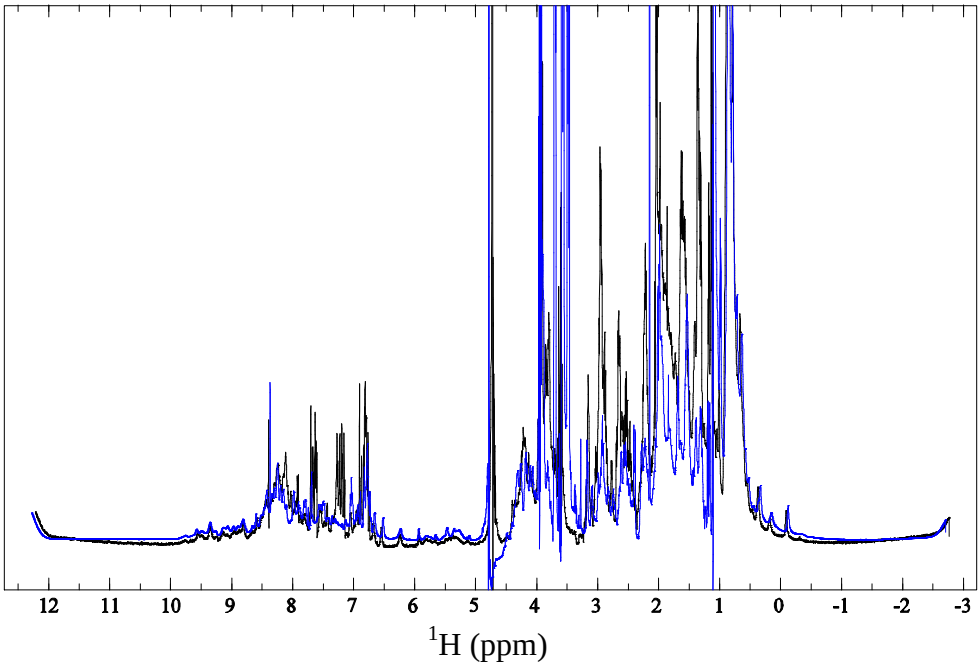


Figure 2.10- 1D ^{1}H NMR of Rsn-1 samples. Scale in ppm. Sample produced from sRsn-1/pET-30/BL21(DE3) construct is shown in black. Rsn-1 sample produced by Dr. MacKenzie is shown in blue. The well dispersed signals in the methyl (-0.5–1.5ppm) and the amide (7-10ppm) regions are indicative of folded protein. Minor differences in the two spectra can be rationalised to variation in fusion tags or contaminants. The two spectra match well confirming that the purified protein is Rsn-1. Experiments were conducted with double-pulse field gradient water suppression [138].

The spectra from the two samples are almost identical, with any minor differences likely to be a consequence of differences in the attached fusion tag or co-purified contaminants. The

NMR data, therefore, indicated that the protein was indeed Rsn-1. As the gene sequencing, Western Blot and NMR results all implied that the expressed protein was Rsn-1, it would appear that for some reason, the peptide mass fingerprinting incorrectly failed to identify the protein as Rsn-1. As there was sufficient convincing evidence that the over-expressed protein was indeed Rsn-1 the peptide mass fingerprinting was not repeated.

A 1D NMR spectrum is also indicative of the structure of the sampled protein. The resonant frequencies of the NMR active nuclei within the protein are highly dependent upon the local environment created by secondary, tertiary and quaternary structure. Therefore, an unfolded protein shows low dispersion of chemical shifts in nuclei such as protons in backbone amide or aliphatic groups [139]. In particular, a broad peak at a chemical shift of approximately 8.3 ppm is a strong indicator of disordered protein, as this is a characteristic shift of backbone amides in random-coil. Alternatively, dispersed signals beyond 8.5 ppm are proof of varying chemical environments experienced by amide protons in a folded protein. Similarly a range of signals within the -1.0 to 1.0 ppm region, as oppose to a broad peak at 1.0 ppm, is indicative of aliphatic protons in a folded protein. Figure 2.10 shows dispersed resonance frequencies in both the aliphatic (-1.0 to 1.0 ppm) and the amide (7-10 ppm) regions confirming that Rsn-1 is folded.

2.2.6 Rsn-1 Protease Cleavage

As mentioned earlier, the *sRsn-1*-/pET-30 construct codes for Rsn-1 protein with attached polyhistidine and S·Tag fusion tags which are cleavable via two protease cleavage sites: a thrombin cleavage site between the His-tag and the S-tag and an enterokinase cleavage site between the S-tag and Rsn-1 (Figure 2.5). By utilising the enterokinase site, the protease cleavage should produce untagged Rsn-1 with only two “cloning artefact” residues, an alanine and methionine, at the N-terminus.

Protease cleavage was carried out with enterokinase as described in the materials and methods section. However, the cleavage did not result in two products of mass 11.5kDa and 4.8kDa (the mass of the cleaved tag) as expected but at least three different species (Figure 2.11).

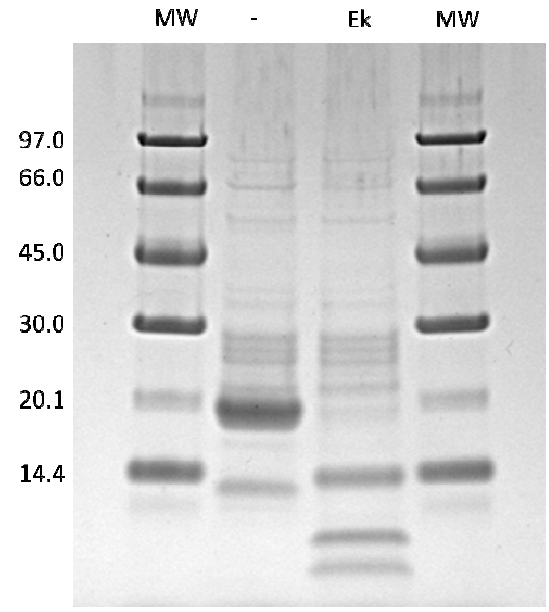


Figure 2.11- SDS-PAGE analysis of Rsn-1 enterokinase cleavage. Where - = the tagged Rsn-1 with no protease control & Ek = tagged Rsn-1 incubated with enterokinase. Protease cleavage was carried out at 4°C for 48 h with 1U enterokinase per mg of protein. MW= molecular weight markers (GE Healthcare, #17-0446-01). The tagged Rsn-1 band, aligned with the 20.1kDa marker is converted into three clear bands by enterokinase.

It was concluded that Rsn-1 must be being cut at some alternative enterokinase cleavage site within its sequence, which is a common observation in proteins where the alternative site is more accessible than the primary recognition site [140]. Various conditions such as temperature and protease concentration were trialled to investigate if the specificity of the protease could be restricted to the desired site, without success. It was concluded that an alternative strategy would be to replace the enterokinase cleavage site with an alternative protease site. A selection of common proteases: thrombin (Novagen, #69671-3), Factor Xa, (Novagen #69036-3) and HRV 3C (Novagen #71493) were trialled to ensure that Rsn-1 did not contain any other hidden cleavage sites. SDS-PAGE was used to analyse the products of incubation of Rsn-1 with each protease (Figure 2.12).

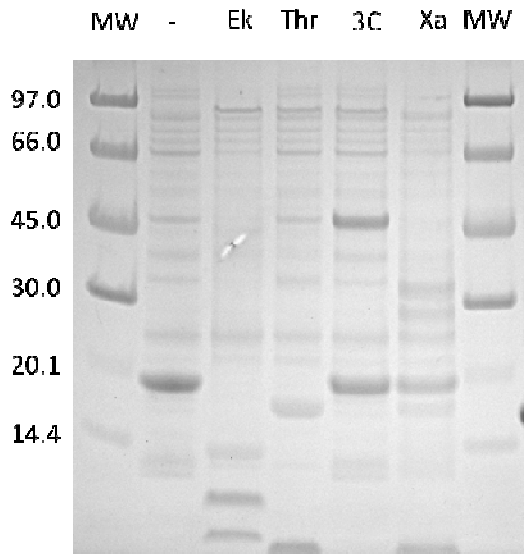


Figure 2.12- SDS-PAGE analysis of Rsn-1 with various proteases. – = the tagged Rsn-1 with no protease control, Ek= enterokinase cleavage, Thr= thrombin cleavage, 3C= HRV 3C cleavage & Xa= factor Xa cleavage. MW= molecular weight markers (GE Healthcare, #17-0446-01). All reactions were carried out under optimal conditions as described by the suppliers for a period of 48 h..

The enterokinase cleavage again showed multiple products confirming what had been observed already. The thrombin digest produces protein bands in agreement with the theoretical products (14.7kDa for Rsn-2+S•Tag and 1.7kDa for cleaved His-tag). Despite the absence of a factor Xa recognition sequence (Ile-Glu/Asp-Gly-Arg↓) within the tagged Rsn-1 sequence, factor Xa causes partial cleavage producing products very similar to that of the thrombin protease. The ability of factor Xa to cleave at the same recognition sequence as thrombin has been reported and would explain this observation [141]. This, as with thrombin, makes it difficult to completely rule out secondary cleavage sites. An additional disadvantage to the insertion of either a thrombin or factor Xa recognition sequence in place of enterokinase was the existing thrombin site. Thrombin or factor Xa cleaving at both the existing thrombin and replacement site would produce three products: His-tag, S•tag and Rsn-1. Of which only the His-tag could be easily extracted. The original thrombin site could be mutated or removed altogether, but this would require an additional, potentially time-consuming step. HRV 3C was the only protease tested to have no effect upon Rsn-1 and was therefore selected as the preferred replacement for enterokinase.

HRV 3C protease is a cysteine protease found in Human Rhinovirus (HRV), the main cause of the common cold. Recombinant 3C is commonly used in protein purification due to its high selectivity for its recognition site (LeuGluValLeuPheGln↓GlyPro) [142]. The existing enterokinase site within the sRsn-1/pET-30 construct was flanked by KpnI and

NcoI restriction sites (Figure 2.4) and therefore could be easily removed and replaced with the relevant 3C recognition site.

Primers were designed to contain the 3C recognition sequence with KpnI and NcoI "sticky end" overhangs to allow simple cloning into the desired vector (Figure 2.13). The resultant DNA was then ligated into a KpnI and NcoI digested Rsn-1/pET-30 vector. Sequencing confirmed the construct was as designed.

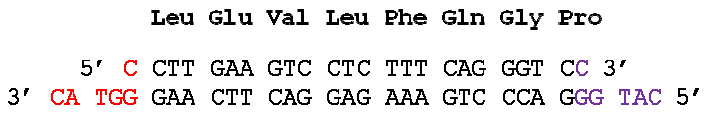


Figure 2.13- HRV 3C Primers. The DNA sequence of the two complementary primers and the translated amino acid sequence: a 3C protease recognition sequence. The KpnI and NcoI "sticky end" overhangs are shown in red and purple respectively.

2.2.7 *sRsn-1*/pET-30(3C) Expression

The *sRsn-1*/pET-30(3C) construct was transformed into Tuner(DE3) expression cells. Tuner(DE3) cells are a BL21(DE3) based strain that contain all the characteristics of BL21(DE3) with the advantage of greater control over expression levels through IPTG concentration. This is a consequence of a mutation in the lac permease gene that encodes a membrane bound protein and pumps lactose into the cell. The mutation allows uniform entry of IPTG into all cells in the culture, which allows for a homogeneous, concentration dependant, control over expression level. Initial expression trials were carried out. Results were, as expected, similar to that observed for *sRsn-1*/pET-30: high levels of over-expression, only a small proportion of soluble protein (Figure 2.14).

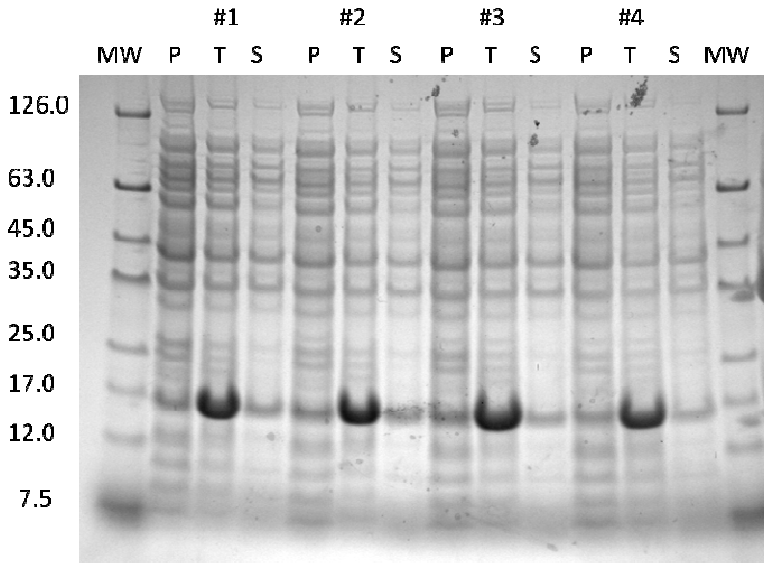


Figure 2.14- SDS-PAGE analysis of *sRsn-1*/pET-30(3C)/BL21(DE3) expression. Where #1, 2, 3 & 4 refer to the four cultures assessed. P= preinduction sample, T= total sample & S= soluble sample. MW= molecular weight markers with values in kDa (NEB, #P7708). Cultures were induced with IPTG to a final concentration of 1mM and left to express at 20°C for 16 h. The Rsn-1+tag expression band can be observed adjacent to the 17.0kDa MW marker. Some Rsn-1 is observed in the soluble fraction for all four cultures.

Although this amount of Rsn-1 may have been sufficient for NMR experimentation the levels were significantly reduced when expression was carried out in the minimal media required for isotopic labelling. As a consequence good amounts of unlabelled Rsn-1 were produced for characterisation work but sufficient amounts of isotopically labelled Rsn-1 to allow structure determination by NMR were never achieved.

2.3 Rsn-2

Two Rsn-2 expression constructs were used to produce recombinant Rsn-2 at different stages throughout the project. Initial work was carried out using the *Rsn-2*/pCRT7/BL21(DE3) construct created and used by Dr. MacKenzie in previous work [130].

However, at the initiation of the mutagenesis project it was decided that work to create a new expression construct would be relatively trivial and could potentially improve Rsn-2 yield, reducing time and expense of protein production.

2.3.1 Rsn-2 Constructs

2.3.1.1 Existing *Rsn-2*/pCRT7 Expression Construct

The existing Rsn-2 expression vector consisted of *Rsn-2* cloned into the pCRT7/NT-TOPO vector (Invitrogen, #K4200-01). This plasmid allows for a single, straight-forward cloning step directly following PCR utilising the TOPO system. Expression of the *Rsn-2*/pCRT7 construct produced Rsn-2 with a thrombin cleavable His-tag (Figure 2.15). The pCRT7/NT-TOPO vector is supplied with an enterokinase recognition sequence, however this was modified to thrombin by Dr. MacKenzie who found that enterokinase cleaved at a secondary site within the Rsn-2 sequence [130].

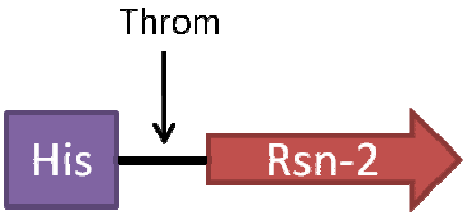


Figure 2.15- Diagram of *Rsn-2*/pCRT7 expression product. Where His= His-tag and Throm= thrombin protease recognition sequence.

Dr MacKenzie found optimal conditions for protein expression to be induction with IPTG at a final concentration of 1 mM and expression at 37°C for 4 h. Recombinant protein was then purified by Ni-affinity chromatography followed by thrombin protease cleavage of the fusion tag. Size exclusion chromatography was then carried out as a final polishing step.

The expression and purification procedures designed by Dr MacKenzie were followed in this work. Yields of 10-12 mg/L of culture media were achieved under these conditions. The identity of the purified protein was confirmed as Rsn-2 by NMR.

2.3.1.2 *Rsn-2*/pET-28 Expression Construct

The new Rsn-2 expression product was created by PCR. Two primers were designed: one to anneal to the 5' end of the complementary strand of the Rsn-2 gene and containing an NdeI restriction site (Figure 2.16), the other to the 3' end of the coding strand incorporating a BamHI restriction site (Figure 2.17).

```
                 His Met Leu Ile Leu Asp Gly Asp Leu Leu
Rsn2_FL_5p  5' CAT ATG TTA ATA TTA GAT GGG GAC CTA CTA 3'

 5'... GAT CCA ACC CTT TTA ATA TTA GAT GGG GAC CTA CTA AAG GAC ...3'
 3'... CTA GGT TGG GAA AAT TAT AAT CTA CCC CTG GAT GAT TTC CTG ...5'
   ... Asp Pro Thr Leu Leu Ile Leu Asp Gly Asp Leu Leu Lys Asp ...
```

Figure 2.16- Rsn2_FL_5p primer and its annealing site on the Rsn-2 gene. The NdeI restriction site is underlined.

```
  ... Cys Arg Lys Asp Asp Asp Asp Asp Asp Gly Tyr Stop

5'... TGT AGG AAG GAC GAT GAT GAT GAT GAT GGA TAT TAG AAG GGC ... 3'
3'... ACA TCC TTC CTG CTA CTA CTA CTA CTA CCT ATA ATC TTC CCG ... 5'

       Rsn2_FL_3p  3' CTA CTA CTA CTA CTA CCT ATA ATC CTA GG 5'

                   Asp Asp Asp Asp Asp Gly Tyr Stop
```

Figure 2.17- Rsn2_FL_3p primer and its annealing site on the Rsn-2 gene. The BamHI restriction site is underlined.

PCR was carried out with these primers using the Rsn-2/pCRT7 construct as a template to produce an Rsn-2 insert flanked by NdeI and BamHI restriction sites. The gene was first cloned into pCR2.1-TOPO (Invitrogen, # K4510-20) before it was transferred into pET-28 expression plasmid (Novagen, #69864) using NdeI and BamHI. pET-28 was selected for expression trials as it contains a thrombin rather than an enterokinase recognition sequence between the His-tag and the MCS (Figure 2.18). This was necessary as there is a secondary enterokinase cleavage site within Rsn-2 [130]. DNA sequencing confirmed the construct was as designed.

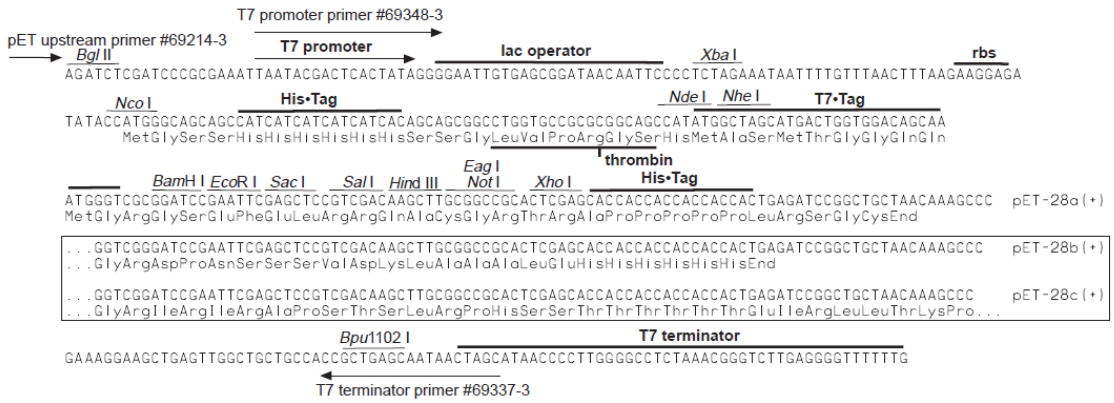


Figure 2.18- pET-28 cloning/expression region. Image reproduced from Novagen pET-28 vector map (Novagen, #TB074).

2.3.2 Rsn-2 Expression

The Rsn-2/pET-28 plasmid was transformed into BL21(DE3) cells. Initial expression trials were carried out under the conditions found previously to be optimal for the *Rsn-2*/pCRT7 construct (induction with IPTG at a final concentration of 1 mM and expression at 37°C for 4 h), and gave high production levels of recombinant Rsn-2 (Figure 2.19) [130]. Various adjustments to the expression conditions such as IPTG concentration, temperature and length of induction were investigated, but no significant improvement relative to the original conditions was observed.

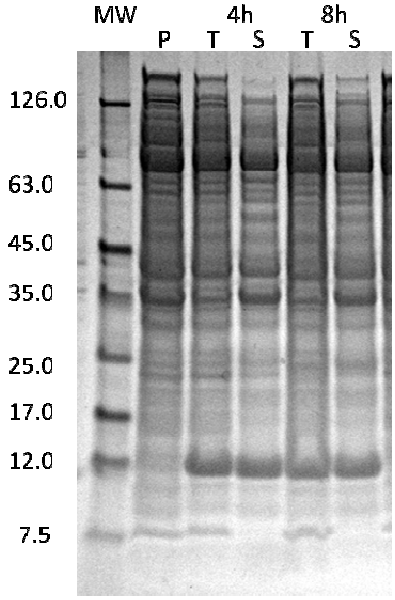


Figure 2.19- SDS-PAGE analysis of *Rsn-2*/pET-28/BL21(DE3) expression 4 and 8h refer to the length of time following addition of IPTG. P= preinduction sample, T= total sample and S= soluble fraction. MW= molecular weight markers (NEB, #P7708). Cultures were induced with IPTG at a final concentration of 1 mM and left to make Rsn-2 at 37°C with samples taken 4 and 8 h after induction. The Rsn-2 expression band can be observed adjacent to the 12 kDa MW marker. All Rsn-2 was located in the soluble fraction. No increase in Rsn-2 level was observed after 4h.

2.3.3 Rsn-2 Purification

Relatively pure Rsn-2 could be obtained from the cell lysate by Ni^{2+}-affinity chromatography.

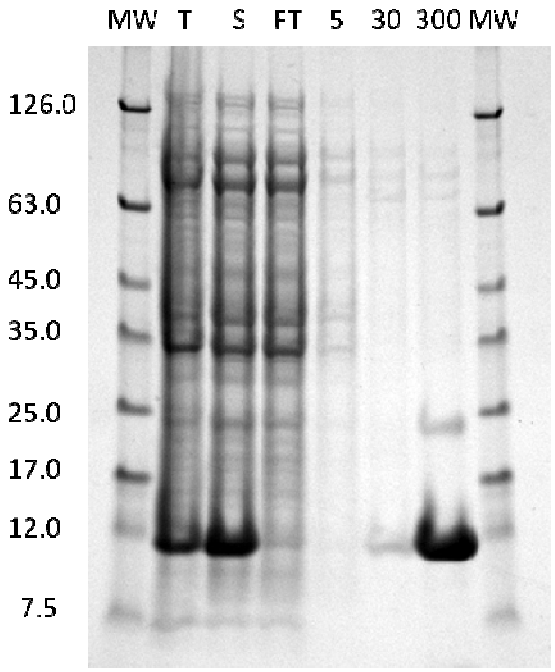


Figure 2.20- SDS-PAGE analysis of Rsn-2 Ni^{2+}-affinity chromatography. T= total sample, S= soluble fraction, FT= flow-through, 5= 5mM imidazole wash, 30= 30mM imidazole wash, 300= 300mM imidazole wash and MW= molecular weight markers with values in kDa (NEB, #P7708). All Rsn-2 (adjacent to the 12 kDa MW marker) binds to the column and is eluted as expected in the 300 mM imidazole wash.

Following Ni^{2+}-affinity chromatography it was deemed wise to remove the fusion tag to ensure its presence did not affect subsequent activity assays or make the NMR spectra unnecessarily complicated. Protease cleavage of the fusion tag was carried out according to the optimal protocol developed previously by Dr MacKenzie [130]. In brief, Rsn-2 was incubated with 5 units of restriction grade thrombin (Novagen, #69671-3) at 4°C. The reaction was monitored by SDS-PAGE until cleavage was complete. Two cleavage products were observed corresponding to the expected masses of Rsn-2 and the cleaved fusion tag (Figure 2.21). The cleaved tag was then removed by passing the completed thrombin reaction over a fresh Ni^{2+}-affinity column.

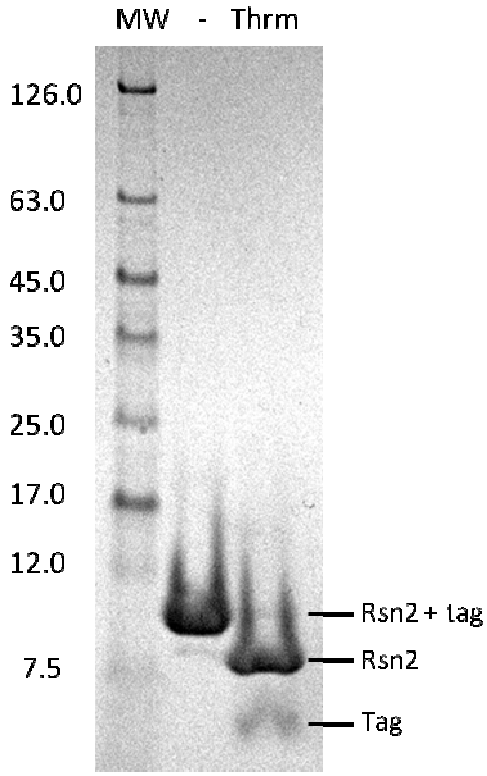


Figure 2.21- SDS-PAGE analysis of Rsn-2 thrombin cleavage. - = negative control and Thrm = thrombin reaction. MW= molecular weight markers with values in kDa (NEB, #P7708). The cleavage reaction results in two major products, a band between the 7.5 and 12 kDa markers and a band below the 7.5 kDa marker. These products correspond well to the expected masses of Rsn-2 (11.5 kDa) and cleaved tag (1.9 kDa).The reverse-purified Rsn-2 was passed down a size exclusion chromatography column as a final polishing step to remove any residual small molecule and/or protein contaminants. A single elution peak at a volume expected for a molecule of Rsn-2's molecular weight was observed. The results of SEC are analysed and discussed in detail in Chapter 3.

2.3.4 Confirming identity of Rsn-2

The purified protein was confirmed as Rsn-2 by 2D NMR. A ^{15}N-HSQC spectrum recorded from a ^{15}N labelled Rsn-2 sample produced from the *Rsn-2*/pET-28/BL21(DE3) construct was compared to previous spectra recorded by Dr. MacKenzie during the structure determination process [62]. The two spectra were essentially superimposable and subsequent assignment of the spectrum confirmed that the protein produced in this work was Rsn-2.

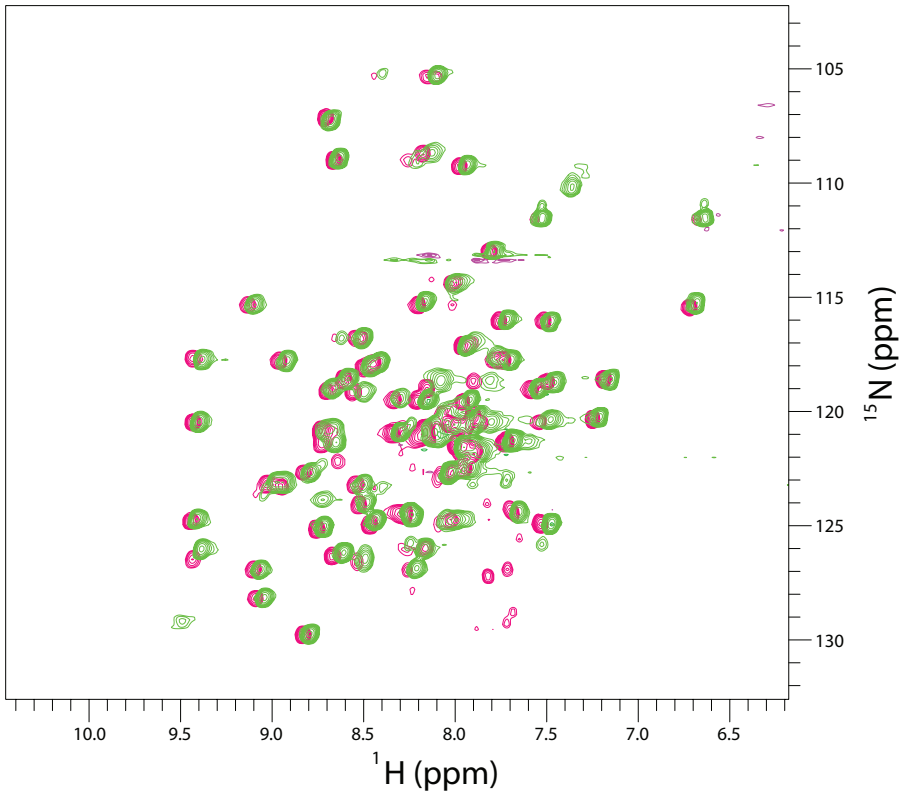


Figure 2.22- ^{15}N-HSQC spectra of Rsn-2 samples. The spectrum from the Rsn-2 sample created from the *Rsn-2*/pET-28/BL21(DE3) construct in this work (green) is essentially identical to the spectrum recorded by Dr. MacKenzie during the structure determination process (pink). The slight change in chemical shift in the ^{1}H dimension is likely to be a result of minor differences in buffer composition between the two NMR samples. The sharp peaks at a 1H shift of 7.5-8.0 ppm and 15N shift of 125 to 130 ppm are almost certainly due to the presence of degradation in Dr. MacKenzie's sample.

2.4 Latherin

Previous work successfully produced isotopically labelled latherin to allow NMR structure determination to be initiated [73]. However, yields were not sufficient to produce large amounts of latherin economically and resulted in limiting sample concentrations. As a result, the assignment process was impeded and various proposed isotope labelling techniques, which require expensive substrates for the assignment of the large proportion of leucine residues, were not viable. A new higher yielding system for the production of isotopically labelled latherin was required.

2.4.1 Synthetic Latherin Gene

A synthetic version of the latherin gene had been produced and was ready for subcloning into the vector of choice. The synthetic Latherin (*sLath*) gene was designed with a flanking NcoI and BamHI restriction sites to allow routine subcloning into the MCS of the destination vector of choice (Figure 2.23).

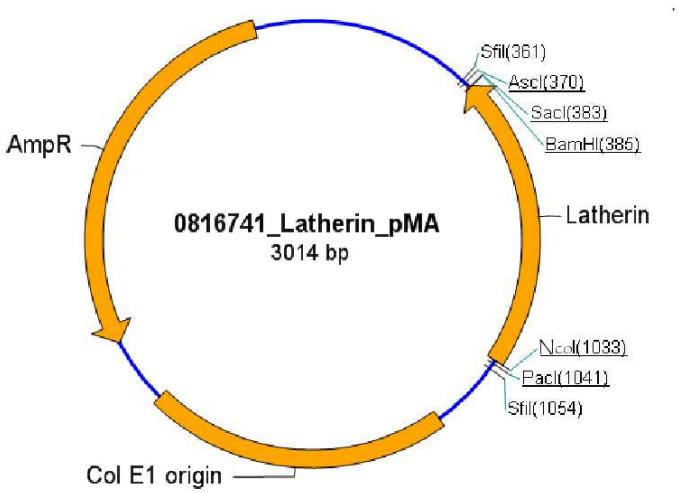


Figure 2.23- Vector map of *sLath*/pMA plasmid. Cloning vector containing *sLath* gene and requested NcoI and BamHI restriction sites.

2.4.2 *sLath*/pET-32 Expression Construct

The *sLath* gene was subcloned into Novagen's pET-32 expression plasmid (Novagen, #69015) using the NcoI and BamHI restriction sites. DNA sequencing was carried out confirming the *sLath* gene had been incorporated into the pET-32 vector as expected.

Figure 2.24- pET-32 cloning/expression region. Image reproduced from Novagen pET-32 vector map (Novagen, #TB122).

The pET-32 plasmid codes for a thioredoxin tag (Trx-Tag) to aid the formation of native disulphide bonds as well as a His-Tag and S•Tag for purification and detection, all of which are enterokinase cleavable (Figure 2.25) [143]. The tagged latherin has a MW of 40kDa and the cleaved a MW of 23kDa. However it was observed, both in this work and in the initial characterisation of latherin purified directly from sweat, that latherin migrates at an unusually fast rate in SDS-PAGE and therefore appears to have a much lower MW than expected. Beeley *et al*. proposed that the high occurrence of hydrophobic residues within the latherin sequence resulted in enhanced levels of SDS binding and hence increased mobility over that observed in an average protein [70]. In this work the tagged latherin was observed on SDS-PAGE at an apparent MW≈ 34kDa and untagged at MW≈ 15kDa.

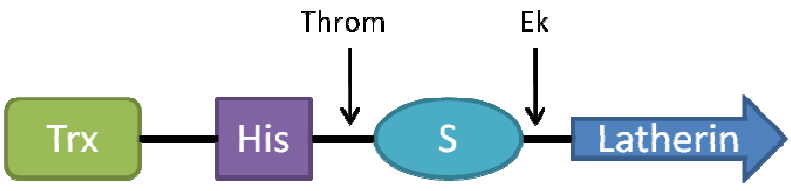


Figure 2.25- Diagram of *sLath*/pET-32 expression product. Trx= thioredoxin tag, His= His-tag and S= S•Tag. Throm= thrombin protease recognition sequence and Ek= enterokinase protease recognition sequence.

2.4.3 Latherin Expression

The *sLath*/pET-32 vector was transformed into Novagen's Tuner(DE3) cells. Expression trials indicated latherin was being expressed at low levels. A range of expression conditions were trialled to identify the optimum for soluble latherin production. It was concluded by carrying out expression at 20°C rather than 37°C, significantly higher proportions of soluble protein were achievable. It was also observed that extended durations of expression (>12Hrs) produced a greater level of total latherin (Figure 2.26).

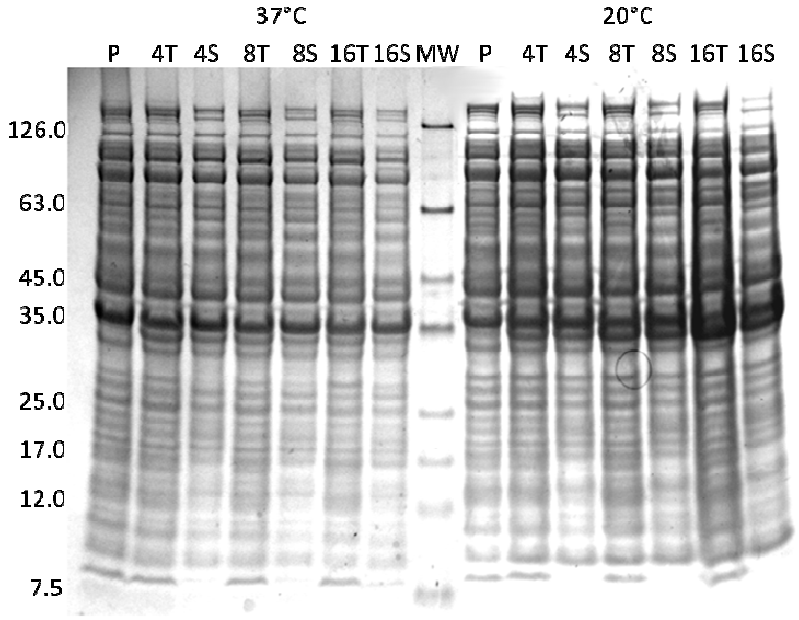


Figure 2.26- SDS-PAGE analysis of sLath/pET-32/Tuner(DE3) expression. Cultures were induced with IPTG to a final concentration of 1 mM and left to express at either 20 or 37℃ with samples taken at the desired intervals. Where Pre= preinduced; 4, 8 & 16 are the length of expression in hours and T & S refer to the total and soluble fractions respectively. MW= molecular weight markers with values in kDa (NEB, #P7708). The latherin expression band can be observed at an apparent MW of 35kDa.

2.4.4 Latherin Ni^{2+}-Affinity Chromatography

The latherin purification strategy followed the general purification procedure. However it was observed that significant amounts of protein were being lost in the post cleavage Ni^{2+}-affinity chromatography step. A large proportion of the untagged latherin, rather than passing directly through, was binding to the column and being eluted alongside the cleaved His-tag in the 300mM imidazole wash (Figure 2.27).

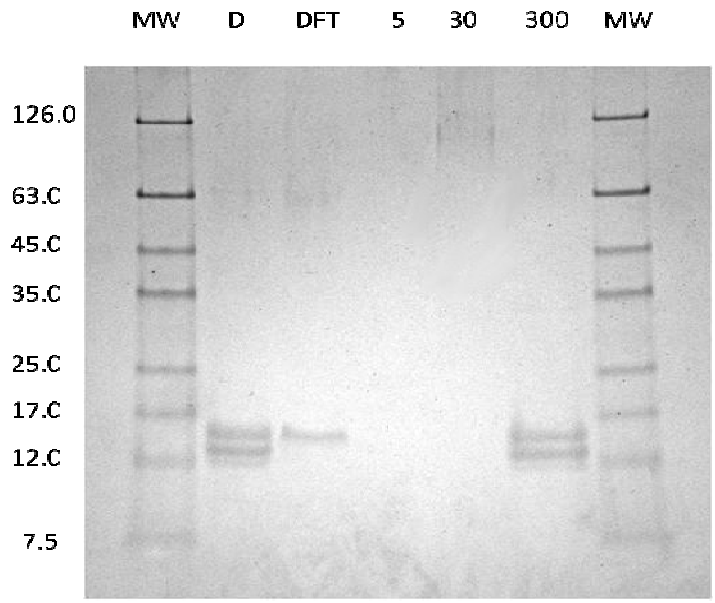


Figure 2.27- SDS-PAGE analysis of latherin Ek digest Ni^{2+}-affinity chromatography. Where D= digest products, DFT= digest flow through, 5= 5mM imidazole wash, 30= 30mM imidazole wash, 300= 300mM imidazole wash and MW= molecular weight markers with values in kDa (NEB, #P7708). A large proportion of latherin (heavier band; MW= 23kDa)is retained on the column and only eluted in the 300mM imidazole wash alongside the cleaved tag (lighter band: MW= 17kDa).

Latherin with a pI of 4.11 would have an estimated charge of -15 at pH=8.0 (the pH of the Ni-affinity chromatography buffers). An initial hypothesis was that it was likely that electrostatic interactions between the anionic latherin and the cationic Ni ions may be the cause of the untagged proteins affinity for the column. The NaCl concentration of the buffers was adjusted from 300 to 500mM in an attempt to shield these effects. However rather than improving the effect, the higher salt concentration increased the degree of untagged latherin binding to the column. To counter this effect the chromatography was repeated with buffers containing no NaCl and all latherin passed directly through the column, successfully separating the protein from the cleaved fusion tag (Figure 2.28).

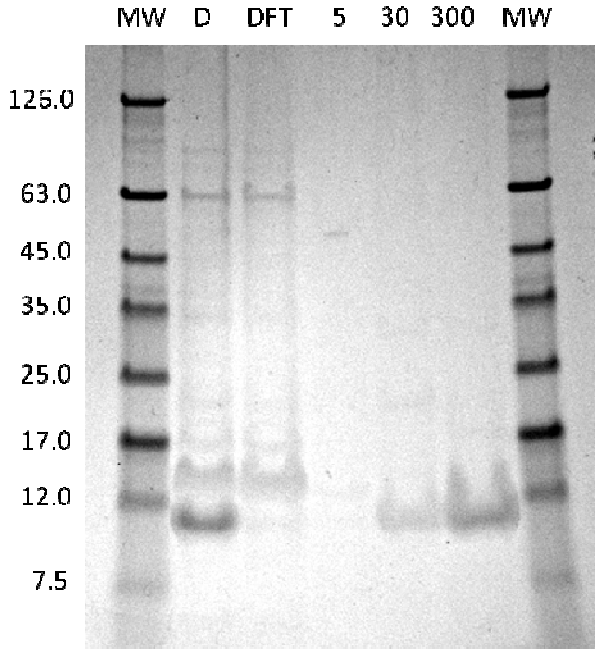


Figure 2.28- SDS-PAGE analysis of latherin Ek digest Ni-affinity chromatography using no salt buffers. Where D= digest products, DFT= digest flow through, 5= 5mM imidazole wash, 30= 30mM imidazole wash, 300= 300mM imidazole wash and MW= molecular weight markers with values in kDa (NEB, #P7708). All latherin (heavier band; MW= 23kDa) passes through in the DFT. The cleaved tag (lighter band: MW= 17kDa) is eluted in the 30mM and 300mM imidazole washes.

This ionic strength dependence is indicative of hydrophobic interactions, rather than electrostatic, causing the protein to interact with the column matrix, perhaps as a consequence of some degree of surface exposed hydrophobicity upon an amphiphilic molecule. That the protein was eluted on addition of imidazole may indicate that the interaction was with the cleaved tag rather than the column.

2.4.5 Latherin Size Exclusion Chromatography

Size exclusion chromatography of what appeared to be relatively pure latherin from the digest flow through (DFT) fraction from the post-cleavage Ni-affinity chromatography step, produced three major peaks (Figure 2.29).

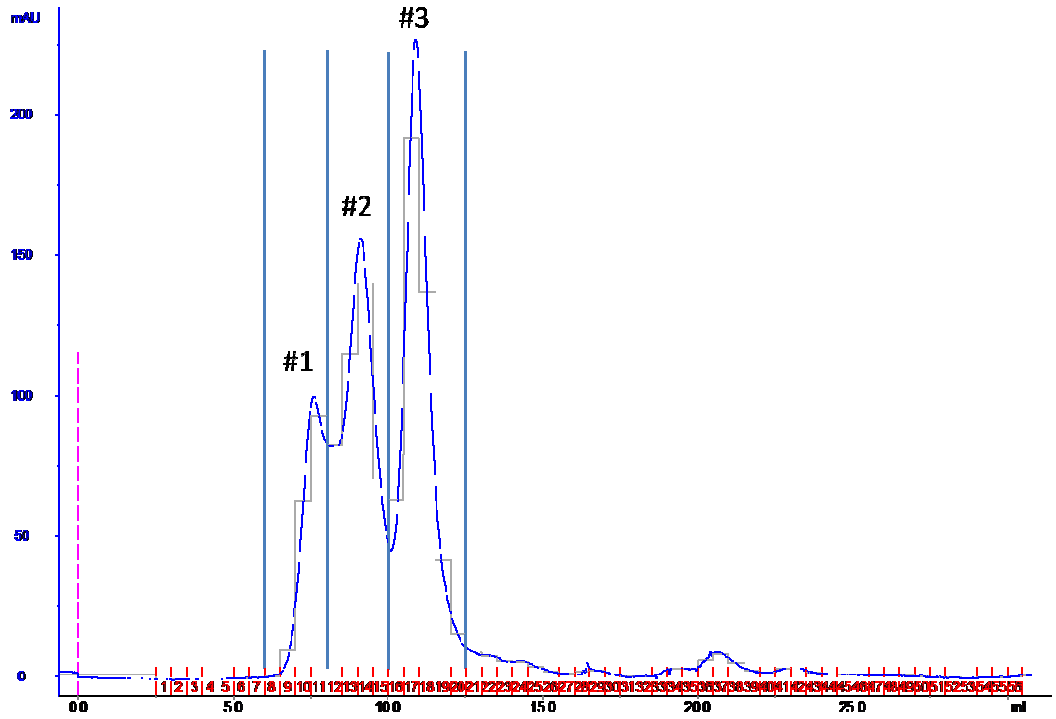


Figure 2.29- Latherin size exclusion chromatography analysis. The absorbance at a wavelength of 280nm is shown as the blue line. The fractions are shown in red and the V_e (mL) in black. The three major peaks were labelled #1, 2 & 3 and blue lines separate their main constituent fractions.

SDS-PAGE analysis of the size exclusion chromatography fractions was carried out to probe the constituents of each peak further. No bands were observed for peak #1, indicating that any protein present was either too small or too large (more likely from the V_e of the peak) to run on the gel. Both peaks #2 and #3 contained a single protein band at a molecular weight corresponding to that expected for latherin. Analysis of the size exclusion chromatography results, identified peak #2 as having a V_e much lower than expected for a protein of latherin's molecular weight. Peak #3 had a V_e far closer to that of a protein of latherin's size (these results are investigated in more detail in Chapter 3). The data indicated that peak #2 consisted of some multimer of latherin. This species was not observed in routine SDS-PAGE analysis as the protein is denatured and reduced, therefore migrating as a monomer. However, when SDS-PAGE was carried out under non-reducing conditions, the bands from peak #2 were then found at a higher MW than those of peak #3, which were not affected (Figure 2.30). These results as well as those from size exclusion chromatography indicate a species with a MW of approximately double that of monomeric latherin and is therefore almost certainly some sort of homodimer. The different results observed between reduced and non-reduced SDS-PAGE indicate that disulphide bonds play a role in the interaction between the latherin molecules that comprise the proposed dimer found in peak #2.

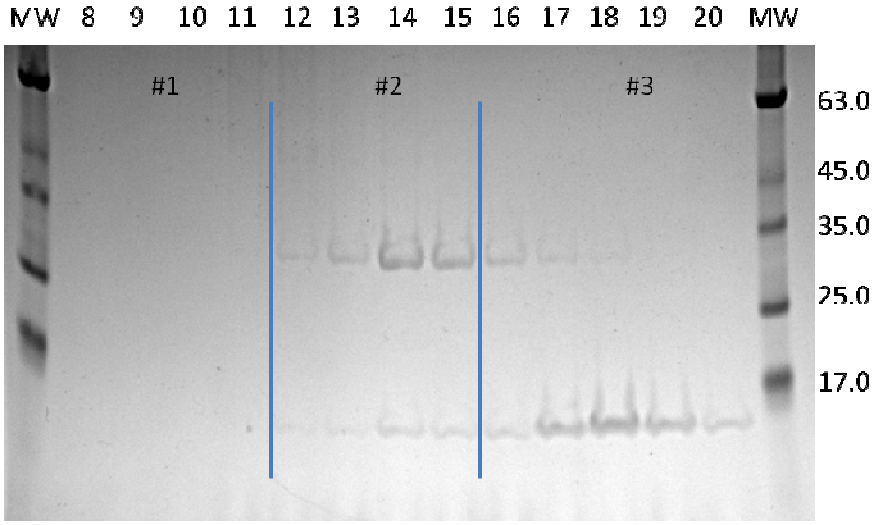


Figure 2.30- SDS-PAGE analysis of non-reduced latherin SEC fractions. The number of the fraction is shown at the top of the gel. MW= molecular weight markers with values in kDa (NEB, #P7708). The fractions have been labelled according to the peak they belong to and are separated by the blue lines. Fractions from peak #1 do not show any detectable protein content. Fractions from peak #2 show protein bands at a MW much greater than expected for untagged latherin. Fractions from peak #3 contain pure, monomeric latherin.

The characterisation of latherin purified directly from horse sweat carried out by Beely *et al*. found that latherin migrated as a single, monomeric species in sedimentation-velocity ultracentrifugation, SEC and in either reduced or non-reduced SDS-PAGE [70]. This would suggest that the redox sensitive, multimeric form of latherin found in peak #2 is an artefact of recombinant protein expression in *E.coli* as opposed to a secondary form of latherin and was therefore deemed of no relevance within this study. Therefore no further analysis of this protein was carried out and care was taken to extract SEC fractions containing this unwanted form from all subsequent latherin preparations.

2.4.6 Latherin NMR

NMR carried out on ^{15}N labelled latherin from peak #3 indicated pure, folded latherin, identical to previous NMR samples produced by MacDonald *et al*. (Figure 2.31) [73].

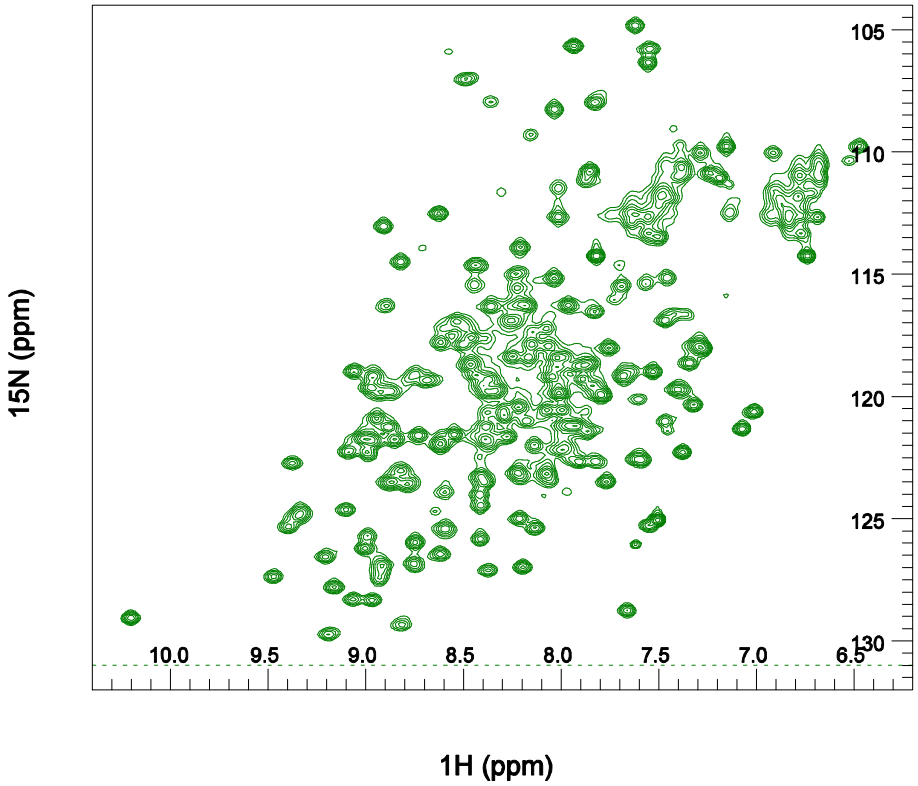


Figure 2.31- ^{15}N-HSQC spectrum of latherin from peak #3. The spectrum shows good dispersion of resonances in both the ^{1}H & ^{15}N dimensions and there is approximately one peak per residue. Both observations are indicative of folded protein.

2.4.7 Latherin Quantification

To investigate whether the yield of pure monomeric latherin from minimal media would be sufficient for full structure elucidation the yield purified from 1L of M9 was estimated by UV/Vis absorbance spectroscopy. Untagged latherin was calculated to have an extinction coefficient at 280 nm (ε_{280nm}) of 10095 $M^{-1}cm^{-1}$. The estimated yield of latherin was approximately 8.5 mgL^{-1} M9. In a 600 µL NMR sample this corresponds to a concentration of just over 600 µM, which should be sufficient for NMR structure elucidation. This yield was found to be consistent over several expressions.

3 Characterisation of Surfactant Proteins

The surfactant proteins that form the focus of this work have distinguished themselves from conventional surfactants and proteins in various ways. The differences include surfactant activity, biocompatibility and behaviour in solution. The aim of this chapter is to study and characterise these properties that make these proteins both unusual and potentially useful materials.

3.1 Surfactant Activity

Rsn-2 and latherin have been identified as highly surface active agents in previous work [55,73]. In this thesis, the activity of these proteins, as well that of protein standards: bovine serum albumin (BSA) and lysozyme, will be studied in greater detail, utilising a standard methodology to allow direct comparison of the activities of all proteins.

Rsn-1 is of unknown function within the foam nests and therefore any evidence of surfactant activity will be probed to investigate the possibility that Rsn-1 is a second surface active agent within the foam nests.

3.1.1 Method

A number of methods were employed throughout this work for the investigation of surface tension. Simple qualitative methods such as agitation of samples to determine if foam was produced were often useful. However, for more quantitative analysis three main techniques were employed: the Bikerman Assay, the Du Noüy ring, and a MicroTrough as a tensiometer. Each technique was found to offer advantages and disadvantages. However, for this work it was determined that the MicroTrough tensiometer was the most appropriate technique for the study of surfactant proteins.

3.1.1.1 Bikerman Assay

The Bikerman assay is a simple technique for quantifying the foam-ability of a solution [144]. A volume of liquid is poured into a graduated flask and gas is then bubbled through at a constant rate. Foam is produced and accumulates into a column, with older foam being pushed up the column by fresh foam. As the foam rises the liquid dispersed between the bubbles drains to the bottom until at some point the foam breaks. Steady state, when the rate of foam formation matches the rate of foam breakage, is reached when the foam column height is constant with time. This point, measured as the height of foam column from the liquid surface, is then used as an estimate of the solution's foam-ability.

Although surfactant activity is only one factor of many involved in foam formation, it was concluded that the insight obtained from this relatively straight forward technique may complement that from surface tension measurements, providing additional insight into the activity of these surfactant proteins.

It has been observed that a potential problem in interpreting the results of the Bikerman assay is that it is both dependent upon the foam-ability and foam-stability of the solution. Iglesias *et al*. proposed an amendment to the Bikerman assay whereby after reaching steady state equilibrium the gas-flow is turned off and the time for the foam to collapse from the maximum, equilibrium height, to half that height is recorded as an estimate of only foam-stability [145]. This was attempted with the protein surfactants. However, rather than collapsing in a uniform manner, the foams adhered to the side of the column, making it extremely difficult to ascertain when the foam column had reduced by half. Therefore, the conventional Bikerman approach was carried out rather than the Iglesias amendment.

All surfactant protein solutions produced good sized foam columns, however a major issue with the Bikerman assay was reproducibility. Samples from one preparation of recombinant protein gave reasonably consistent results, however when samples from different preparations were compared large variation was found (Figure 3.1).

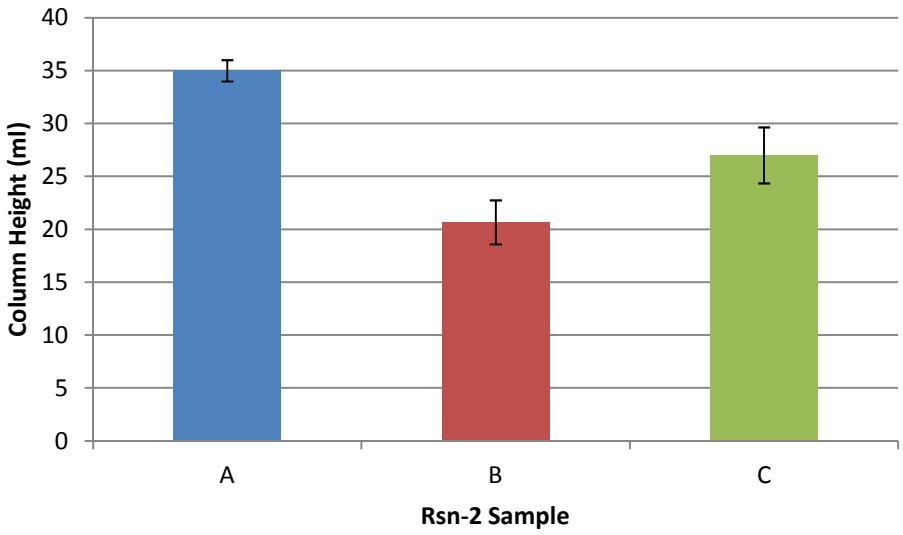


Figure 3.1- Graph showing variation in results from Bikerman assay. Measurements for three Rsn-2 preparations each at a concentration of 100 µg/mL, were carried out in triplicate to produce a mean estimate of column height and standard deviation (shown by the error bar). Details of the methodology are described in Chapter 10.

Measurements on the same samples using techniques that investigated purely surface activity produced no evidence of this variability suggesting that some other factor involved in foam-ability or foam-stability was the source of variation. Although each preparation was carried out under as identical conditions as possible, a potential source of variation may be some co-purified, foam stabilising or anti-foaming agent that was removed more effectively in some preparations than others.

Due to the variability observed in results from the Bikerman assay, its use was abandoned in favour of more reproducible techniques which measured only surface tension.

3.1.1.2 Du Noüy Ring

The du Noüy ring method was devised by Pierre Letcombe du Noüy in 1925 for the measurement of surface tension of liquids [146]. A metal ring, traditionally platinum due to its exceptional wettabilty and chemical stability, is submerged under the surface of the interface to be tested. The ring is then lifted upwards, raising a meniscus of liquid along with it. The force required to raise the ring increases until the meniscus breaks, releasing the ring. The maximum force encountered by the ring is measured by a balance and used to calculate surface tension.

The du Noüy ring has been successfully used in previous work for the measurement of protein surface tension [55,62,73]. However, in this work it was observed that even after

several minutes of equilibration, results were inconsistent at some Rsn-2 concentrations (5-10 µg/mL), with an apparent trend for the surface tension of the protein solutions to decrease with time. It was also observed that although measurements taken over time displayed variation, measurements taken from different samples at identical time-points were in good agreement. The conclusion was that the activity of the proteins was developing over time, which agreed with results from work carried out by Dr. C. MacKenzie where the MicroTrough tensiometer was utilised to observe surface tension against time [130]. Although theoretically it should be possible to measure surface tension at numerous time-points with the du Noüy ring, the manual nature of the measurement, as well as the low sampling rate would have introduced large variation in addition to being highly user intensive. It was concluded that the measurement of surface tension utilising the MicroTrough tensiometer would be a more appropriate approach.

3.1.1.3 MicroTrough as a Tensiometer

A Kibron MicroTrough S (Kibron, Espoo, Finland) was used as a tensiometer to measure the surface tension of protein solutions. A platinum rod attached to a computer-monitored microbalance is lowered into contact with the surface of interest. The downward force exerted upon the rod by the surface is then monitored. In principle the technique is similar to that of the du Noüy ring, however the ability for surface tension to be monitored continuously, in a time-dependant manner was found to be useful for the surfactant proteins, where at certain concentrations large changes in surface tension were observed over time.

One concern was the reproducibility of this method. There were many steps at which error could be introduced: variations in preparation, accuracy of sample concentration, time from deposition of the sample to beginning of measurement etc. Therefore, to obtain an appreciation of the error within the methodology three preparations of Rsn-2 were created at different stages of the project and their activity measured (Figure 3.2). Even at concentrations where the activity was developing the methodology was found to be reproducible and it was therefore selected as the best technique for the assessment of surface tension of protein solutions of those available. The methodology is described in detail in Chapter 10.

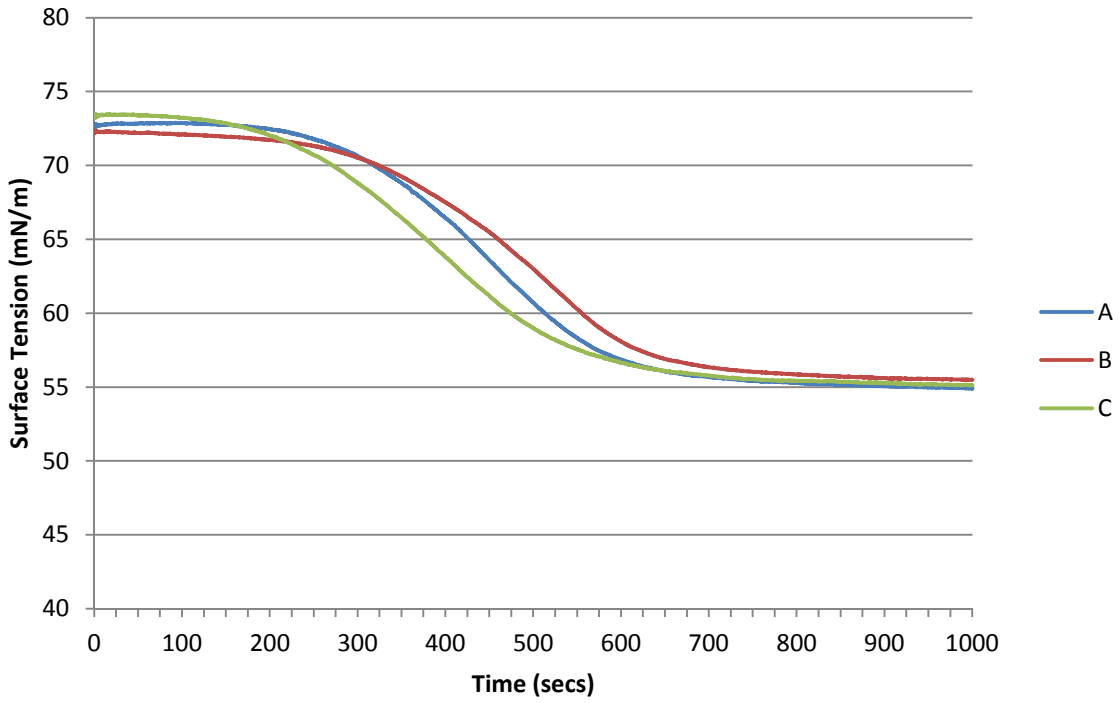


Figure 3.2- Reproducibility of MicroTrough tensiometer. Graph showing development of surface tension of three 5 µg/mL Rsn-2 samples produced in three different preparations A, B and C, to show the reproducibility of the MicroTrough tensiometer method for assessing the surface tension of protein solutions.

3.1.2 Results

3.1.2.1 Protein Standards

As discussed, every globular protein has an inherent amphiphilicity and therefore has the potential to be surface active. In order to better understand the activity of the surfactant proteins, the activity of previously studied standard proteins bovine serum albumin (BSA) and lysozyme were investigated for comparison [55,62,73].

BSA samples were produced from serial dilutions of a 10 mg/mL stock produced from commercial BSA powder (Sigma, A-7906) in 20 mM NaPi, 50 mM NaCl, 1mM sodium azide, pH 7.5.

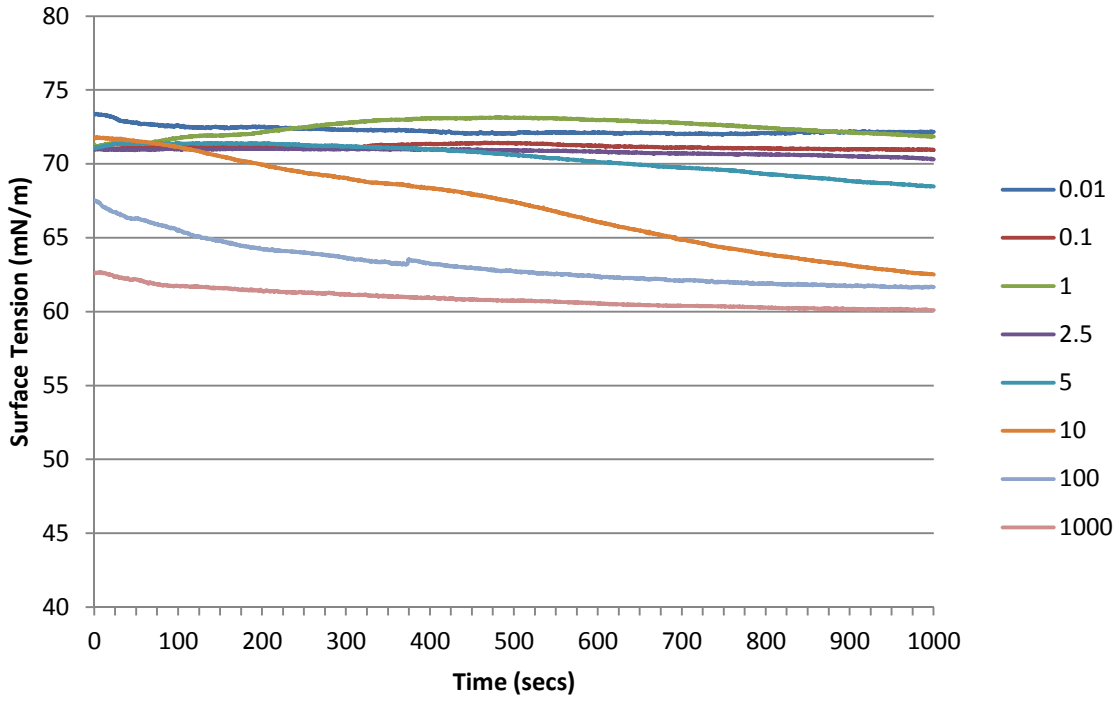


Figure 3.3- Surface tension of BSA solutions at various concentrations from the MicroTrough tensiometer technique. Protein concentrations in µg/mL are shown to the right of the graph.

BSA was found to show steady development of surfactant activity at concentrations of 10 µg/mL and above. At a concentration of 1000 µg/mL levels as low as 60 mN/m were achieved. Surfactant activity has been reported for BSA previously and was postulated to be a consequence of lipid contamination or surface proximal hydrophobic substrate binding sites [73].

Lysozyme samples were produced from serial dilutions of a 10 mg/mL stock prepared from dried Lysozyme (Sigma, L-6876) in 20 mM NaPi, 50 mM NaCl, 1mM sodium azide, pH 7.5.

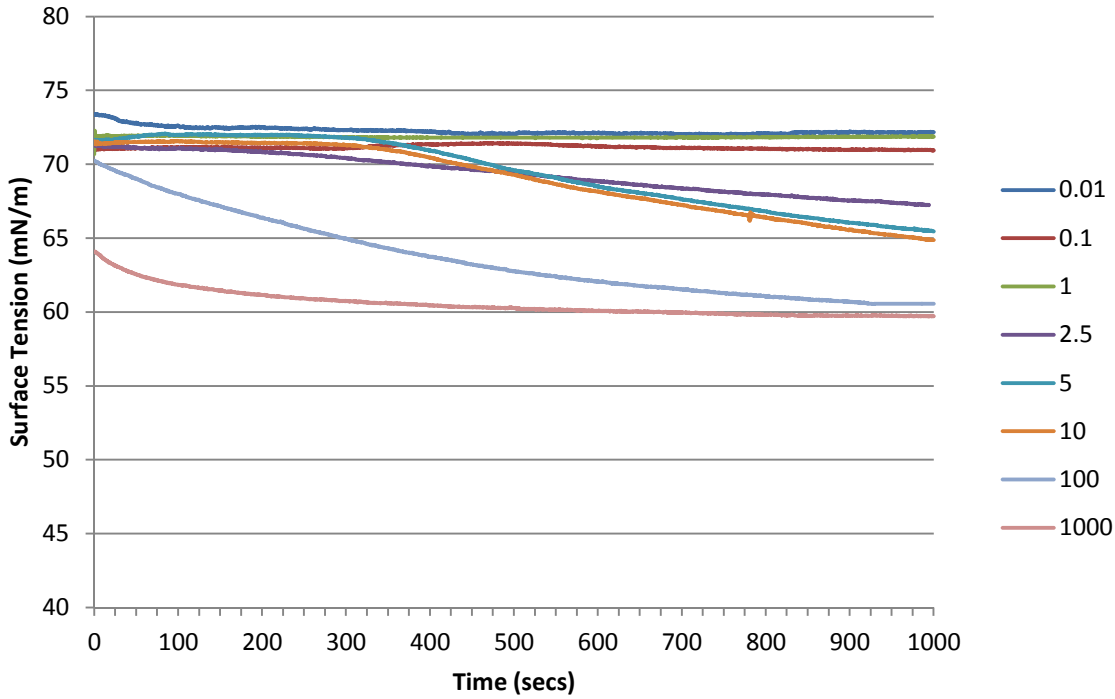


Figure 3.4- Surface tension of lysozyme solutions at various concentrations from the MicroTrough tensiometer technique. Protein concentrations in µg/mL are shown to the right of the graph.

Small reductions in surface tension are observed at lysozyme concentrations of 2.5-10 µg/mL and continue to develop over time. At higher concentrations: 100 µg/mL and above, more significant effects upon surface tension are observed, reaching concentrations as low as 60 mN/m.

3.1.2.2 Rsn-1

No surfactant activity for Rsn-1 has been reported previously, however its predicted similarity to members of the cystatin family and the cystatin-like fold seen in the surface active Rsn-2 would indicate that the structures of the two ranaspumins are similar. This in turn suggests that Rsn-1 may have a similar function to that of Rsn-2. Therefore, the surfactant activity of Rsn-1 was assessed.

In work to produce recombinant protein for NMR studies, samples containing pure Rsn-1 were observed to foam readily under agitation. Therefore to investigate Rsn-1's activity relative to other proteins it was analysed using the MicroTrough tensiometer methodology.

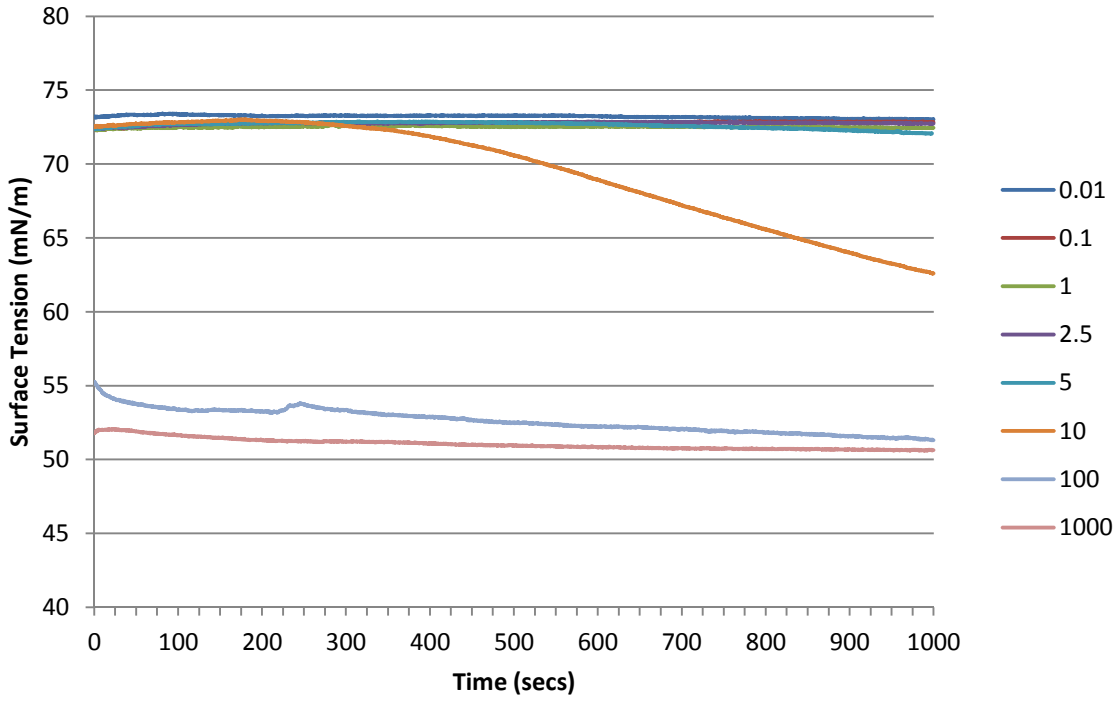


Figure 3.5- Surface tension of Rsn-1 solutions at various concentrations from the MicroTrough tensiometer technique. Protein concentrations in µg/mL are shown to the right of the graph.

No activity is observed for Rsn-1 at concentrations below 10 µg/mL. At a concentration of 10 µg/mL activity starts to develop at a steady rate after approximately 300 seconds, although not reaching equilibrium within the lifetime of the experiment a maximum reduction to 63 mN/m is achieved. At higher concentrations (100 – 1000 µg/mL) surface tensions as low as 51 mN/m are reached.

3.1.2.3 Rsn-2

The surfactant activity of Rsn-2 has been investigated by both the du Noüy ring and MicroTrough tensiometer previously [62]. This work repeated using the MicroTrough tensiometer experiment in order to investigate how the two measurements compare as well as to act as a benchmark for the mutagenesis work described elsewhere in this thesis.

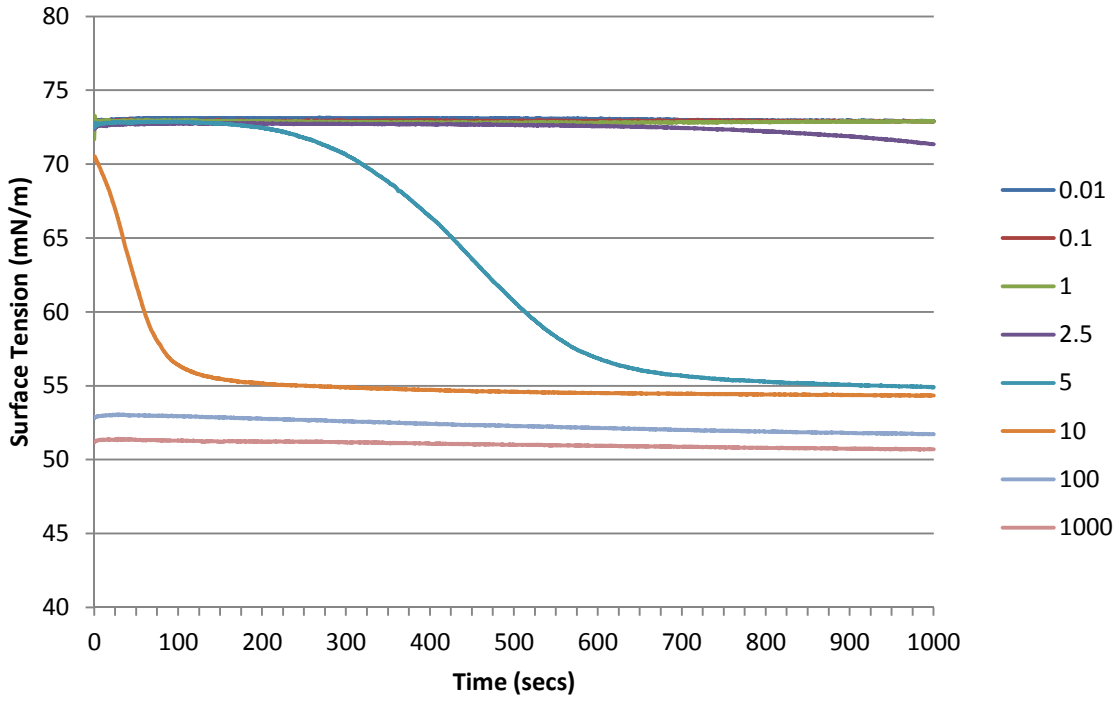


Figure 3.6- Surface tension of Rsn-2 solutions at various concentrations from the MicroTrough tensiometer technique. Protein concentrations in µg/mL are shown to the right of the graph.

The results of this work correlate well to previously reported values for Rsn-2 activity. Significant levels of surfactant activity appear at of 5 µg/mL and tensions as low as 51 mN/m are achieved at higher concentrations.

3.1.2.4 Latherin

The surface tension of latherin had been assessed in previous work using the du Noüy ring method [73]. Here the MicroTrough tensiometer method was employed.

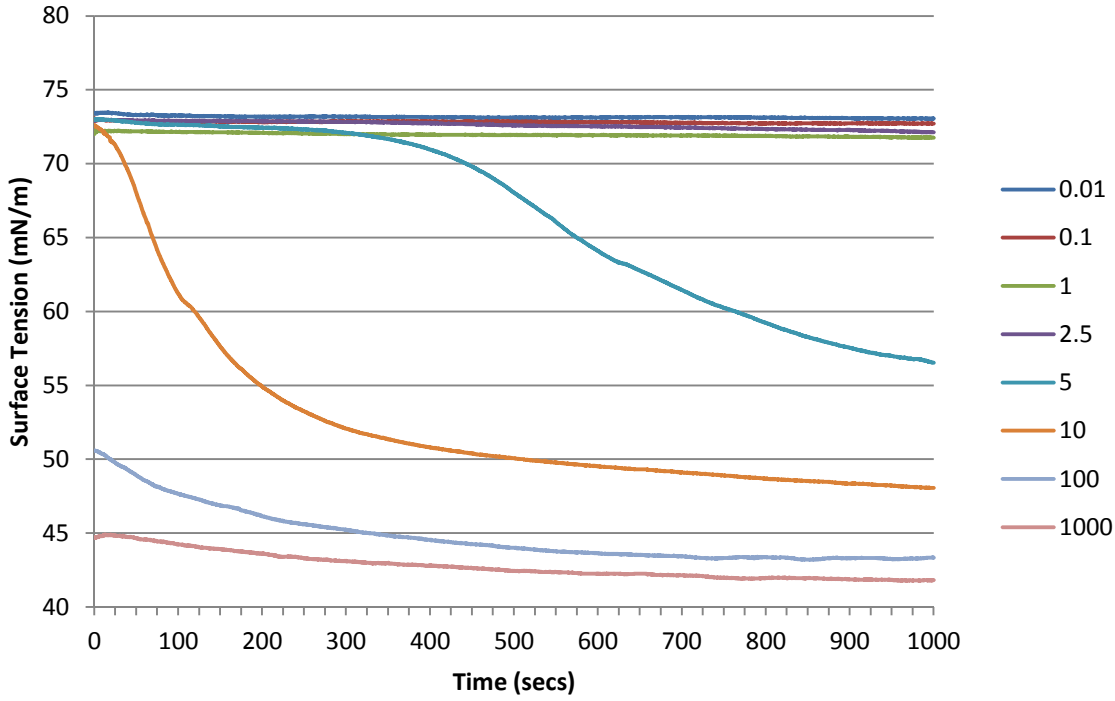


Figure 3.7- Surface tension of latherin solutions at various concentrations from the MicroTrough tensiometer technique. Concentrations in µg/mL are shown to the right of the graph.

At several concentrations (5-1000 µg/mL), latherin was found to be highly surface active. At concentrations of 5-10 µg/mL activity develops over time reaching values considerably lower than that of pure water (73 mN/m). At high concentrations, surface tensions as low as 43 mN/m are attained.

3.1.2.5 Summary

To aid comparison of surfactant activity of the various proteins, a graph of minimum observed surface tension at each concentration is shown in Figure 3.8.

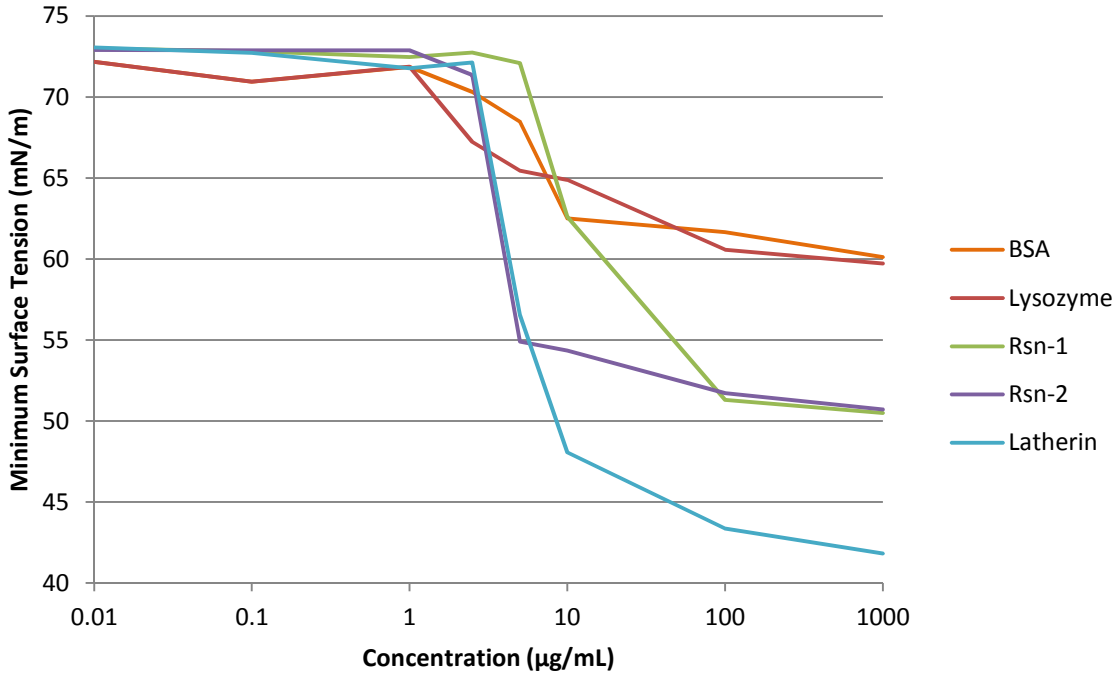


Figure 3.8- Summary of protein surfactant activity. The minimum surface tension observed for the various proteins investigated at each concentration. Concentration is displayed on a logarithmic scale.

No activity is observed for any of the proteins at concentrations of 1 µg/mL or below. In every protein, activity arises at concentrations between 1–10 µg/mL. The two standards: lysozyme and BSA, show similar activities, both reaching a minimum surface tension of 60 mN/m. The two ranaspumins although displaying large disparities in activity at concentrations of 5 and 10 µg/mL show a remarkably similar level of minimum surface tension at both 100 and 1000 µg/mL. Latherin shows by far the greatest effect on surface tension of any of the proteins tested reaching tensions as low as 42 mN/m.

3.1.3 Discussion

Both standard proteins show significant levels of surfactant activity, with BSA and lysozyme both reaching surface tensions as low as 60 mN/m at a concentration of 1000 µg/mL. These results were in good agreement to results from separate studies [55,73,147,148,149].

The activity of Rsn-2 found within this work is very similar to that reported previously by the Glasgow group [62,130]. At a concentration of 5 µg/mL the full development from zero to near full activity over the 1000 s period can be observed. It is also worth noting that at concentrations at which the protein is found in the foam nest fluid (> 100 µg/mL) almost maximum activity is observed immediately. This may be relevant in a biological context where it is important that surface tension reduction occurs sufficiently rapidly to allow for effective foam nest formation, before dilution and dispersal of natural foam fluid into the surrounding pond water.

The activity observed for latherin in this work is significantly higher than that reported in previous experiments where a maximum reduction in surface tension to 56 mN/m at a concentration of 1000 µg/mL was observed via the du Noüy ring method [73]. A possible explanation could be some discrepancy between the two techniques. However, as both instruments are calibrated before use, this would not appear sufficient to explain these large differences. Alternatively, it can be observed from the graph of latherin surfactant activity (Figure 3.7) that surface tension often takes several minutes to develop. If activity was measured before the protein was fully equilibrated, the degree of activity would be underestimated. Another potential disparity between the two experiments may have been the protein sample. In this work a multimeric latherin aggregate was detected and removed during the protein production process. It is unclear if this species was present in previous work, where latherin was produced from an alternative construct. However if it was, it would have led to a miscalculation of latherin (monomeric form) concentration, that may have affected any subsequent assay results.

There are clear differences in the activity of the proteins already identified as being surface active, Rsn-2 and latherin, and that of the standard proteins. Both standard proteins show a minimum surface tension of 60 mN/m whereas Rsn-2 and latherin reach levels of 51 mN/m, and 42 mN/m respectively. This suggests that whatever the form of the surfactant active proteins at the interface they are able to reduce surface energy by a far greater extent

than regular globular proteins, perhaps as a result of a greater fraction of hydrophobic surface area for adsorption per molecule. The two groups also show opposing behaviour in the concentrations and rate at which activity arises. In the standard proteins at concentrations between 1-100 μg/mL, a slow, near linear, increase in surfactant activity is observed over time. Whereas at concentrations as low as 5 μg/mL, after an initial delay, both surfactant proteins show a rapid transition from no, to almost full activity, giving the plots of activity over time sigmoid-like curves. This indicates that there is some feature of the surfactant proteins that allows them to achieve their surface active conformation at the interface more rapidly than standard proteins. This may be a result of enhanced adsorption of the molecule into the interface, or perhaps a lower energy barrier for any conformational exchange required for activity, relative to an equivalent process in the standard proteins. What is clear is that there are some intrinsic properties of these surface active proteins that are not present in all globular proteins that allow them to function as superior surface active agents.

At low protein concentrations the surface activity of the proteins displays sigmoid-like curves, where activity develops rapidly after an initial lag phase. That this lag is absent in the higher concentration samples and decreases with higher protein concentration suggests that this is a concentration dependant effect, specifically surface concentration, rather than a delay from conformational exchange processes.

Rsn-1 appears to share some of these properties. One of which is demonstrated by its ability to reduce surface tensions to almost identical levels to those measured for Rsn-2. This indicates that the surface active form of the two molecules has similar properties. As discussed in Chapter 1, sequence similarity suggests, that like Rsn-2, Rsn-1 forms a cystatin-like fold. It therefore seems likely that the two similar molecules may function by a common process. There are however also differences between the activity of the two proteins at lower concentrations. Unlike in Rsn-2, no activity is observed for Rsn-1 at concentration below 10 μg/mL. Comparison between the two protein samples at 10 μg/mL shows that while Rsn-2 is close to reaching full activity within 100 seconds, Rsn-1 displays no activity for an initial 400 second delay after which activity arises in a slow linear manner similar to that observed in the standard proteins. This observation may indicate that although Rsn-1 is capable of achieving similar levels to those observed for Rsn-2, it lacks the features that cause Rsn-2 activity to develop so rapidly. It has been proposed that in Rsn-2 the dynamic, unstructured, hydrophobic N-terminus and the polar C-terminus act to orientate the protein at the air-water interface, where the opposing properties of the two

termini facilitates unfolding to the active conformation [62]. This suggested role of the two termini is supported by experimental data in Chapter 4. Analysis of the corresponding regions in the Rsn-1 sequence identifies that like Rsn-2, Rsn-1 contains a polar C-terminus with a number of aspartic and glutamic acid residues. However as discussed in Chapter 1, Rsn-1 differs from Rsn-2 through the presence of a section of α-helix which may be a factor in the differing activities.

3.2 Analytical Size Exclusion Chromatography

The most studied examples of surface active proteins are the hydrophobins, which function via a surface exposed hydrophobic patch that is shielded from the solvent in solution by forming dimers. This dimerisation is not believed to occur in either Rsn-2 or latherin, however the behaviour of these proteins, as well as Rsn-1, in solution was investigated to gain further insight into how these proteins function.

3.2.1 Theory

Size exclusion chromatography, one of the techniques utilised to purify the recombinant protein in this work. The principle of the technique is to separate molecules according to apparent size by passing them through a column containing pores of various sizes. Large molecules that do not enter the pores are eluted first in the *void volume,* V_0, as they pass straight through the column without entering the pores. Small molecules which have full access are eluted last in approximately one *total column volume,* V_t. Molecules of intermediate size, with partial access to the matrix are separated in order of decreasing size. The point at which each of these molecules is eluted is described as the *elution volume,* V_e. This technique is not only a useful purification step but also allows estimation of molecular size and shape. These experimentally determined values V_e, V_0 and V_t can be used to calculate the fraction of the matrix available to each elutant, K_{av}:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \tag{3.1}$$

K_{av} is in turn proportional to the Stokes radius (R_s) of the molecule [150]. A calibration curve of the $\sqrt{-log.k_{av}}$ against R_s for protein standards of a known size can be used to estimate R_s for the protein of interest. In addition a theoretical Stokes radius: R_s^{theo}, can be calculated based upon the molecular weight, M_r and the partial specific volume, $\bar{v}$, of the molecule:

$$R_s^{theo} = \sqrt[3]{\frac{3\bar{v}M_r}{4\pi N_A}} + R_w \tag{3.2}$$

Where R_w accounts for a hydration layer surrounding the protein [151]. By comparing the ratio of the theoretical and observed R_s values, a frictional ratio, f/f_0, for the molecule of

interest can be produced. The closer this frictional ratio to 1, the more spherical the protein is. Using Perrin's equations [152] this frictional ratio can also be related to the axial ratio to determine an estimation of macromolecular shape [153].

Protein partial specific volumes were calculated from amino acid sequence using the method described by Zamatnin [154]. In each protein a full hydration layer of a thickness of 3.2 Å was assumed [151]. All size exclusion chromatography was carried out on a pre-packed Superdex 75 10/300 GL column using 20 mM NaPi, 50 mM NaCl, 1 mM NaN_3, pH 7.5, buffer. The methodology and creation of a calibration curve is described in detail in Chapter 10.

3.2.2 Results.

3.2.2.1 Rsn-1

Rsn-1 eluted at a V_e= 12.02 mL producing a K_{av} of 0.204. From the calibration curve this is indicative of an R_s = 2.23 nm. The R_s^{theo} of various possible forms of Rsn-1 were calculated, based upon a M_r of 11582 g mol^{-1} (Rsn-1 + cloning artefacts) and a $\bar{v}$ of 0.741 mL g^{-1}. These values were then used to calculate frictional and axial ratios based upon the observed R_s (Table 3.1). Multimers of three or more units can be ruled out as the minimum theoretical R_s is sufficiently larger than the observed R_s.

Rsn-1 Form	R_s^{theo} (nm)	f/f_0	Approximate Axial Ratio
Monomer	1.82	1.23	5:1
Dimer	2.21	1.01	1:1

Table 3.1- Theoretical Stokes radii, frictional coefficients and axial ratios of various possible Rsn-1 forms.

3.2.2.2 Rsn-2

Rsn-2 eluted at a 11.17 mL which gives a K_{av} of 0.221, indicative of an R_s = 2.64 nm. The $R_{s\text{-}theo}$ of various possible forms of Rsn-2 was calculated based upon a monomeric M_r of 11533 g mol^{-1} (Rsn-2 + cloning artefacts) and a $\bar{v}$ of 0.726 mL g^{-1}. These values were then used to calculate frictional and axial ratios based upon the observed R_s (Table 3.2).

Rsn-2 Form	R_s^{theo} (nm)	f/f_0	Approximate Axial Ratio
Monomer	1.81	1.46	8:1
Dimer	2.20	1.20	4:1
Trimer	2.47	1.06	2:1

Table 3.2- Theoretical Stokes radii, frictional coefficients and axial ratios of various possible Rsn-2 forms.

3.2.2.3 Latherin

In this work latherin was measured to have an R_s = 2.76 nm. Earlier work by Beeley *et al.*, working with latherin purified directly from sweat, reported a R_s of 2.47 nm by similar techniques [70]. Based upon a M_r of 22899 g mol^{-1} and a $\bar{v}$ of 0.746 mL g^{-1}, R_s^{theo} (nm) of various possible latherin forms were calculated. These values were then used to calculate frictional and axial ratios based upon both of the observed R_s (Table 3.3).

Latherin Form	R_s^{theo} (nm)	f/f_0		Approximate Axial Ratio	
		A	B	A	B
Monomer	2.21	1.24	1.11	5:1	3:1
Dimer	2.70	1.02	-	1:1	-

Table 3.3- Theoretical Stokes radii, frictional coefficients and axial ratios of various possible latherin forms. A = calculations based upon R_s observed in this work. B = calculations based upon R_s observed in separate work [70].

3.2.3 Discussion

All three surfactant proteins show faster than expected progression through the size exclusion column. The column, buffer and flow conditions used were identical in the creation of the standard curve and therefore there should be no inconsistencies in the methodology. The proteins used to create the standard curve are all globular proteins of defined R_s. The results would therefore suggest the proteins either have an elongated shape, are associated as multimers or their entry into the pores is inhibited in some manner relative to the standard proteins.

This accelerated progression of Rsn-2 in size exclusion chromatography was also observed by Dr Cameron MacKenzie in previous work, who came to similar conclusions as those stated above [130]. The solution structure, in which there is no evidence of oligomerisation, makes the prospect of multimers unlikely. NMR studies also gave a rotational correlation time (τ_m) of 5 ns for Rsn-2, entirely typical for a protein of 11 kDa, further reinforcing the conclusion that Rsn-2 is monomeric in solution. The conditions in this size exclusion chromatography are almost identical to those used in the NMR structure determination. The only difference was a slight increase in pH (7.0 → 7.5) in the size

exclusion chromatography relative to the NMR experiments. However it does not seem likely that this pH change affected Rsn-2's progression as the chromatography carried out previously by Dr MacKenzie was carried out at pH 7.0. Although the protein has a globular fold and is almost spherical in shape, the N-terminus (residues 1–17) is unstructured and dynamic and is therefore likely to result in a larger than expected Stokes radius, causing the molecule to progress through size exclusion chromatography at an advanced rate.

As very little experimental data is available with regards to the structure or behaviour of Rsn-1 it is difficult to rationalise its behaviour in solution. However its sequence similarity to the cystatin family would suggest a similar fold to that of Rsn-2. That in Rsn-1 the N-terminus is predicted to form a section of helix would appear to counter the proposition of a dynamic terminus as suggested to be responsible for the equivalent effect in Rsn-2. However, without a high resolution structure it is difficult to predict how the helix interacts with the rest of the protein and how it may affect the Stokes radius of the molecule.

The frictional ratio for latherin in either study is lower than that observed in the ranaspumins. Therefore the possibility of an elongated monomeric structure is more realistic in this case. However the possibility of the formation of a dimer cannot be completely ruled out at this stage. The solution structure should provide more insight into which of these propositions is the more likely. This is discussed with reference to the calculated solution structure in Chapter 8.

Another consideration is that there is some property of these surfactant proteins that gives rise to contrasting behaviour in solution relative to the standard proteins and impedes their entry into the pores of the matrix. The proposed hypothesis of Rsn-2's activity requires the molecule to undergo large-scale conformational exchange at the interface. If this exchange occurs within solution, the open form of the molecule would have a significantly increased Stokes radius and would advance at a rate much greater than expected based upon its molecular weight. The lack of evidence of a secondary conformation of Rsn-2 in previous NMR studies and the results of the CPMG-relaxation experiment in Chapter 4 would suggest that this hypothesis is unlikely.

3.3 Surfactant Protein Interaction with Lipid Vesicles

Frog foam nests support the development of the reproductive cells deposited within them, whilst simultaneously not succumbing to microbial degradation or colonisation. Similarly, the sweaty pelt of a horse does not succumb to microbial infection. The surface active protein Rsn-2 has been identified as unusual compared to conventional surfactant molecules because of its biocompatibility. This property is potentially exploitable within the relevant systems as illustrated by the use of Rsn-2 foams in artificial photosynthesis, discussed in Chapter 1 [63]. However, little is understood with regards to the molecular basis for this compatibility. The protein may not interact with cells at all or there may be some interaction that does not result in cell lysis. The source of the foam nests' antimicrobial activity is still not fully understood and the proposition that any Rsn-2 cell disruption is specific to bacterial cells has not thus far been investigated. No work has thus far investigated how latherin interacts with cells. This biocompatibility may be a common property of all surfactant proteins.

To gain insight into these processes, isothermal titration calorimetry was utilised to probe for evidence of interactions between the surfactant proteins and lipid vesicles.

3.3.1 Method

3.3.1.1 ITC

Any interaction between molecules can usually be detected and measured by isothermal titration calorimetry (ITC). ITC measures any changes in heat that occur as a ligand reacts with a target species. An ITC instrument consists of 2 cells. One is a reference cell, which is kept constant. The other is the sample cell. This contains a known concentration of target species in a buffer solution. The temperature between these two cells is monitored and kept constant by the feedback heaters. Any difference activates the feedback heaters to apply either increase or decrease in energy to the sample cell depending on what is required to ensure it matches the reference. This ensures that any temperature difference in the two cells is kept as minimal as possible. The energy required to keep the two cells constant is called the baseline energy. To the sample cell, an injection syringe, containing the ligand, also in buffer solution, is connected. Pre-selected volumes of the ligand are then injected into the target solution. Any binding interactions between the ligand and target will result in a temperature change in the sample cell. If the binding reaction is endothermic, the sample cell will be cooler than the reference cell and energy will be added by the feedback

heater until the sample matches the reference. If the reaction is exothermic the energy being supplied by the feedback heaters will decrease until the sample cell matches the reference. The energy increase/decrease required to correct any imbalance is measured and recorded for numerous additions of known volumes of ligand until all the binding sites are occupied. Each of these deviations from the baseline are plotted as energy vs. time resulting in a peak for each injection. The area of these peaks then indicates the energy change from the baseline at each injection. This can be calculated by integration of the graph to give the heat evolved or absorbed by the binding reaction with respect to time. This must also be run for the ligand titrated into buffer so the heat of dilution for the ligand can also be calculated and subtracted to give the true binding energies. Once this has been completed a binding curve can be created. From this, various binding parameters, such as binding constants (K_A), reaction stoichiometry (n), enthalpy (ΔH) and entropy (ΔS) can all be extracted [155].

In this work ITC served as a highly sensitive technique for the determination of the presence of interactions between surfactant proteins and lipid vesicles.

3.3.1.2 Synthetic Lipid Vesicles

Large unilamellar lipid vesicles were utilised rather than biological cells to simplify interpretation of the results. The complexities of natural cell membranes, which contain numerous components, make them particularly difficult substrates for investigation by ITC. By using lipid vesicles of defined composition, only the potential interaction between the protein and the lipid bilayer is probed, removing interference from any competing processes.

To investigate any selectivity of the proteins towards bacterial cells, two types of lipid vesicles designed to mimic both eukaryotic and bacterial cell membranes were utilised in this work. Eukaryotic membranes contain neutral and negatively charged phospholipids but the negatively charged phospholipids are located mainly in the inner layer with exposed outer layer containing neutral or zwitterionic lipid headgroups [156,157]. In contrast bacterial membranes contain more lipids with anionic headgroups, resulting in an overall negative charge [158,159]. This difference in charge on the outer surface is proposed to be the basis for selectivity of antimicrobial peptides [160].

As a model of the eukaryotic cell membrane synthetic vesicles were produced purely from dimyristoylphosphatidylcholine (DMPC) lipids. This creates lipid vesicles with an

uncharged surface due to the zwitterionic head groups on the DMPC lipids. To mimic bacterial cell membranes vesicles were produced from a 2:1 mix of DMPC and dimyristoylphosphatidylglycerol (DMPG) lipids, resulting in vesicles with a negative surface due to the anionic head groups on the DMPG lipids. The method for preparation of lipid vesicles is described in detail in Chapter 10.

3.3.2 Results

These experiments were performed in collaboration with Dr. G. Brand. A peptide fragment (residues 33-62) of human metastasis suppressor protein (UniProt #043312) known to interact with and disrupt biological membranes was used as a positive control to ensure the methodology was appropriate [161].

A negative peak indicating the evolution of heat can be observed for each titration of vesicles into the peptide solution. This evolved heat decreases as the process reaches saturation.

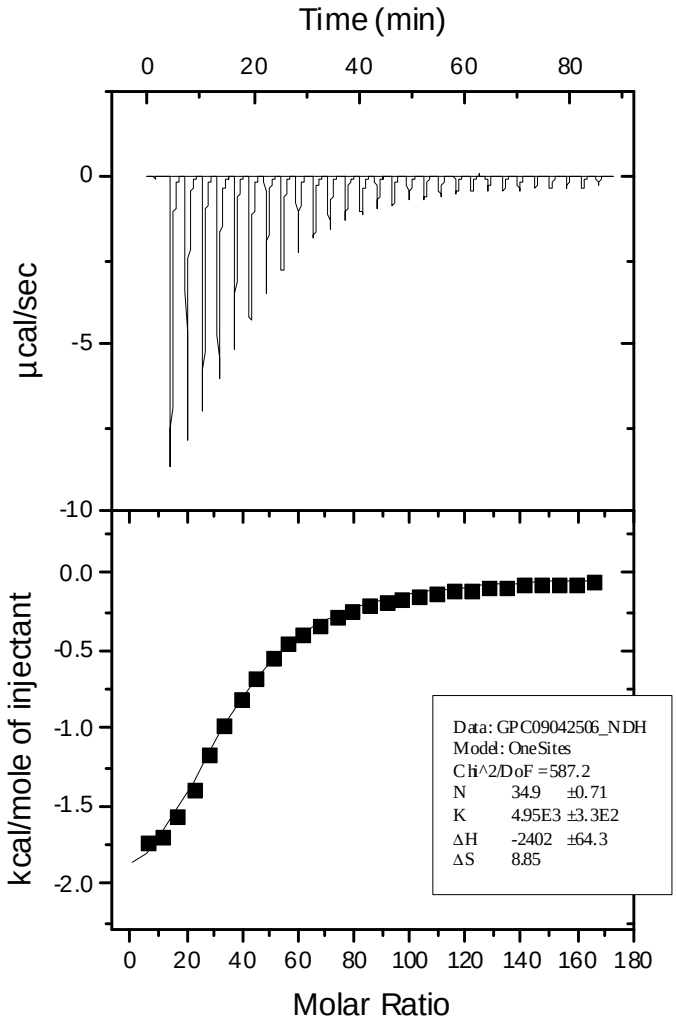


Figure 3.9- ITC results for titration of DMPC/DMPG vesicles into control peptide. A solution of 10 mM DMPC/DMPG vesicles was injected into a solution of 20 µM 0904, in 29 injections (1 x 1 µL and 28 x 10 µL).

There was no evidence of any interaction between latherin and the DMPC/DMPG lipid vesicles (Figure 3.10). Note that the heat effects here are significantly smaller than observed in the positive control (Figure 3.9), and there is no saturation apparent over this wide concentration range.

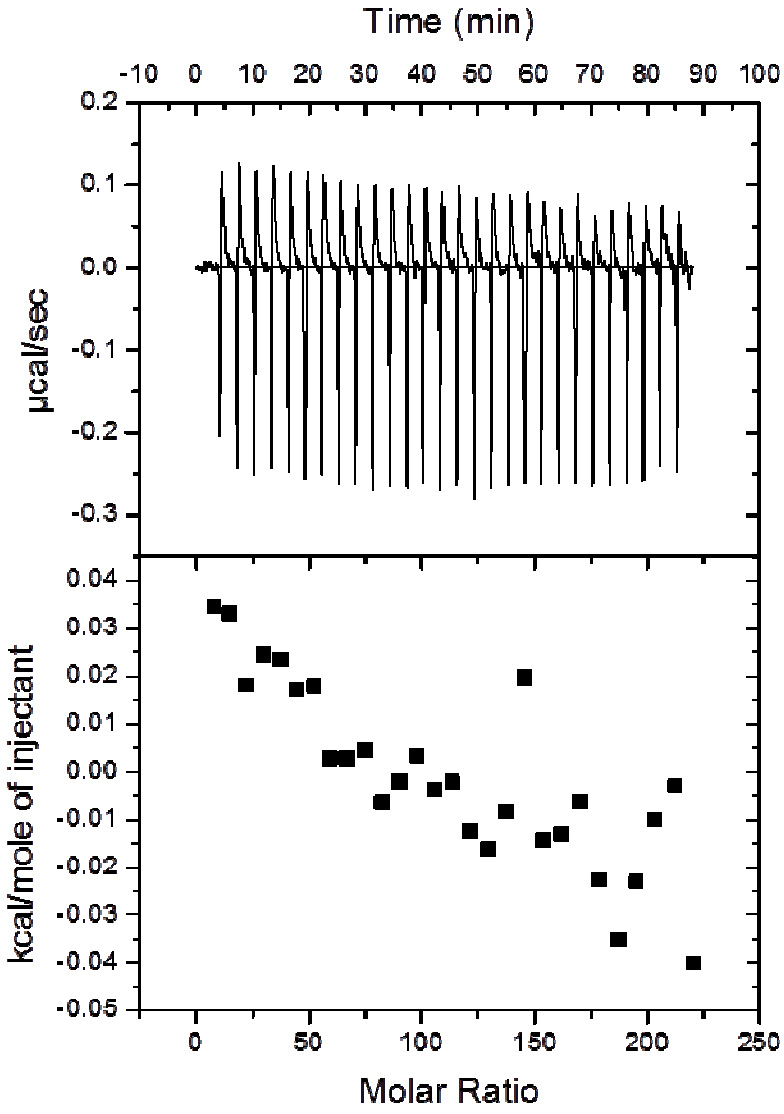


Figure 3.10- ITC results for titration of DMPC/DMPG vesicles into latherin. A solution of 10 mM DMPC/DMPG lipid vesicles was injected into a solution of 5 µM latherin, in 29 injections (1 x 1 µL and 28 x 10 µL).

Rsn-2 displayed no interaction with DMPC lipid vesicles illustrated by heat effects near zero (Figure 3.11).

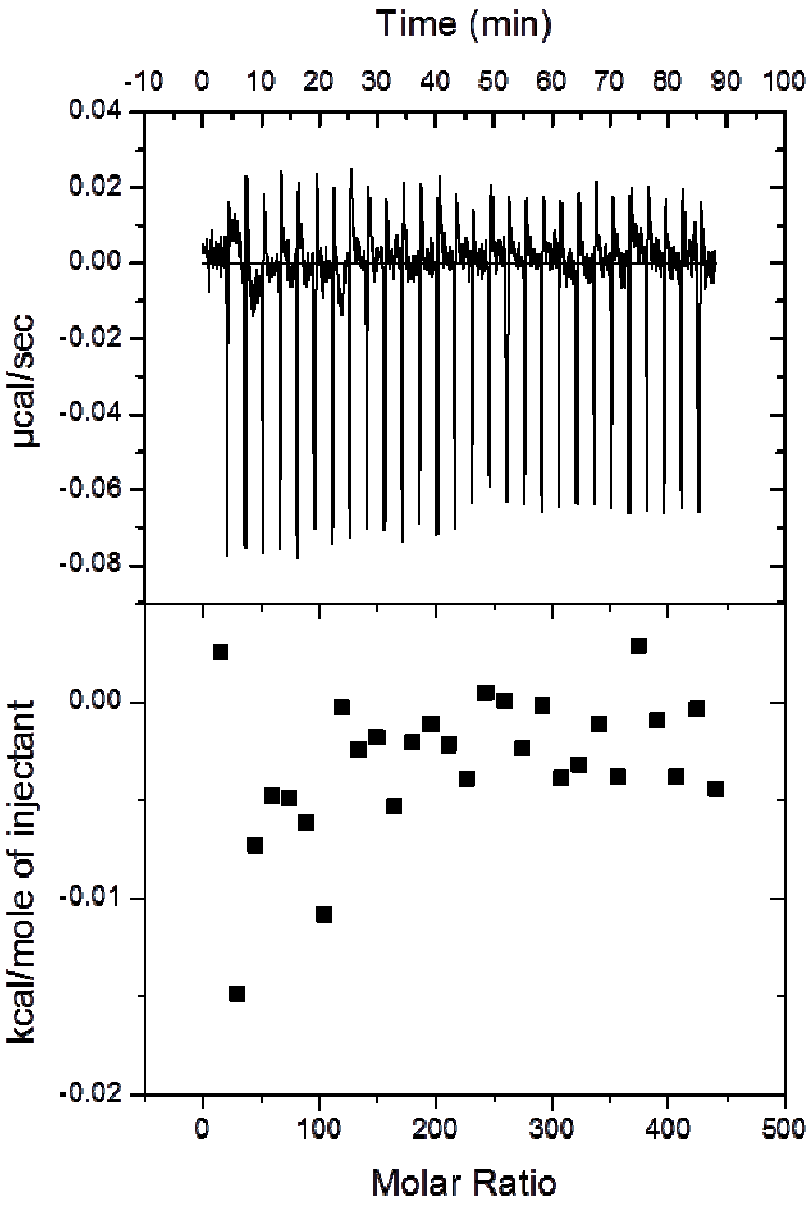


Figure 3.11- ITC results for titration of DMPC vesicles into Rsn-2. A solution of 10 mM DMPC vesicles injected into a solution of 5 µM Rsn-2, in 29 injections (1 x 1 µL and 28 x 10 µL).

Similar results are observed for Rsn-2 with DMPC/DMPG lipid vesicles (Figure 3.12).

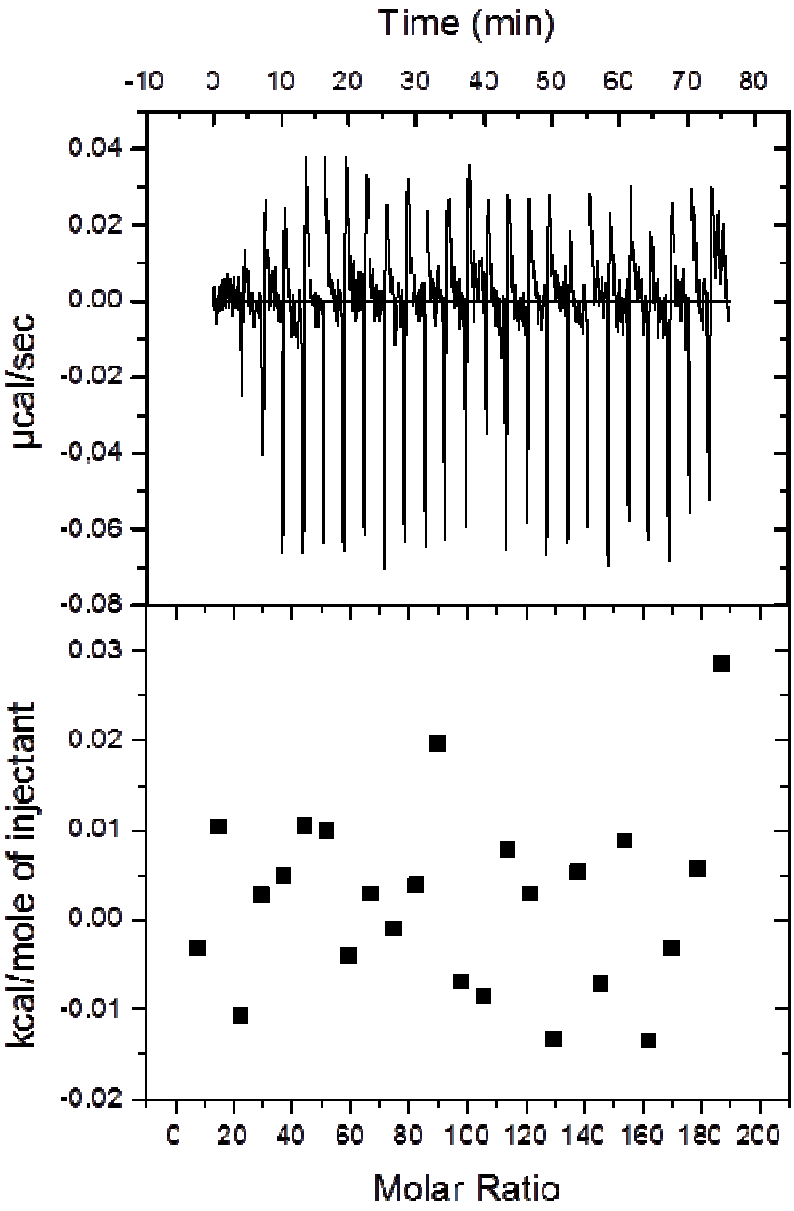


Figure 3.12- ITC results for titration of DMPC/DMPG vesicles into Rsn-2. A solution of 10 mM DMPC/DMPG vesicles was injected into a solution of 5 µM Rsn-2, in 29 injections (1 x 1 µL and 28 x 10 µL).

3.3.3 Discussion

The positive control with a known membrane interacting peptide indicates that the methodology and vesicles are suitable to study protein-membrane interactions. However, no interaction of any of the surfactant proteins with either type of lipid vesicle was observed. This therefore suggests that in contrast to cell disrupting peptides, the surfactant proteins tested here do not interact with lipid membranes.

A possible explanation for this is the size of the proteins relative to small molecule surfactants and peptides. The proteins may be too bulky to pierce the hydrophilic exterior of the bilayer and gain access to its hydrophobic interior. This seems unlikely when the activity of pore-forming proteins such as the cholesteryl dependant cytolysins [162] and the complement proteins that form the membrane attack complex are considered [163]. These soluble proteins associate with a target membrane where they subsequently form a water filled pore [164].

The pore-forming proteins interact and insert into the lipid bilayer via common features which display high interfacial solubility. Most prevalent are tryptophan rich loops and amphiphilic helices consisting of cationic residues that complement the net negative charge found in biological membranes [165]. Analysis of the solution structures of the surfactant proteins confirms that neither contains any loops or motifs rich in aromatic residues. Additionally any regions of helix display a primarily anionic character on their polar side. Therefore, the lack of interaction between the surfactant proteins and the lipid membranes appears to be a consequence of a lack of features with interfacial solubility that could mediate initial interactions. It would be interesting to investigate how the mainly acidic Rsn-2 (pI 5.16) and latherin (pI 4.11) molecules interact with lipid vesicles with a positively charged outer surface.

Although the lipid vesicles used in this work represent a simple model membrane and more work is required to probe how the proteins interact with natural cell membranes, the compatibility of the proteins with the cell mimics demonstrates the potential of the surfactant proteins for use alongside eukaryotic and prokaryotic cells and more directly, synthetic lipid vesicles.

4 Relating Rsn-2 Structure to Function

The solution structure of Rsn-2 provides no evidence of surface amphiphilicity [62]. It has however led to a hypothesis that fulfils all experimental data gathered thus far. It has been proposed that Rsn-2 unfolds at the air-water interface in a "clamshell-like" motion, exposing its hydrophobic core to the air and thus producing an amphiphilic structure with a large surface area [62]. Proteins are known to be highly dynamic molecules capable of undergoing significant structural rearrangements to fulfil their functions. These conformational exchanges within proteins can be extremely difficult to study as they normally involve the transition between a low energy, "ground" state and a high energy, "excited" state with an uneven population distribution, favouring the ground state as described by the Boltzmann distribution. Therefore, in many techniques including conventional NMR experimentation, the excited state is often too sparsely populated to be observed or is obscured from view by a more densely populated ground state. Knowledge of the ground state alone is often insufficient to fully explain the molecular basis for activity, therefore a more complete view of the relationship between structure and function of a protein requires data from more of the conformations that the protein populates.

The aim of the work described here is to probe for the existence of an alternative state of Rsn-2 that fulfils the criteria of an amphiphile and therefore provides an explanation of its surfactant activity. This goal is complicated by the possibility that this alternative conformation may only exist at the air-water interface.

4.1 CPMG-relaxation dispersion

Although conventional NMR experimentation is not optimal for studying low populations of transient states, the development of new techniques, the ability to sample wide range of timescales (picoseconds – seconds) and the availability of data from almost every atom within a molecule have allowed specialised NMR techniques to emerge as powerful methods for the study of dynamic processes [166]. One such technique is Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion, where the presence of dynamic processes occurring on a µs to ms scale is investigated by observing the effects that the exchange between states inflicts upon the spectra of the ground state [167,168,169].

4.1.1 Theory

The chemical shift of a nucleus observed in an NMR spectrum is a measurement of the Larmor frequency (υ), the rate of precession within a magnetic field of that particular nucleus. In turn the Larmor frequency is a consequence of both the gyromagnetic ratio of that nucleus (γ) and the effective magnetic field that the nucleus experiences (B_{eff}):

$$\upsilon(Hz) = \gamma B_{eff}/2\pi \tag{4.1}$$

The B_{eff} that any nucleus experiences depends upon the external magnetic field (B_0) and the environment of that particular nucleus within the external field. It is small differences in the environment of each nucleus that gives rise to the unique chemical shifts that make NMR such a valuable technique for the study of molecular structure. Consider a protein with two conformational states, A & B. Where k_B is the rate constant of the exchange from A to B and k_A is the rate constant of the reverse reaction. Therefore the overall chemical exchange can be described as:

$$k_{ex} = k_A + k_B \tag{4.2}$$

The equilibrium populations in each state, P_A and P_B can be defined by:

$$P_A = [A]/([A] + [B]) \tag{4.3}$$

$$P_B = [B]/([A] + [B]) \tag{4.4}$$

$$P_A + P_B = 1 \tag{4.5}$$

Nuclei in the ground state A and excited state B will have different characteristic Larmor frequencies depending on their individual environment. In the vector model this can be considered as two vectors rotating at slightly different offsets (Ω), the apparent Larmor frequency when taking into account the rotating frame (Figure 4.1).

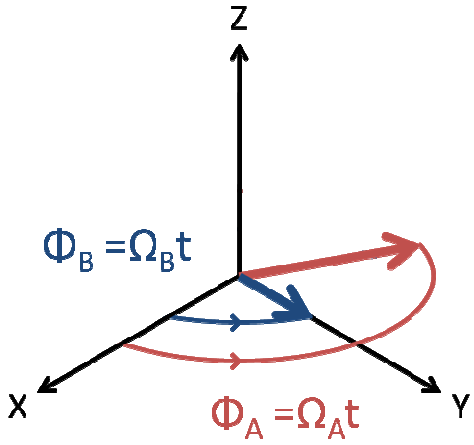


Figure 4.1- Vector model representation of two states of a perturbed nucleus precessing at different offset frequencies. Vector A, shown in red, has a higher offset than vector B, shown in blue. After time t, vector A will have rotated in the xy plane by a greater angle giving rise to two different phases (ϕ).

Without any exchange between states, these would be observed as two sharp peaks in the spectra at their respective offsets, Ω_A and Ω_B, with intensities corresponding to their populations, P_A & P_B (Figure 4.2). If $P_A \approx P_B$ and providing the two nuclei have similar relaxation properties, the peak at Ω_A will be approximately equal in intensity to that at Ω_B. However, if $P_A \approx 10P_B$ the peak at Ω_A will be approximately ten times the height of the one at Ω_B.

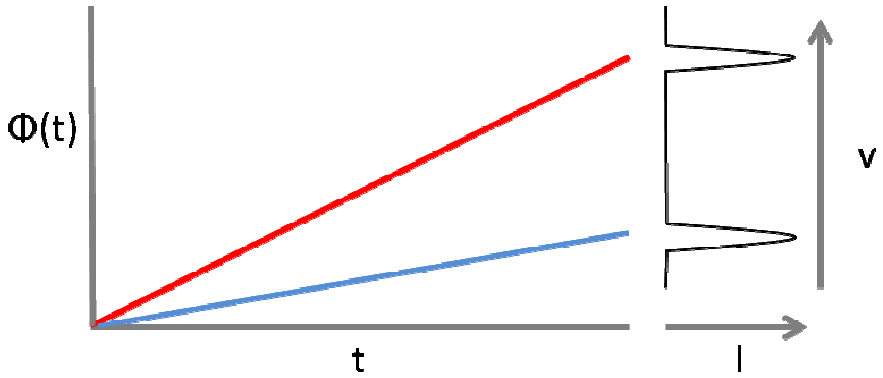


Figure 4.2- Schematic representation of the development of phase over time in two nuclei without conformational exchange. The red line represents state A and the blue line the state B. In this representation $P_A = P_B$. Each state with its individual offset, ω_A & ω_B, gives rise to an individual peak at that frequency (ν) with equal intensities (I).

Under exchange between states the system becomes more complicated. If a nucleus in state A with an offset of Ω_A exchanges to state B, it now adopts an offset of Ω_B. Hence instead of two resonances with distinct resonance frequencies, numerous amalgamations of the spins from each state, dependent upon the rate of exchange will be observed (Figure 4.3).

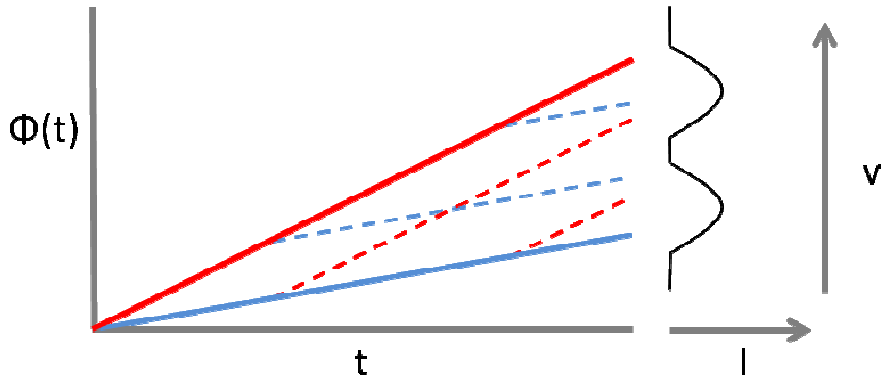


Figure 4.3- Schematic representation of the development of phase over time in two nuclei undergoing conformational exchange. The red line represents state A and the blue line the state B. Dashed lines indicate spins that have exchanged state. In this representation $P_A = P_B$. Each exchange event produces a spin with a phase that is an amalgamation of the two offsets, the extent of which depends upon the time spent in each state. This gives rise to peaks that are broadened out over a range of frequencies (ν) with equal intensities (I).

With increasing rates of exchange the peaks broaden as the two individual resonances average out over the two offsets, until under fast exchange a single resonance line is observed at the population weighted average, Ω_{av}:

$$\Omega_{av} = P_A\Omega_A + P_B\Omega_B \qquad (4.6)$$

If the populations of the two states are similar, two separate narrow peaks can identify slow exchange, as can a single average resonance indicate fast exchange. However, in the case of ground and excited states with a large energy difference, the population balance is considerably skewed to one side ($P_A >> P_B$), making any resonance at Ω_B so weak as to be undetectable. In this case a single peak would be observed that was not necessarily indicative of fast exchange as any sparsely populated excited state undergoing exchange is essentially invisible with conventional NMR experimentation [169]. Exchange can however be probed via the effects that chemical exchange inflicts upon the effective rate of transverse relaxation (the loss of phase homogeneity once the magnetisation is in the xy plane) R_2^{eff}, of the ground state. The rate R_2^{eff} is dependent upon the true rate of transverse relaxation R_2 (discussed in detail in Chapter 7), as well as the effects of static filed inhomogeneity which arises from variations in the magnetic field experienced throughout the sample. Static field broadening is an effect of nuclei experiencing varying

environments and therefore possessing varying Larmor frequencies. As a result an initially coherent vector becomes unfocused with increasing time giving rise to normally undesirable broadened, "relaxed" spectra [170]. These effects can be caused chemical exchange between conformations or experimental factors such as a non-uniform external magnetic field. The majority of effects of static field broadening can be removed, or at least reduced, by utilising a simple spin echo pulse sequence (Figure 4.4).

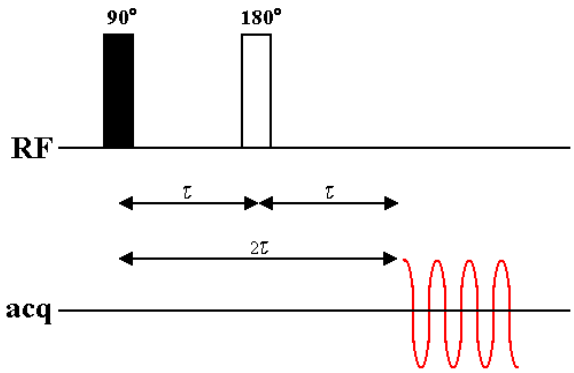


Figure 4.4- Spin echo pulse sequence. 180º pulse (indicated by open rectangle is at the centre of duration 2τ, dividing the pulse sequence into two equal segments after initial 90º RF pulse (indicated by closed rectangle). Signal is acquired at end of second delay of time τ.[171]

The spin echo consists of a 90° pulse around the x or y-axis (in the example in Figure 4.5, the x-axis) which aligns the vector along the perpendicular axis within the xy-plane (Figure 4.5, B). This is then followed by a delay of time τ. During this first delay the vector precesses round in the x-y plane at a constant speed, the offset (Ω). At the end of delay τ, the vector will have rotated to a phase (ϕ) described by Ωτ. The effects static field broadening causes variations in Ω throughout the magnetisation to develop (Figure 4.5, C). The application of the 180° pulse around the x-axis flips the vector to its mirror image or an angle of Π - Ωt from where it started (Figure 4.5, D). A second delay of time τ allows the vector to advance towards the –y-axis, with the dephasing caused by static field broadening progressively removed until coherence is regained at time 2τ (Figure 4.5, E). This produces an exact mirror image of the vector from where it started with the effects of static field broadening reversed [170].

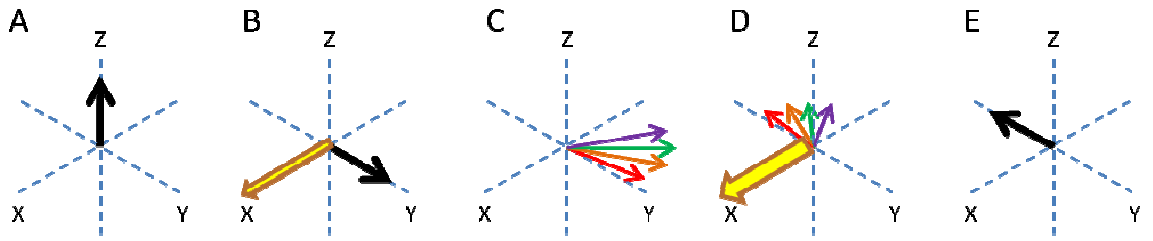


Figure 4.5- Vector model representation of a spin echo experiment. A: coherent magnetisation (represented by black arrow) aligned within the applied magnetic field B_0. B: Application of a 90° pulse (represented by thin yellow arrow) along the x-axis which flips the magnetization into the xy-plane along the y-axis. C: as the vector precesses around in the xy-plane (from x to y to –x to –y) at an average offset frequency Ω, inhomogeneity leads to fanning of the magnetisation (represented by coloured arrows (red $\Omega < \Omega_{ave}$, purple $\Omega > \Omega_{ave}$)). D: 180° refocusing pulse around the x-axis (represented by thick yellow arrow) after time τ, flips the broadened magnetisation into its mirror image. The magnetisation with slightly lower Ω, is now at the front and the higher Ω at the rear. E: another delay of time τ allows the faster precessing magnetisation to catch up with the slower, refocusing the magnetisation and hence removal of the effects of static field broadening.

A single spin echo is effective at removing the majority of the effects of inhomogeneity caused by experimental factors as they tend to be constant throughout the experiment. However the random unpredictable nature of conformational exchange means that a single spin echo cannot easily reverse its effects [172]. But by applying a series of spin echoes the effects can be reduced. This is the principle behind a Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiment. By applying a series of spin echoes at various frequencies (ν_{CPMG}) the contribution of exchange towards the effective relaxation effects can be probed. If the gap between refocusing, spin echos (τ_{CPMG}) is large, numerous exchanges can occur randomly in this time, allowing a broad range of phases to build up and hence contribute to R_2^{eff}. However, if τ_{CPMG} is short relative to the rate of exchange between states, very few or no exchanges can occur between refocusing pulses, maintaining a uniform phase and hence dramatically reducing signal loss (Figure 4.6).

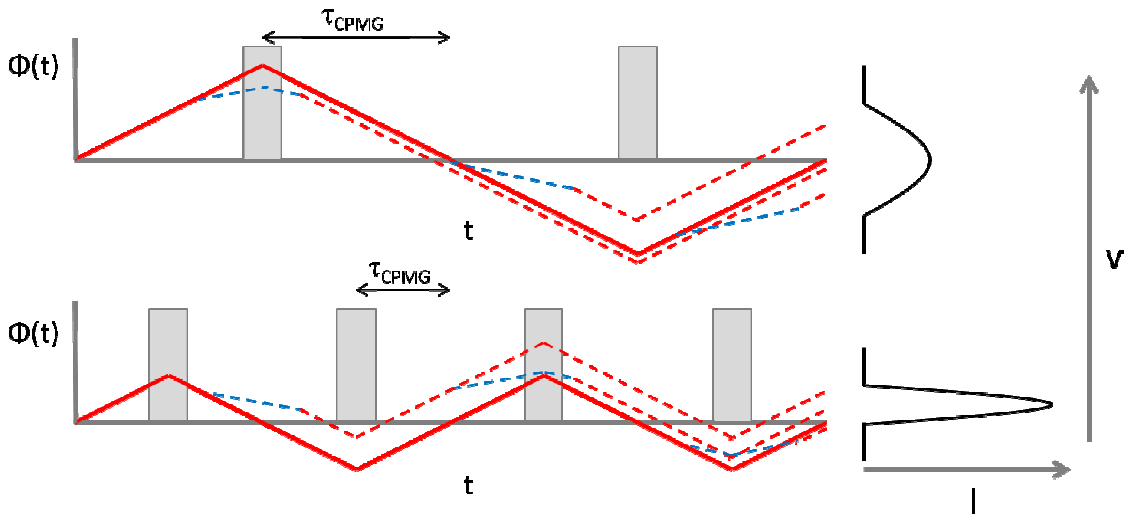


Figure 4.6- Schematic representation of effect of ν_{CPMG} upon signal relaxation caused by chemical exchange. Where: the solid red line represents the spin of the ground state, blue dashed line represents nuclei exchanged in the excited state and the red dashed line nuclei that have returned to the ground state. 180° refocusing pulses are represented by a grey box, each separated by 2 τ_{CPMG}. Figure adapted from [173].

The CPMG experiment can determine the presence of exchange, since in the absence of exchange, transverse relaxation will not be affected by the frequency of CPMG pulses. If however, R_2^{eff} is ν_{CPMG} dependent, this is indicative of chemical exchange. Furthermore, if exchange is present, complex kinetic information on the process can be derived from the experimental data. The effect of ν_{CPMG} upon transverse relaxation R_2^{eff} can be calculated for each peak via the equation:

$$R_2^{eff}(\nu_{CPMG}) = \frac{-1}{T}\ln\frac{I(\nu_{CPMG})}{I_0} \text{ [174]} \qquad (4.7)$$

where $I(\nu_{CPMG})$ & I_0 are the intensities of a given cross peak in the presence and absence (respectively) of a CPMG pulse sequence at frequency ν_{CPMG}. $T/2$ is the length of each CPMG pulse train. And ν_{CPMG} is $1/(4\tau_{CPMG})$, with $2\tau_{CPMG}$ the separation between the centres of successive refocusing pulses [174]. By fitting curves for relaxation, R_2^{eff} vs. refocusing pulse frequency, ν_{CPMG}, kinetic data such as k_A, k_B & k_{ex}, the populations P_A & P_B, the relaxation not contributed by exchange R_2^0 and the difference in Larmor frequency of the two nuclei in the different states ($\Delta\omega$), can in favourable cases all be extracted [169,175,176,177].

4.1.2 Method

A ^{15}N labelled recombinant Rsn-2 sample was produced from the *Rsn-2*/pCRT7/BL21(DE3) in ^{15}N M9 minimal media following the protocol described in Chapter 2. Pure ^{15}N Rsn-2 was concentrated to 190 µM (2.1 mg/mL) in the same buffer as was used in Rsn-2 structure determination by NMR (20mM NaPi, 50mM NaCl, 1 mM sodium azide, pH 7.0, 5% D_2O) [130].

Experiments were carried out on a Bruker Avance 600 MHz spectrometer equipped with triple resonance cryoprobe and gradient channel. The pulse scheme described by Mulder *et al*. [178] was used at 293, 298, 303, 308 (the temperature used for structure determination [62]) & 312K with 23 ν_{CPMG} values, ranging from 50 – 1000Hz (or 1 – 20 ms delays) at increments of 50Hz. Measurements at 500Hz were repeated three times and 550Hz twice to estimate the inherent error in the calculation of cross-peak intensity. Reference spectra were also recorded by omitting the CPMG pulses.

Spectra were processed using the *Azara* suite of programmes (Dr. Wayne Boucher, www.ccpn.ac.uk/azara). Initial processing scripts were created from the Bruker data files via the programme *ReferenceB* written by Dr. Krystyna Bromek. Assignment and peak intensity (by height) calculations were carried out in the programme CCPNMR *Analysis v.2* [179]. Backbone assignments were copied from an assigned ^{15}N-HSQC used by Dr. MacKenzie in the structure determination process [130].

4.1.3 Results

55 of the Rsn-2 backbone residues were identified as having sufficiently unique chemical shifts (not overlapped) to provide reliable data. As an initial analysis, the rate of change of the peak intensity (ΔI) with refocusing pulse frequency (ν_{CPMG}) for each backbone amide was calculated by fitting the function $I = \frac{\Delta I}{\Delta \nu_{CPMG}} \nu_{CPMG} + c$ within the CCPNMR *Analysis* programme for experiments carried out at all five temperatures. This was normalised by dividing by the peak intensity in the absence of the CPMG pulse (I_0) and plotted by residue number (Figure 4.7). An error was also calculated from the normalised error of the fit.

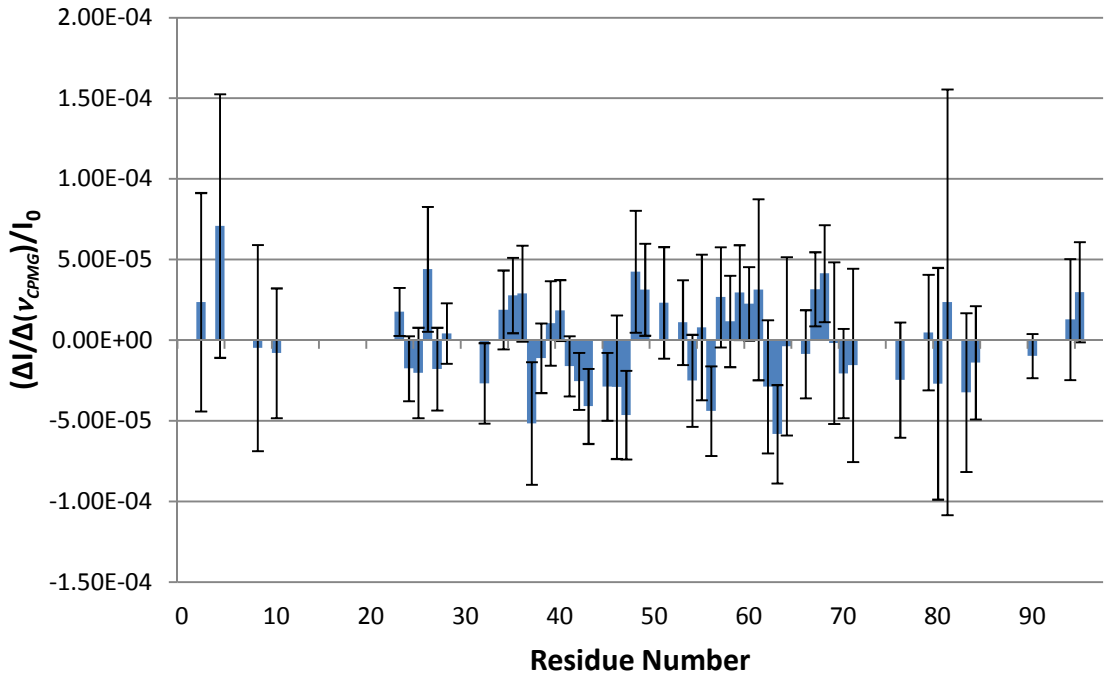


Figure 4.7- The normalised dependence of peak intensity upon ν_{CPMG} at 308K. The linear rate of change of peak intensity against ν_{CPMG} normalised against peak intensity in the absence of the CPMG pulse is shown as bars. The error of the fit is shown as error bars. Almost all dependence upon ν_{CPMG} is within the error of the fit.

The graph indicated that for almost all of the peaks sampled any change in peak intensity was within the error of the fit. Only a small number of peaks showed any significant increase in peak intensity and in each of these cases the data was examined and no significant trend indicating dependence on ν_{CPMG} was observed. Nor was there any concentration of these peaks to specific regions of the protein that would potentially indicate an area undergoing conformational exchange.

4.1.4 Discussion

There is no evidence of any dependence of peak intensity upon ν_{CPMG} outside of the calculated experimental error. Therefore no conformational exchange has been observed on the timescales and populations that this experiment is sensitive to. There are several reasons that this may be the case. Primarily the prospect that Rsn-2 does not undergo exchange between states must be considered. It would however contradict the IRRAS and neutron reflectivity data.

An alternative explanation is that conformational exchange in Rsn-2 is occurring outwith the timescales investigated in the CPMG experiment (1-20 ms). Exchange rates outside this range would not be affected by the CPMG pulses and no change in peak intensity

would be observed. Signals from both conformations from any nuclei undergoing exchange at rates slower than 20 ms would be expected to be observed in conventional NMR experimentation, providing the population/concentration of the alternate conformation was sufficient for detection. The appearance of secondary conformations for a group of residues observed in the assignment process would indicate that some exchange between states could be occurring on a timescale slower than that probed by the CPMG experiment, although these peaks may arise from a contaminant or degradation product [52].

Another explanation as to why no evidence of exchange was observed is that the proposed active form of Rsn-2 is believed to exist at the air-water interface and the “open” form may therefore not occur in the bulk. The geometry of an NMR tube provides a very small area of air-water interface relative to the bulk volume of the sample. Also the probe in a spectrometer is designed to sample only the bulk solution to ensure a homogeneous magnetic field is experienced by all observed molecules. Therefore if this alternate conformation exists exclusively at the interface, exchange between the two states is likely to only occur at the interface or in the immediate sample area. Therefore even if exchange is occurring within the sample, it may be occurring outwith the sample volume probed by the spectrometer.

In conclusion although the CPMG relaxation dispersion experiments failed to detect the presence of conformational exchange, this does not necessarily indicate that this exchange does not exist. It seems probable that solution state NMR spectroscopy is not a suitable technique to investigate an alternate conformation that may only exist at the air-water interface.

Attempts were made to produce stable emulsions of hydrophobic solvents including cyclohexane and chlororform in H_2O stabilised by Rsn-2 with the aim of studying the form of the protein at a hydrophobic interface by solution state NMR. However the emulsions were not sufficiently stable for use within the experiment timescales. Even if this obstacle could be overcome, various other challenges such as obtaining a sufficient concentration at the interface and the correlation time at the interface may prevent the measurement of quality spectra.

4.2 Rsn-2 Mutagenesis

A common technique for investigating protein activity is mutagenesis. Amino acids can be mutated to an alternative residue or removed altogether in either a rational or random manner to identify motifs/residues that are critical for protein function. Specific regions of the Rsn-2 structure have been proposed to be key to the observed surfactant activity [62]. The work described here aims to probe specific structural features of Rsn-2 and their potential roles, in order to increase understanding of the molecular basis of surfactant activity.

4.2.1 Disulphide Bond Introduction/Removal Mutations

The proposed clamshell opening of Rsn-2 requires the ability of the alpha helix to re-orientate independent of the β-sheet. If this hypothesised change in conformation is indeed the source of activity, any inhibition of the dynamics of such a process should have an effect upon activity.

Throughout nature disulphide bonds are incorporated to stabilise three dimensional structures of extracellular proteins by covalently binding sections of the polypeptide chain together. This principle was utilised here in an attempt to restrict the possible conformational changes in Rsn-2. Pairs of cysteines were introduced into specific locations of the Rsn-2 sequence to covalently bind the alpha-helix and beta-sheet together and thus prevent the hypothesised “clam-shell” opening.

4.2.1.1 Rationale

Due to the high dependence of structure upon amino acid sequence, care was taken to ensure that any effect upon activity was due to the disulphide-linkages and not to unwanted structural consequences of the mutagenesis. The cysteine residues were therefore placed in locations thought most likely to have minimal effect on the folded Rsn-2 structure. Natural cysteine bonds have C_β-C_β distances ranging from 3.45 to 4.40Å [180]. Therefore residues on alternate sides of Rsn-2’s hydrophobic core with inter C_β distances within this range were required.

N19 at the N-terminus of the α-helix and I46 on the first strand of the β-sheet (Figure 4.8) were identified as fulfilling all criteria for potential mutagenesis sites, in particular having

an inter C_β distance of 3.8Å. At this position the proposed disulphide would bridge the hinge region of the hypothesised clam-shell opening.

The area where the introduction of a disulphide linkage would be expected to have greatest effect would be on the opposing “lips” of the proposed conformational exchange (the C-terminus of the α-helix and the fourth strand of the β-sheet). In this area of the protein no C_βs fall within the designated range. The closest are residues D32 and M81 which have an inter C_β distance of 4.8Å (Figure 4.8). This is larger than would normally be desired however the M81 residue is on an unstructured loop, which should mean that, the additional restraint from the disulphide bond will not drastically affect the structured areas of the protein.

The introduction of a single mutant disulphide at one end of the molecule may not be sufficient to inhibit the hypothesised dynamics. In an attempt to fully restrict any conformational exchange a mutant was created to contain all four cysteine mutations. As this form of Rsn-2 should contain two mutant disulphide linkages “pinning” either end of the α-helix to the β-sheet, any dynamics involving the unfolding of these features away from one another should be prevented.

An additional concern was the potential for formation of unwanted disulphides between the introduced cysteines and the native cysteines. Therefore, a mutant with the native cysteines replaced with other residues was created that could subsequently be mutated to include the designed disulphides. This also provided an opportunity to investigate the role of the native disulphide with respect to surfactant activity. To remove the existing cysteines C68 and C86 (Figure 4.8), they must be replaced with alternate residues. This is a rather complex process as the introduced residues must have similar properties to that of the oxidised cysteines in order to have minimal effect upon the overall protein structure. Cysteines in a disulphide bond are considered hydrophobic and are therefore commonly mutated to hydrophobic residues such as alanine or methionine. However, the disulphide bond in Rsn-2 is slightly surface exposed, particularly around the location of C68, raising concern with regards to how mutating hydrophobic residues into this area may affect both structure and activity of Rsn-2. The conservative approach of producing both a “hydrophobic”, valine and “less hydrophobic”, threonine mutant was employed. Mutation to an alanine was selected as the most appropriate choice for the buried C86.

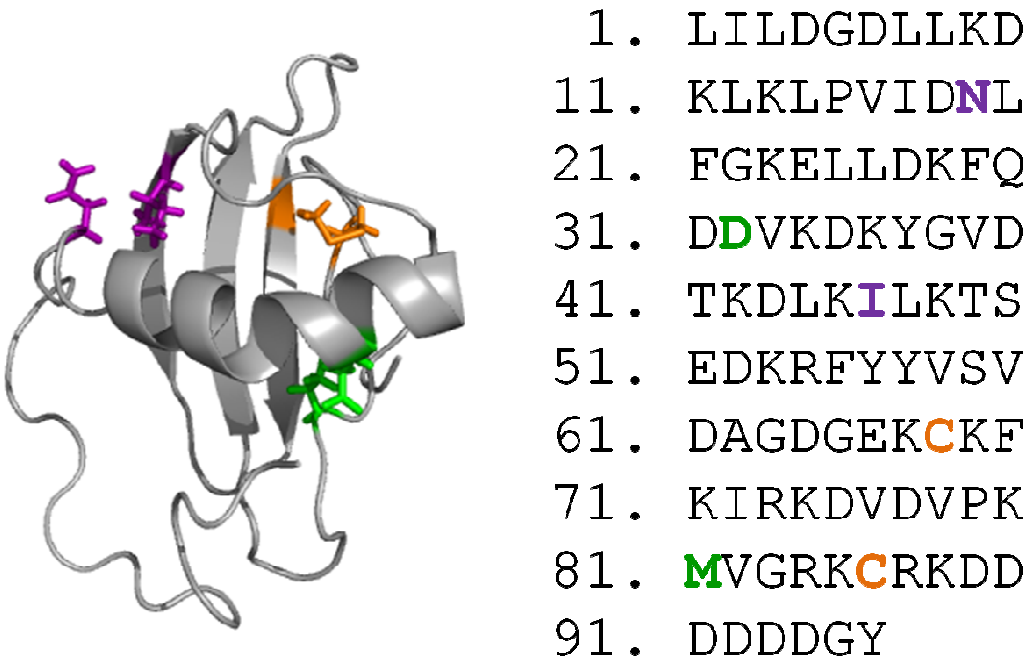


Figure 4.8- The location of the residues to be mutated to, or from, cysteines on both the structure and sequence of Rsn-2. The N19 and I46 pair of residues to be mutated to cysteines are shown in purple. The D32 and M81 pair of residues to be mutated to cysteines are shown in green. The C68 and C86 native disulphide to be mutated is shown in orange.

4.2.1.2 Method

This work was carried out with the assistance of Miss Sallie Antill. Disulphide bonds were introduced or removed by site-directed mutagenesis using a QuikChange II Site Directed Mutagenesis kit (Stratagene, #200524). The QuikChange II kit allows for point mutations to be introduced into plasmid DNA quickly and efficiently.

Primers were designed to anneal to both strands of the template DNA at the area surrounding the codon to be mutated. Within the primers the relevant codon was replaced with DNA encoding the desired mutation. PCR carried out utilising these primers produced plasmid DNA with the mutation incorporated. The template DNA was then digested using DpnI, a restriction enzyme that digests methylated DNA at the sequence GA↓TC. In *E.coli* DNA is methylated by DNA adenine methylase whereas that produced from PCR *in vitro* is not. DpnI therefore selectively digests any template DNA leaving only plasmid containing the mutation preventing false positive transformations arising from residual template DNA. The plasmid was then transformed into competent cells, ready for use.

In this work *Rsn-2*/pET-28 plasmid was used as the template. Primers Rsn2_N19C_f, Rsn2_N19C_r, Rsn2_I46C_f and Rsn2_I46C_r were designed for the mutagenesis of N19 to C19 and I46 to C46 (Figure 4.9).

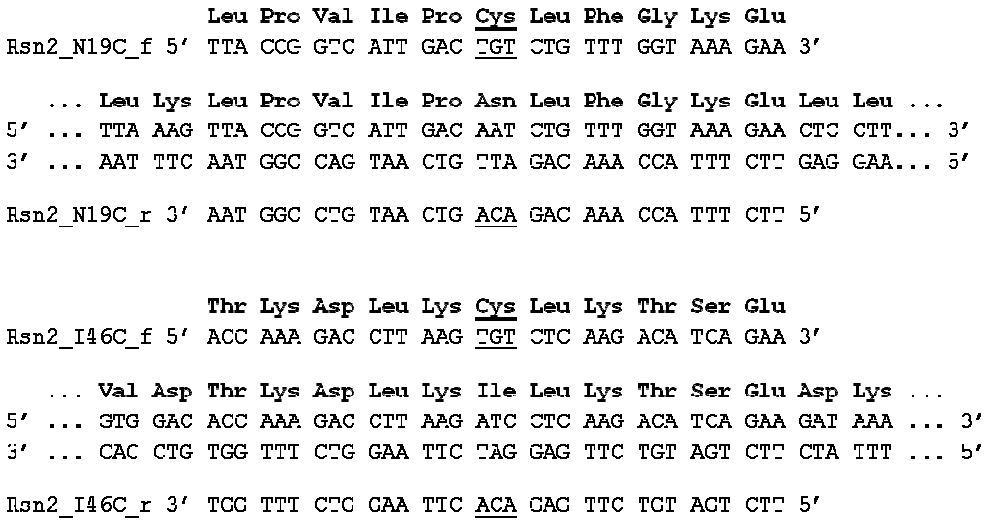


Figure 4.9- Rsn2_N19C_f, Rsn2_N19C_r, Rsn2_I46C_f & Rsn2_I46C_r primers and their annealing sites on the Rsn-2 gene. The mutant codons and resultant residues are underlined.

Primers Rsn2_D32C_f, Rsn2_D32C_r, Rsn2_M81C_f and Rsn2_M81C_r were designed for the mutagenesis of D32 to C32 and M81 to C81 (Figure 4.10):

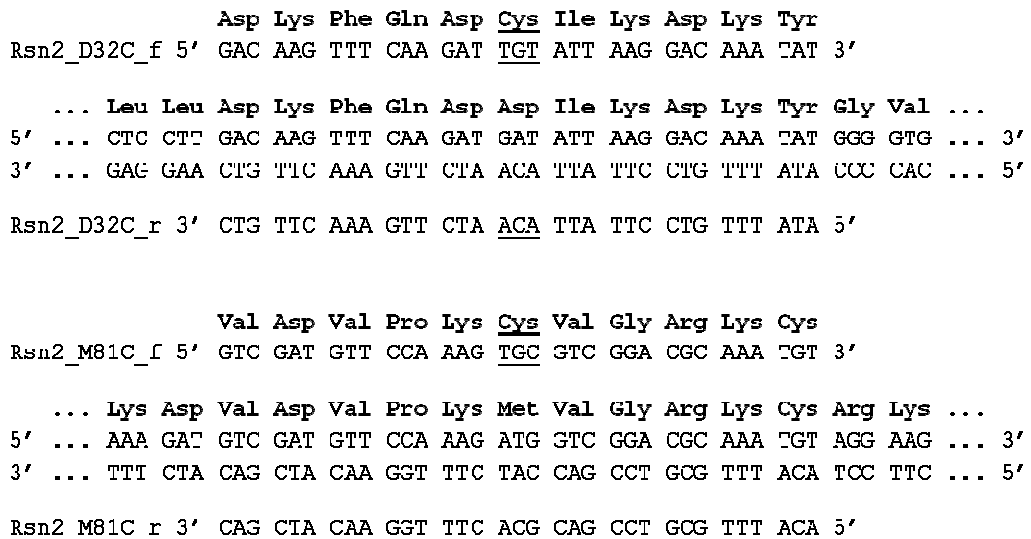


Figure 4.10- Rsn2_D32C_f, Rsn2_D32C_r, Rsn2_M81C_f & Rsn2_M81C_r primers and their annealing sites on the Rsn-2 gene. The mutant codons and resultant residues are underlined.

Primers Rsn2_C68T_f, Rsn2_C68T_r, Rsn2_C68V_f, Rsn2_C68V_r, Rsn2_C86A_f and Rsn2_C86A_r were designed for the mutagenesis of C68 to T68 & V68 and C86 to A86 (Figure 4.11).

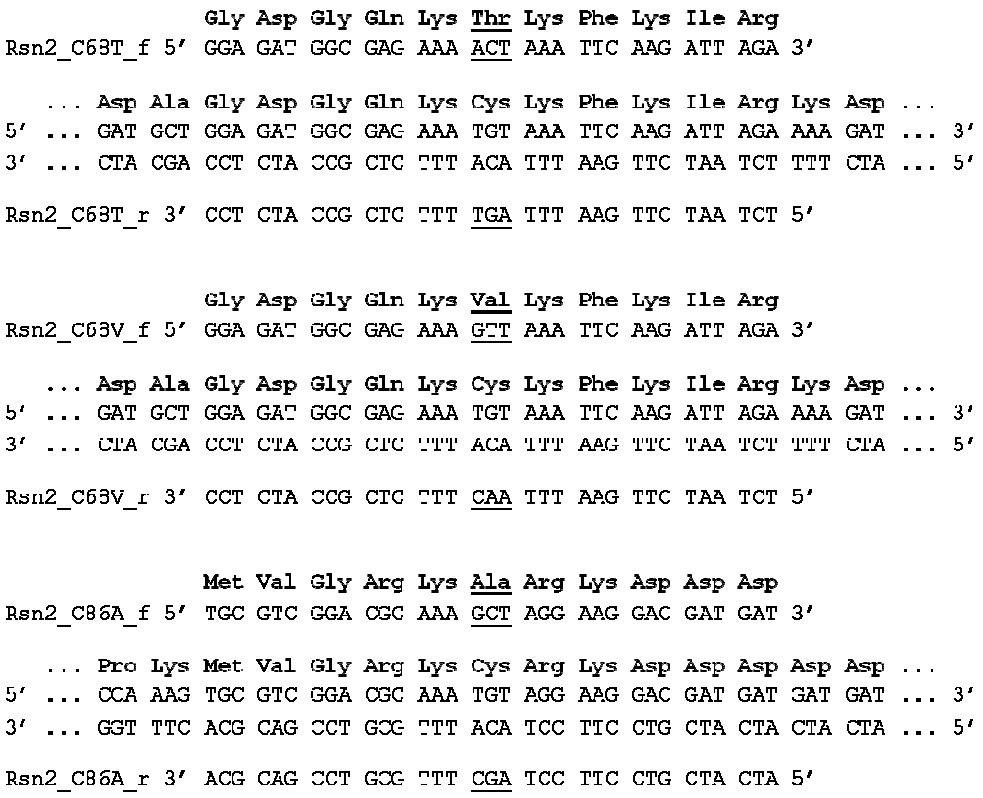


Figure 4.11- Rsn2_C68T_f, Rsn2_C68T_r, Rsn2_C68V_f, Rsn2_C68V_r, Rsn2_C86A_f & Rsn2_C86A_r primers and their annealing sites on the Rsn-2 gene. The mutant codons and resultant residues are underlined.

Site-directed mutagenesis was carried out using the relevant primers to create N19C-I46C, D32C-M81C, N19C-D32C-I46C-M81C, C68T-C86A and C68V-C86A mutant Rsn-2 genes. All plasmids were analysed by DNA sequencing to confirm the presence and correct orientation of the cloned gene. Routine issues were encountered in the site-directed mutagenesis that slowed the creation of the Rsn-2(C68V-C86A) gene. Meanwhile the Rsn-2(C68T-C86A) mutant had been produced and the mutation shown to have a minimal effect upon protein structure (discussed in section 4.2.4). Therefore effort to create the alternative Rsn-2(C68V-C86A) mutant was abandoned.

4.2.2 Removal of Rsn-2 Termini

An apparent degree of amphiphilicity can be observed in the primary sequence of Rsn-2, in particular a hydrophobic N-terminus and a hydrophilic C-terminus which includes the unique occurrence of six consecutive aspartic acids. This has led to the proposition that the two termini act to orientate the protein at the air-water interface, where the opposing properties of the termini provide a stretching force that facilitates unfolding to the active conformation [62]. To investigate the involvement of the termini in activity, mutants of Rsn-2 with the termini removed were created. Both termini are unstructured and dynamic in solution, making it unlikely that their removal would affect the structured region of the protein.

4.2.2.1 Rationale

The most obvious exclusively hydrophobic section of the N-terminus, or any other part of the protein, according to amino acid sequence, is the Leu-Ile-Leu at the extreme N-terminus (Figure 4.12). If the N-terminus does indeed act as a hydrophobic trigger to induce the proposed unfolding, these residues are likely to be key to that action and their removal should affect the activity of Rsn-2.

The N-terminus of Rsn-2 is unstructured until the start of the α-helix at V16. After the LIL the N-terminus is a combination of hydrophobic leucines and charged lysines and aspartic acids (Figure 4.12). To investigate if this region has any role in surfactant activity a mutant with the entire dynamic N-terminus dL1-P15 removed, was required.

The NMR studies found evidence indicating that the tyrosine at the extreme end of the C-terminus (Y96) was bound back onto the surface of the 4^{th} β-sheet [62]. It seems this interaction would restrain the otherwise dynamic C-terminus. The function of this feature is not understood but its affect upon activity was assessed by creating a mutant with the tyrosine removed.

The series of six aspartic acids at positions 89-94 form a motif not observed in any other protein sequence reported thus far. This anionic feature has been suggested as acting in the opposing manner to the hydrophobic N-terminus, pulling the protein into solution and hence aiding unfolding [62]. A mutant with residues D89-Y96 removed was created to investigate any role of the poly-aspartic acid motif upon activity.

In the hypothesis both termini act as triggers and therefore the removal of one terminus alone may not be enough to fully inhibit activity. Therefore a mutant with both termini removed was created to investigate the effects upon activity.

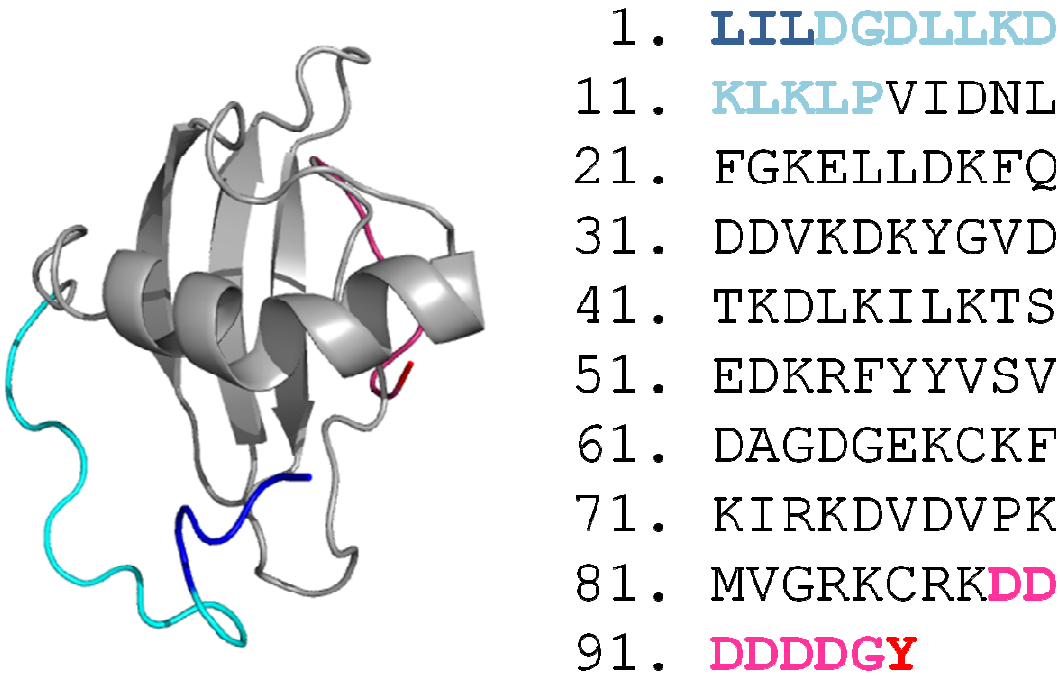


Figure 4.12- The location of termini removal mutations on both the structure and sequence of Rsn-2. The residues removed by the dL1-L3 mutation are shown in dark blue. The residues removed by the dL1-P15 are shown in light and dark blue. The residue removed by the dY96 mutation is shown in red. The residues removed by the dD89-Y96 are shown in pink and red.

4.2.2.2 Method

A similar approach to that used in the creation of the *Rsn-2*/pET-28 construct in Chapter 2 was employed for the removal of the termini. Alongside the primers designed for cloning of the full Rsn-2 gene (Chapter 2), new primers were designed to anneal to specific regions upon the Rsn-2 gene that would exclude the areas that coded for unwanted termini regions as well as to incorporate 5' NdeI or 3' BamHI restriction sites. These primers were then used within PCR cloning (as described in Chapter 2) to produce a mutated gene that was subsequently transferred into the expression vector pET-28.

The primer Rsn-2_dL1-L3_5p was designed for the creation of an Rsn-2 mutant with the LIL sequence removed (Figure 4.13).

```
                    His Met Asp Gly Asp Leu Leu Lys Asp Lys Leu
 Rsn2_dL1-L3_5p 5′ CAT ATG GAT GGG GAC CTA CTA AAG GAC AAG TTA 3′

5′ ... CAT ATG TTA ATA TTA GAT GGG GAC CTA CTA AAG GAC AAG TTA AAG ... 3′
3′ ... GTA TAC AAT TAT AAT CTA CCC CTG GAT GAT TTC CTG TTC AAT TTC ... 5′
   ... His Met Leu Ile Leu Asp Gly Asp Leu Leu Lys Asp Lys Leu Lys ...
```

Figure 4.13- The Rsn2_dL1-L3_5p primer and its annealing site on the Rsn-2 gene. The NdeI restriction site utilised for subcloning is underlined.

The primer Rsn-2_dL1-P15_5p was designed to create an Rsn-2 mutant with the entire N-terminus L1-P15 sequence removed (Figure 4.14).

```
                    His Met Val Ile Asp Asn Leu Phe Gly Lys Glu
Rsn2_dL1-P15_5p 5′ CAT ATG GTC ATT GAC AAT CTG TTT GGT AAA GAA 3′

5′ ... AAG TTA AAG TTA CCG GTC ATT GAC AAT CTG TTT GGT AAA GAA CTC ... 3′
3′ ... TTC AAT TTC AAT GGC CAG TAA CTG TTA GAC AAA CCA TTT CTT GAG ... 5′
   ... Lys Leu Lys Leu Pro Val Ile Asp Asn Leu Phe Gly Lys Glu Leu ...
```

Figure 4.14- The Rsn2_dL1-P15_5p primer and its annealing site on the Rsn-2 gene. The NdeI restriction site utilised for subcloning is underlined.

The primer Rsn2_dY96_3p was designed to create an Rsn-2 mutant with the C-terminal tyrosine removed (Figure 4.15).

```
   ... Lys Cys Arg Lys Asp Asp Asp Asp Asp Asp Gly Tyr Stop

5′ ... AAA TGT AGG AAG GAC GAT GAT GAT GAT GAT GGA TAT TAG GAT CCG ... 3′
3′ ... TTT ACA TCC TTC CTG CTA CTA CTA CTA CTA CCT ATA ATC CTA GGC ... 5′

          Rsn2_dY96_3p 3′ CTA CTA CTA CTA CTA CCT ATC CTA GG 5′

                         Asp Asp Asp Asp Asp Gly Stop
```

Figure 4.15- The Rsn2_dY96_3p primer and its annealing site on the Rsn-2 gene. The BamHI restriction site utilised for subcloning is underlined.

The primer Rsn-2_dD89-Y96_3p was designed for the creation of an Rsn-2 mutant with the poly aspartic acid motif as well as G95 and Y97 removed (Figure 4.14).

```
... Lys Met Val Gly Arg Lys Cys Arg Lys Asp Asp Asp Asp Asp Asp ...

5′ ... AAG ATG GTC GGA CGC AAA TGT AGG AAG GAC GAT GAT GAT GAT GAT ... 3′
3′ ... TTC TAC CAG CCT GCG TTT ACA TCC TTC CTG CTG CTG CTG CTG CTG ... 5′

   Rsn2_dD89-Y96_3p 3′ GCG TTT ACA TTC TTC ATC CTA GG 5′

                       Arg Lys Cys Arg Lys Stop
```

Figure 4.16- The Rsn2_dD89-Y96_3p primer and its annealing site on the Rsn-2 gene. The BamHI restriction site utilised for subcloning is underlined.

PCR was carried out using the relevant primers to produce dL1-L3, dL1-P15, dY96, dD89-Y96 and dL1-P15 & dD89-Y96 mutant Rsn-2 genes. All plasmids were analysed by DNA sequencing to confirm the presence and correct orientation of the cloned gene.

4.2.3 Mutant Rsn-2 Expression & Purification

All mutants were expressed and purified in accordance with the Rsn-2/pET-28/BL21(DE3) protocol outlined in Chapter 2. For some mutants containing multiple mutations, only low levels (0.5 – 1.0 mg) of protein were successfully expressed and purified.

4.2.4 Mutant Structure Validation

Before activity assays were carried out, the possible effects of mutagenesis were investigated to ensure that the overall structure of the protein had not been compromised and that the mutations had been introduced as required.

4.2.4.1 Circular Dichroism

Circular dichroism (CD) is a valuable technique for the study of low-resolution protein structure in solution. The principle behind CD is to measure the difference in absorbance of right and left-handed circularly polarised light by chiral chromophores and to extract structural information from the spectral features produced [181]. The composition of secondary structure, such as α-helix, β-sheet etc. can be estimated from the differential peptide bond absorbance, as the various conformations give rise to characteristic CD spectra in the far UV region [181]. The CD absorbance from aromatic residues in the near UV region provides information indicative of the environment of these residues and although analysis of near UV CD spectra alone is not sufficient to provide significant insight into tertiary structure, it does provide a distinctive fingerprint of the tertiary structure that can be utilised to monitor structural changes [181].

Although CD provides low resolution structural data compared with techniques such as X-ray crystallography and NMR, its benefits in terms of time, sample concentration and cost favoured CD as a rapid screening technique for structural assessment of the various Rsn-2 mutants. Although Rsn-2 contains no tryptophan residues, four tyrosine residues dispersed throughout the protein give rise to a significant absorbance at 280 nm (ε_{280}= 6085 $M^{-1}cm^{-1}$) and allow the production of reliable CD data from the near UV region.

CD was carried out with the assistance of Dr. S. Kelly (University of Glasgow). Samples were prepared at concentrations of 1 mg/mL in 20 mM sodium phosphate buffer at pH 7.5 for near UV CD measurements (0.5cm pathlength quartz cuvette) and diluted to 0.5mg/mL for far UV measurements (0.02cm pathlength quartz cuvette). Experiments were carried out using a Jasco J-810 spectropolarimeter. Spectra were averaged over 8 scans using a bandwidth of 1nm.

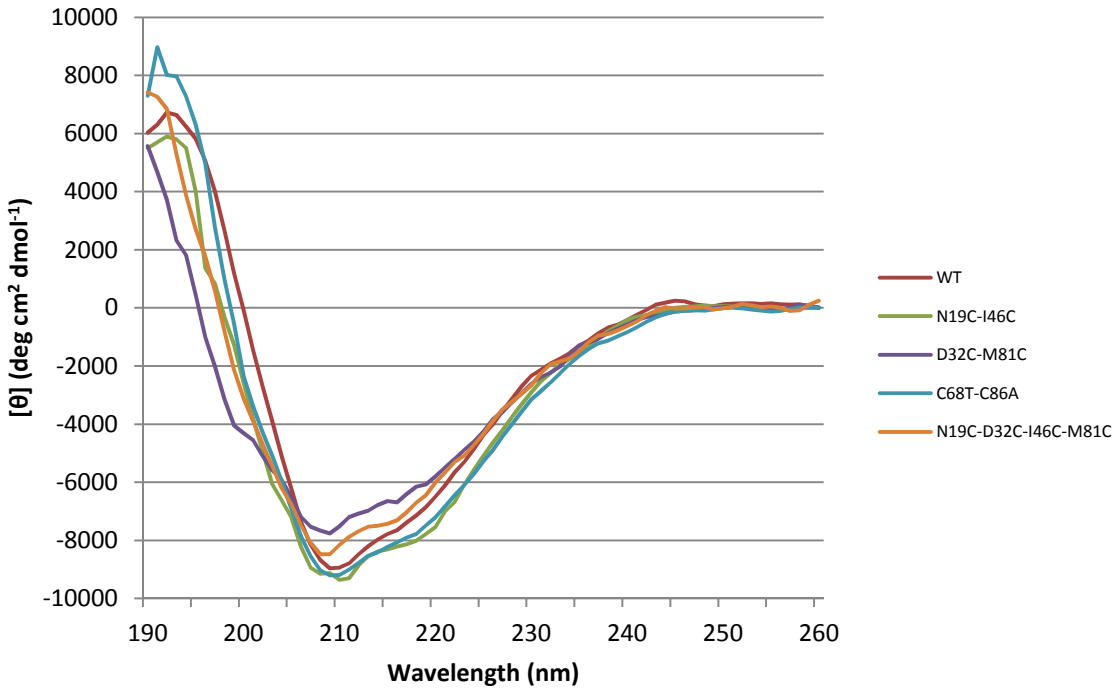


Figure 4.17- Far-UV CD of Rsn-2 (WT) and disulphide mutants.

The far UV spectrum of the various mutants shows little variation indicating the secondary structure composition had not been affected by the introduced mutations.

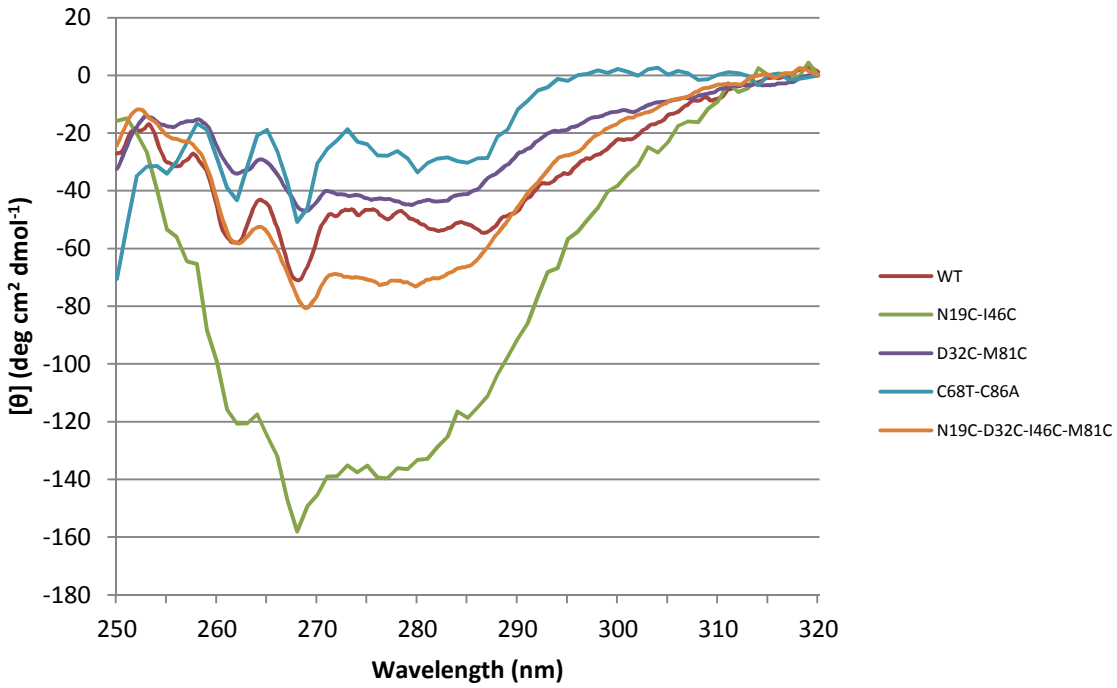


Figure 4.18- Near-UV CD of Rsn-2 (WT) and disulphide mutants.

The near UV spectral data for each of the mutants, exhibit peaks at the same wavelengths as the native protein indicating similar environments of the aromatic residues within the tertiary folds. There are however obvious variations in the intensities at these wavelengths, most obviously in the N19C-I46C mutant. It is possible that these intensity differences reflect changes in the structural rigidity of the protein following the introduction of the disulphide bonds. Increased rigidity will impede the motion of the tyrosine residues and hence increase their asymmetry which can lead to an increase in absorbance. The fact that the N19C-I46C mutant shows a greater effect may be a consequence of the introduced disulphide bond linking two structured regions of the protein. Whereas in the D32C-M81C mutant only one member of the cysteine pair is located on the structured helix, the other resides upon the flexible loop between strands 3 and 4. Contributions from the cysteine bonds themselves, which absorb at wavelengths around 260 nm, at intensities proportional to the dihedral angle of the disulphide bonds, could also cause the variation between the spectra [182].

It was concluded that the four mutants formed a similar overall fold to that of the wild-type protein although changes in the spectral intensities suggest that the introduction of disulphides alters the flexibility of the Rsn-2. It is probable that any effects upon the surfactant activity are as a result of the restricted motion of the molecule due to the introduced disulphides rather than improperly folded protein.

4.2.4.2 Disulphide Bond Formation

The insertion of two cysteines within the sequence does not necessarily result in the formation of a disulphide bond. The introduced residues may merely exist as reduced cysteines and hence have no effect upon any hypothesised conformational interchange. The presence of the disulphide bridge was therefore investigated by Ellman assay. This is a colorimetric assay for the quantification of free thiol groups. The reagent, 5-5'-dithio-bis-2-nitrobenzoic acid (DTNB), also known as Ellman's Reagent, reacts stoichiometrically with free thiol groups to yield a mixed disulphide and 2-nitro-5-thiobenzoic acid (TNB) (Figure 4.19). TNB has a high absorbance in the visible range (412nm) producing a yellow colour. By measuring the concentration of TNB, either from its molar extinction coefficient (ε_{412nm} = 14150 $M^{-1}cm^{-1}$) or from a standard curve of known concentrations of a thiol containing compound, the concentration of free cysteines can be determined.

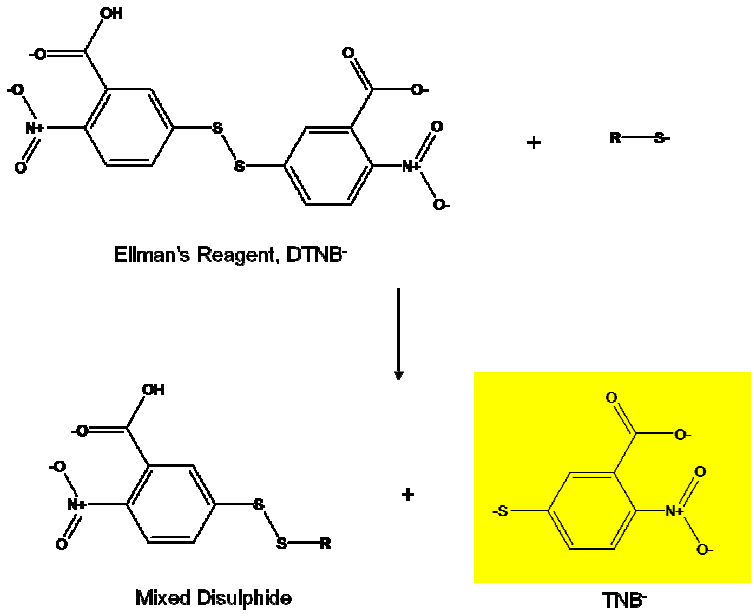


Figure 4.19- Mechanism of Ellman's reagent reacting with free thiol to produce yellow TNB.

Rsn-2 mutants where disulphides had been introduced, as well as the wild-type and C68T-C86A constructs were assayed at a concentration of 178 µM. No evidence of any significant degree of free thiol groups in any of the mutants was detected (Table 4.1).

Rsn-2 Mutant	A_{412nm}
WT	0.00
C68T-C86A	0.00
N19C-I46C	0.00
D32C-M81C	0.00
N19C-D32C-I46C-M81C	0.00

Table 4.1- Table showing free thiol concentration from Ellman assay. The free thiol concentration of each of the Rsn-2 disulphide mutants and the wild-type at a concentration of 178 µM was calculated from the average absorbance at 412 nm (A_{412nm}) from 3 samples and a standard calibration curve.

It appears that there are no free thiols within any of the Rsn-2 mutants indicating that all of the cysteine residues must be in the oxidised disulphide form. As the CD results indicate the three dimensional structure has not been significantly affected by the introduction of the disulphides it would appear that the cysteines have been paired as planned.

4.2.5 Surfactant Activity Assays

To assess the effects of the mutations upon surfactant activity an appropriate assay was required. As described earlier (Chapter 3) the ability to observe the development of activity over time via the Langmuir trough was deemed the most suitable technique for this work.

4.2.5.1 Method

In earlier work, Rsn-2 concentrations of 0.01 µg/mL, 0.1 µg/mL, 1µg/mL, 2.5 µg/mL, 5 µg/mL, 10 µg/mL, 100 µg/mL & 1000 µg/mL were measured. In order to make the data produced in this work as comparable as possible to previous data, this concentration scale was maintained. However, as each mutant was of a different molecular weight to that of the wild-type (in particular the termini removal mutations), to make results comparable between mutants the above concentration scale was converted to moles per litre. Since 1000 µg/mL Rsn-2 is equivalent to 89 µM, a scale of concentrations of 0.89 nM, 8.9 nM, 89 nM, 222 nM, 445 nM, 890 nM, 8.9 µM, 89 µM was used for all mutants where suitable concentrations were available. The samples at the various concentrations were created by serial dilutions into 20 mM NaPi, 50 mM NaCl, 1 mM azide, pH 7.5 buffer from a stock solution in the same buffer. Surface tension was measured over a period of 1000 seconds for several protein concentrations. In some cases surface tension had not reached equilibrium within this timescale, but over longer periods the effects of evaporation interfered with the measurement.

Ideally measurements would have been carried out on several samples from different preparations of each mutant. However, this approach would not have been practical within the time constraints of the project. Another concern is the accuracy of concentration determination. Small changes in the concentration of protein show large effects upon surfactant activity (Chapter 3). Therefore great care was taken to ensure consistency and accuracy in the preparation of all protein samples to be assayed. The entire methodology: production, sample preparation and measurement, was shown to be reproducible over several preparations of WT Rsn-2 (Chapter 3).

Although measurements at each of the above concentrations were carried out where possible for convenient comparison of the properties of each of the mutants, the concentrations 445 nM and 8.9 µM were deemed most informative. At 445 nM (5 µg/mL Rsn-2 (WT)), the rate of the onset of activity can be monitored. This is determined by rate of adsorption of the protein into the interface which must include any conformational exchange required for activity. At 8.9 µM the interface is saturated within the timescale of the experiment and therefore any differences between activities will indicate a change in the conformation of molecule adsorbed at the interface. This is also true for concentrations of 89 µM, however limiting concentrations prevented measurements at this concentration for some mutants.

4.2.5.2 Disulphide Mutants

The surfactant activity of each of the disulphide mutants, N19C-I46C, D32C-M81C, C68T-C86A and N19C-D32C-I46C-M81C was compared to that of the WT Rsn-2 at concentrations of 445 nM and 8.9 µM (Figure 4.20).

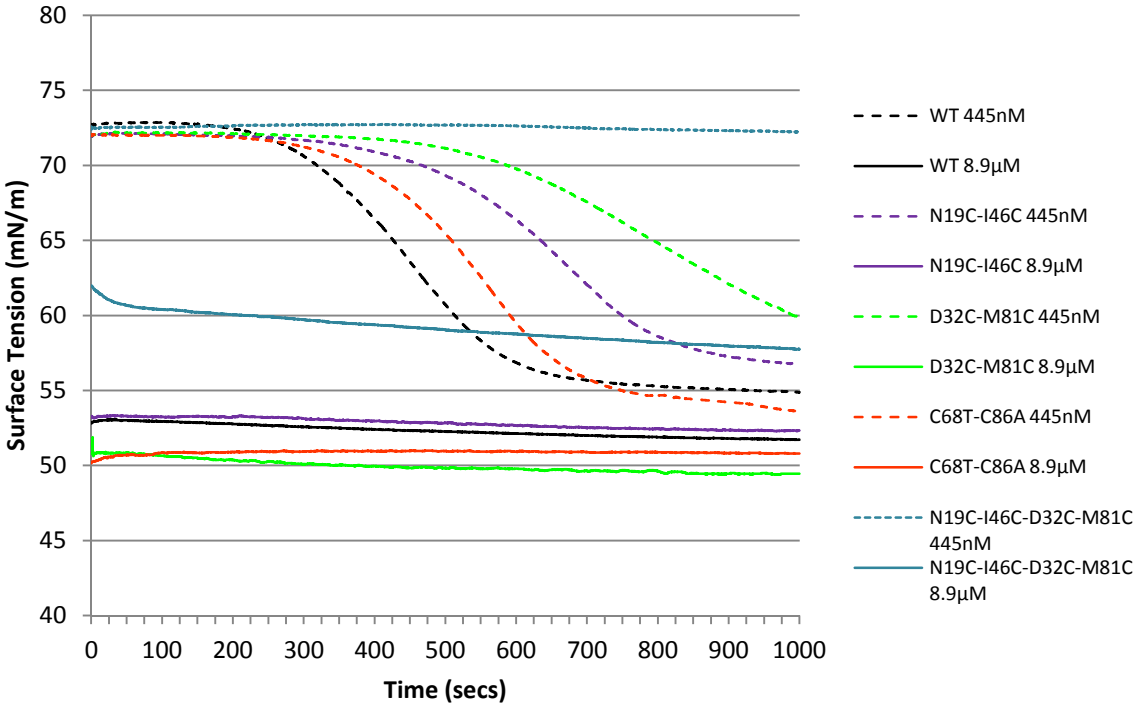


Figure 4.20- Graph showing the development of surface tension over time for the Rsn-2 disulphide mutants at concentrations of 445 pM & 8.9 µM.

Perhaps the most significant observations here are those involving the double-disulphide mutant, N19C-D32C-I46C-M81C. At a concentration of 445 nM there is complete loss of surfactant activity (on this time scale). Furthermore at 8.9 µM although surfactant activity has not been completely inhibited, the extent of activity (maximum reduction to 57 mN/m) is reduced relative to the WT (52 mN/m). These results indicate that the introduced disulphides do have a significant effect upon the surfactant activity of Rsn-2.

The effects of single-disulphide mutants are less clear cut. At lower concentrations (445nM), the onset of surface tension reduction is significantly delayed, with the greatest effect being shown by D32C-M81C. However, eventual surface tension reduction is not abolished. Indeed, at the higher concentrations, surfactant behaviour appears instantaneously within the timescale required to initiate the Langmuir trough measurements (< 10 sec), with surface tension reductions comparable with or, in some

cases, slightly greater than seen with wild-type. Both the D32C-M81C and C68T-C86A mutants produce a larger total reduction in surface tension at 8.9 µM than Rsn-2 WT.

It is, of course, possible in principle that the observed effects here might not be due solely to conformational constraints imposed by –SS- cross-linking, but rather to other indirect effects arising from mutation of what may be key residues. One approach would be to investigate the effects of alternative mutations upon each of the N19, D32, I46, M81 sites upon activity. However this would require extensive effort to create and produce all of the mutations. An alternative approach of analysing activity of the mutants with disulphide bonds reduced was deemed more practical. The results above indicate that the native C68-C86 disulphide bond has a fairly small effect upon activity and the reduction of this disulphide should therefore have little effect, and any changes in activity would be as a consequence of the additional freedom from the reduction of the incorporated disulphide bond. 0.1 M dithiothreitol (DTT) was added to various mutants to probe the effects of reduction upon activity. After unexpected or difficult to rationalise results were observed for numerous reduced mutant samples, DTT at a concentration of 0.1M was found to have some effect upon surface tension of its own that supplemented the protein activity, demonstrated by its effect upon the activity of the C68T-C86A mutant which contains no cysteine residues (Figure 4.21). Therefore it was impossible to determine whether the affects upon activity were a result of the reduced mutant disulphides or the effect of DTT acting alongside the surfactant proteins. In hindsight 0.1 M DTT was likely to be somewhat excessive and lower concentrations may not have had the same effect. Additionally, alternative reducing agents such as tris(2-carboxyethyl)phosphine could have been utilised, however due to time restraints this was never investigated.

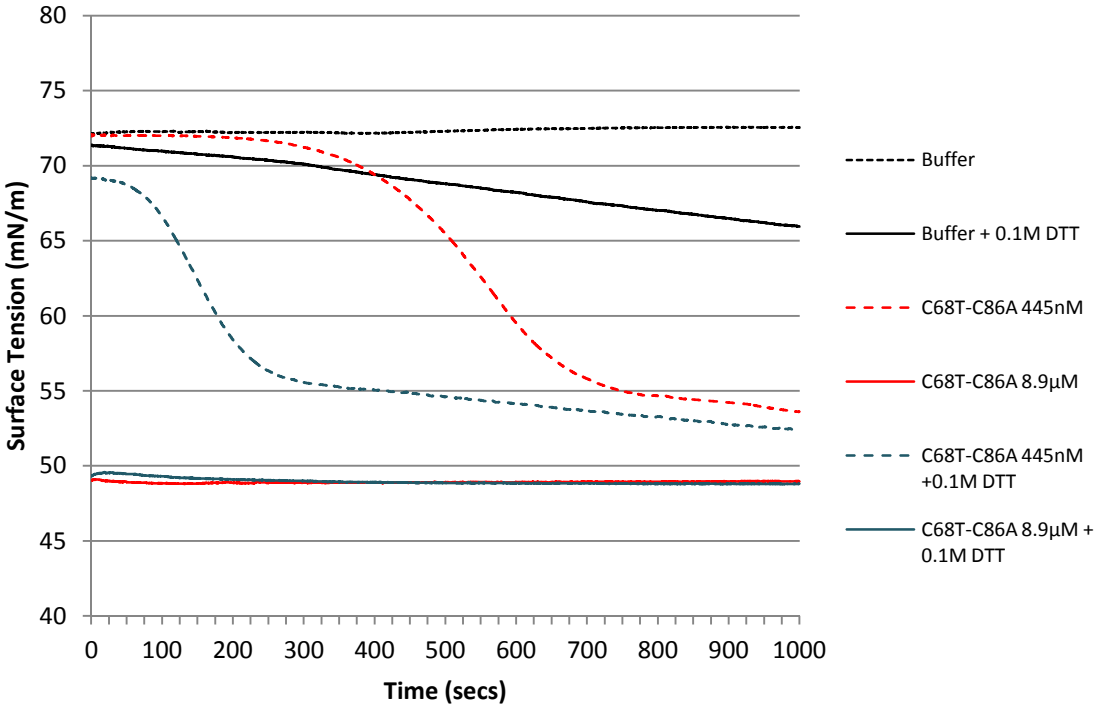


Figure 4.21- The effect of 0.1M DTT upon surface tension and the activity of surfactant proteins. 0.1M DTT was found to reduce the surface tension of the buffer used in the surfactant assays. The ability of this activity to supplement that of Rsn-2 is shown by its effect upon the C68T-C86A mutant, which contains no cysteine residues, at a concentration of 445nM.

4.2.5.3 Termini Mutants

The surfactant activity of each of the termini mutants, dL1-L3, dL1-P15, dY96, dD89-Y96 and dL1-P15 & dD89-Y96 was assessed and compared to that of the WT Rsn-2 at concentrations of 445 nM and 8.9 µM to investigate specific roles for each of the motifs within surfactant activity (Figure 4.22).

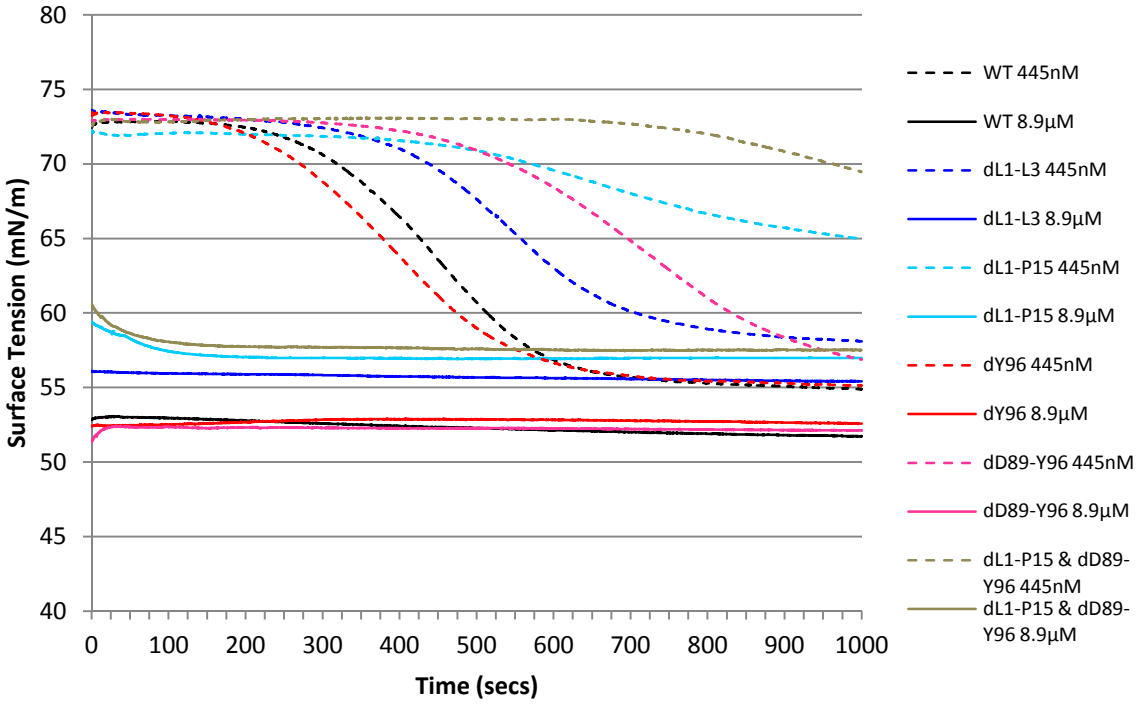


Figure 4.22- The development of surface tension over time for the Rsn-2 termini mutants at concentrations of 445 nM & 89 µM.

At a concentration of 445 nM the dL1-L3 mutant not only shows a reduced onset of activity but a reduced overall activity. This reduced activity is also observed at the higher concentration. Rsn-2 mutant dL1-P15 shows an even greater reduction in onset of activity than the dL1-L3 mutant. At a concentration of 8.9 µM there is a small difference between the minimum surface tension observed in the dL1-L3 (55 mN/m) and dL1-P15 (57 mN/m) mutants. Both of which show a reduced effect compared to the WT at the same concentration (52 mN/m).

The dY96 mutation does not appear to have any significant effect upon surfactant activity. At both concentrations it produced near identical results to that of the WT Rsn-2. Similarly, dD89-Y96 reaches final surface tensions indistinguishable from WT Rsn-2 at both concentrations. However at a concentration of 445 nM a significant reduction in the rate at which the maximum activity is achieved is observed.

The dL1-P15 & dD89-Y96 combination mutant at 445 nM produces the slowest rate of onset of activity of all the termini mutants, where activity only begins to develop after 700 seconds. The minimum surface tension reached by this mutant (57.5 mN/m) is very similar to that of the dL1-P15 mutant alone further supporting the observation that the C-terminal mutants have no effect upon final surfactant activity.

4.2.6 Discussion

The insertion of single mutant disulphides into the Rsn-2 structure, although having little effect on the degree of surfactant activity, does affect the rate of onset of activity at low concentrations. This may indicate that the introduced disulphides do affect the kinetics of any unfolding processes but do not inhibit the protein from achieving its active form. The N19C-D32C-I46C-M81C mutant which contains two inserted disulphides shows no surfactant activity at the low concentration within the time-scale of the experiment. This supports the above idea and indicates that the additional restraint from the extra disulphide has an additive effect on activity. This mutant also shows a large effect upon the extent of reduction in surface tension at both high and low concentrations which is not the case for either of the single disulphide mutants. The final surfactant activity is not however completely abolished, causing a maximum reduction in surface tension to 57 mN/m, nor is it reduced to the levels measured for the standard proteins at the same concentration (60 mN/m). This suggests that the N19C-D32C-I46C-M81C still retains some degree of specialised surfactant activity.

From the results it appears that the Rsn-2 mutants are able to form surface active states, of varying degrees, despite the additional restraints enforced by the disulphide bonds. Proteins are highly dynamic molecules and it is possible that the force of hydrophobic adsorption into the interface is sufficient to reorient the globular protein into a more amphiphilic conformation by various unfolding pathways. By imposing more restraint upon the molecule the dynamics towards and form of, the active conformation appear to be limited, with an effect upon activity. This observation could be probed by the addition of further disulphides. However with increasing numbers of cysteines issues with protein folding and indirect effects upon activity are more likely introduced. The results discussed thus far, do appear to be consistent with the hypothesis that the molecule must undergo significant structural rearrangement in order to function as a surfactant.

A consideration with regards to these observations is whether the observed effects upon activity are a direct result of the introduced disulphide bonds or as a result of the properties introduced or removed by the mutation of the native residues to cysteines. As suggested above, controls where the same residues are mutated to non-cysteine alternatives may be useful to probe this issue as would assays of reduced versions of the mutants. Another concern was whether the disulphide bonds were maintained at the interface. Although the redox state of the cysteine residues was investigated prior to testing surfactant activity, it is possible that these bonds were broken upon adsorption into the air-water interface, allowing full conformational mobility. However as air is an oxidative environment where the formation of disulphide bonds would be favoured, this seems unlikely.

It appears that the native disulphide has no role with respect to the surfactant activity, and is therefore more likely to be involved in structural stability in solution.

The results of the termini mutations are less contentious than those of the disulphide mutations as the simple subtraction of the residues in question provides a more direct comparison to the wild-type protein and therefore a greater confidence in the source of any effects upon activity.

A reduced degree of activity was observed for both N-terminal mutations. This would suggest that some motif directly involved in surfactant activity is contained within the N-terminus. As the greatest effect upon activity occurs upon deletion of the L1-L3 motif it appears that the hydrophobic LIL sequence at the very N-terminus is primarily responsible. The deletion of the following 12 residues, removed only in the dL1-P15 mutant, has a much smaller effect. This contribution towards the level of activity would fit in with the idea that the hydrophobic N-terminus is adsorbed into the interface. Although both N-terminal mutants show a reduction in rate of onset when compared to the WT Rsn-2, the extent of this effect is greater in the dL1-P15 mutant. This suggests that not only the LIL sequence but the entire dynamic N-terminus is involved in promoting activity. This is somewhat surprising as this region contains a range of aliphatic, acidic and basic residues and may indicate a more complicated action than the proposed hydrophobic trigger at the extreme N-terminus.

At the C-terminus the unusual motif of the six consecutive aspartic acids was also shown to facilitate the onset of activity which would support the proposition that this motif is required to orientate the molecule at the air-water interface and provide a stretching force

to aid unfolding of the protein [62]. The C-terminal tyrosine residue (Y96) does not show any influence upon activity and therefore its role, if any, remains unknown. The fact that neither mutation at the C-terminus has any effect upon the final surfactant activity of Rsn-2 supports the idea that the C-terminus remains within the solvent and that the solvent exposed face is sufficiently hydrophilic without the poly-aspartic acid motif.

The dL1-P15 & dD89-Y96 Rsn-2 mutant produced the greatest effect upon surfactant activity of any of the termini mutations. This confirms the previous results and supports the hypothesis that both termini are involved in activity.

Overall it is apparent that a number of residues/motifs located throughout the protein play an active role in Rsn-2 activity. Mutagenesis as an approach to investigate function is more commonly applied to proteins such as enzymes, where only a small number of residues are directly involved in activity. In these examples a mutation to a single residue can have significant effect upon activity. Thus in Rsn-2, where the observed activity is believed to be a result of several, if not the majority, of residues within the sequence, it is extremely difficult to establish clear conclusions from the mutagenic method applied here. This situation is more akin to investigations of protein folding where the contributions of individual residues or motifs are more difficult to deconvolve from one another. Despite this, residues at both termini have been confirmed as playing a key role within activity, and the overall picture is reasonably consistent (or at least not inconsistent) with the simple clam-shell opening model.

5 Rsn-2 Molecular Engineering

Rsn-2's surfactant activity presents numerous potential applications. One potential technique to take advantage of this activity is to co-express Rsn-2 as a fusion tag with selected proteins to create functional surfactants/biofoams. The intricate relationship between activity and structure, demonstrated in Chapter 4, requires that we investigate how such a modification may affect the activity of Rsn-2. This work also provides an opportunity to tag Rsn-2 in a manner that allows the protein to be studied in ways that are not possible with native protein. The fluorescent reporter iLOV was identified as an ideal candidate as it may allow the visualisation of Rsn-2 behaviour in solution.

5.1 iLOV

iLOV is a flavin-based fluorescent protein derived from the LOV (light, oxygen or voltage-sensing) domain of *Arabidopis thaliana* phototropin 2 [183]. The LOV domains form a covalent adduct between a bound flavin mononucleotide (FMN) and a conserved cysteine residue (Cys^{426} of *Arabidopis* phototropin 2) following excitation by blue light. This adduct formation causes a structural change in the LOV domain that in turn initiates a subsequent response in an effector domain. On removal of the photoactive cysteine by mutagenesis (C426A), the adduct formation is inhibited. The initial photoproduct is, however, still formed and subsequently decays back into the ground state resulting in increased fluorescence emission compared to the wild-type. DNA shuffling of the C426A mutant was used to identify three additional mutations in close proximity to the FMN cofactor that gave a further 2-fold increase in fluorescence emission. This mutant was named iLOV. Engineered as an alternative to green fluorescent protein (GFP), iLOV is proposed to offer various advantages over GFP. The significant reduction in size, 10 kDa vs. 25 kDa, not only drastically reduces the steric impedance caused by the marker but also the genetic strain upon the expression system. iLOV is naturally monomeric and is not prone to aggregation unlike GFP. Another benefit is iLOV's ability to recover spontaneously from photobleaching, unlike GFP.

5.2 Molecular Biology

The iLOV gene was supplied in the form of an *E.coli* codon optimised expression plasmid by Dr Andy Roe (University of Glasgow). The gene was flanked by NcoI restriction sites allowing simple subcloning into a single NcoI site located at the 5' end of the Rsn-2 gene in the *Rsn-2*/pET28 plasmid described in Chapter 2. The presence of NcoI sticky ends at both ends of the *iLOV* insert did not allow for directional cloning. Several positive transformants were screened by DNA sequencing to isolate plasmids containing correctly orientated *iLOV*.

A construct for producing untagged iLOV as a control was created by subcloning the *iLOV* insert into an unmodified pET-28b vector, digested at its single NcoI site. This created a vector identical to the iLOV-Rsn-2/pET-28 construct apart from the absence of the Rsn-2 gene.

5.3 Protein Production

5.3.1 Expression and Purification

Both iLOV-Rsn-2 and iLOV proteins were expressed in the same manner as Rsn-2 described in Chapter 2. It was deemed unnecessary to remove the poly-histidine tag, as the very aim of the experiment was to investigate how Rsn-2 behaved in the presence of a fusion tag. Therefore, fusion tag cleavage and cleaved tag removal steps were omitted from the purification strategy.

5.3.2 Quantification

It was observed that the absorbance from the iLOV chromophore at approximately 270 nm [184] overlapped the 280 nm signal rendering any estimate of concentration from conventional UV/Vis absorbance impractical (Figure 5.1). Other measurements, such as Bradford assays, were available but were deemed too insensitive for the highly concentration dependant, surfactant experiments.

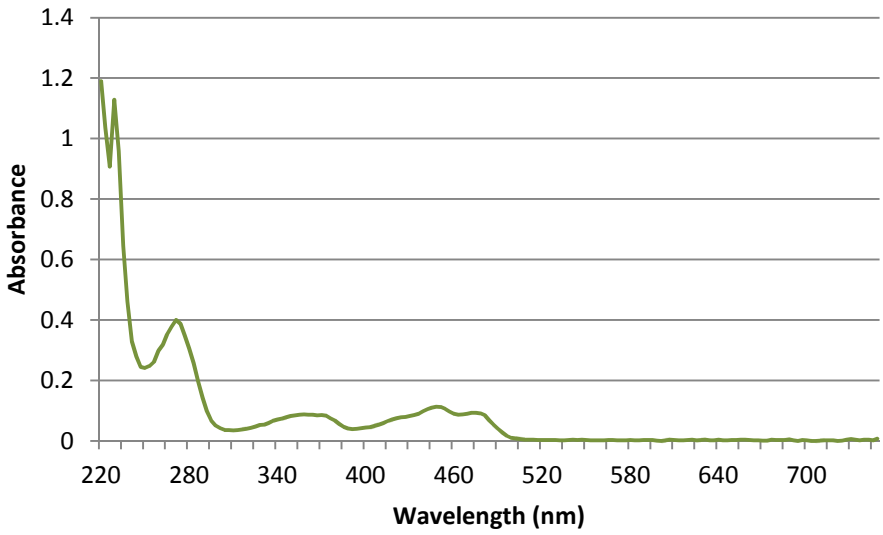


Figure 5.1- Absorption spectrum of iLOV. The absorption of aromatic residues at a wavelength of 280 nm is obscured by the absorption of the fluorophore at 270 nm.

To overcome these issues the concentration of the FMN present in Rsn-2-iLOV and iLOV samples was measured. A volume of iLOV-Rsn-2 was heated at 85˚C for 5 minutes to denature the protein and release the bound FMN. The solution was then centrifuged at 20000 xG for 5 minutes to pellet the denatured protein. The concentration of free FMN in the supernatant was then calculated based on a molar extinction coefficient of 12500 ± 500 M^{-1} cm^{-1} at 450 nm [185]. The concentration of free flavin mononucleotide is equimolar to the concentration of iLOV producing an accurate measurement of protein concentration.

5.4 Surfactant Activity

To investigate whether the presence of iLOV affects Rsn-2's surfactant activity, various tests were carried out. Initially a simple assay was performed by pumping air into a solution of iLOV-Rsn-2 fusion protein to observe if foam was formed. A large stable column of foam was produced indicating that at least some surfactant activity remained. The foam also fluoresced under UV illumination confirming that the iLOV activity had not been affected.

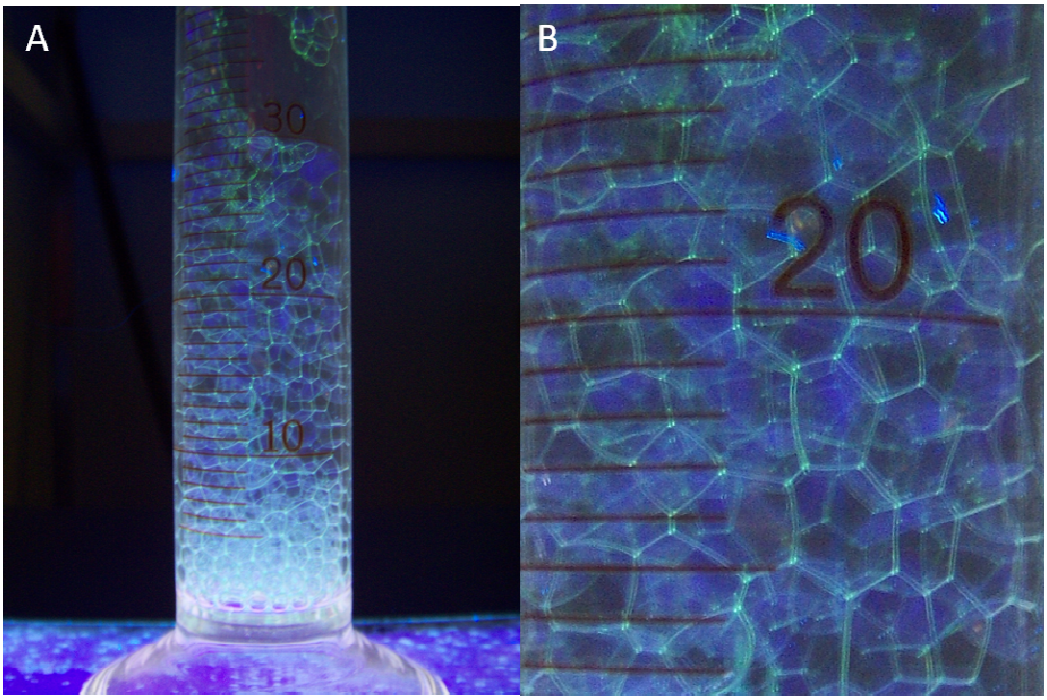


Figure 5.2- iLOV-Rsn-2 foam under UV illumination. Image A shows the entire foam, whereas image B shows a close up of the individual bubbles. Protein solution was at a concentration of 89 μM in 20 NaPi, 50 mM NaCl, 1mM NaN_3, pH 7.5.

The Langmuir trough as a tensiometer method was then used to measure the surfactant activity of iLOV-Rsn-2 and iLOV alone as a control, to estimate what activity, if any, the tag was contributing. The results were compared to unmodified Rsn-2 to investigate any differences between the two proteins (Figure 5.3).

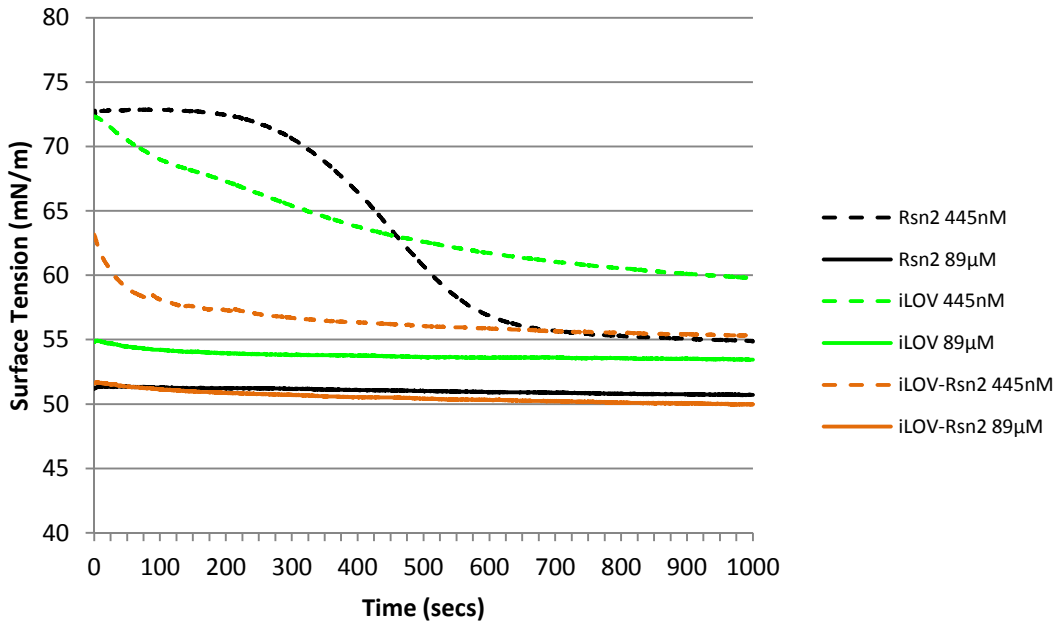


Figure 5.3- Surface activity of Rsn-2, iLOV and iLOV-Rsn-2 fusion protein. Graph depicts surface tension development over time for solutions of Rsn-2, iLOV and iLOV-Rsn-2 at concentrations of 445 nM & 89 µM.

The iLOV protein alone was found to show reasonable levels of surfactant activity surpassing that observed for conventional protein standards, BSA and lysozyme in Chapter 3. At a concentration of 445 nm, activity develops from zero down to a minimum tension of approximately 60 mN/m. At the higher concentration of iLOV a minimum surface tension of 54 mN/m is achieved instantaneously. Comparison of the activity of the iLOV-Rsn-2 fusion protein to that of untagged Rsn-2 at a concentration of 89 µM, indicates that the maximum activity of Rsn-2 is not affected by the presence of the iLOV fusion partner. An intriguing observation is that at a concentration of 445 nM the rate of onset of activity is greater for the iLOV-Rsn-2 than for Rsn-2 alone. There appears to be some additive effect between the two proteins, perhaps as a consequence of the intrinsic surfactant activity of iLOV alone. It is also apparent from Figure 5.3 that while the rate at which the activity appears is increased, as observed at the higher concentration, the final level achieved is identical to that of the untagged Rsn-2. This may suggest that although the iLOV tag increases the speed that Rsn-2 adsorbs and/or unfolds into the surface that the iLOV itself is not at the interface contributing to a reduction in surface tension.

A serious concern with the addition of an N-terminal fusion tag was that the hydrophobic N-terminus of Rsn-2, which has been shown to be key to both the rate of onset and final level of surfactant activity (Chapter 4), may be impeded from adsorption into the interface. This does not appear to be the case.

It is clear from the results that the surfactant activity of Rsn-2 in not inhibited by the addition of iLOV at the N-terminus and that from this experiment Rsn-2 has potential as a fusion tag for the creation of functional surfactants or foams. More work is however required to establish how the surfactant activity of Rsn-2 is affected by a broad range of fusion partners.

5.5 Fluorescent Imaging

The iLOV tag allowed the location and relative concentration of Rsn-2 to be monitored visually. Previous work utilised the fluorescent probe 1-anilinonaphthalene-8-sulphonic acid (ANS) that binds to hydrophobic patches on proteins with a characteristic intensity and wavelength shift [186], to probe the concentration difference of Rsn-2 at the interface versus the bulk solution [130]. An issue with this experiment was that the hydrophobic ANS was itself likely to show preference for the interface regardless of the location of the protein. The iLOV-Rsn-2 fusion gives a direct indication of the location of the fusion protein.

Rsn-2 alone, whilst highly surface active, is not capable of producing foams stable over the timescales required for effective imaging. The native foam mixture on the other hand produces foams that are stable over several days. Therefore, in order to produce stable iLOV-Rsn-2 foams, the native foam mixture was spiked with iLOV-Rsn-2. An additional advantage of using the native foam mixture is that it produces wet foam with thick sections of liquid between bubbles that allows for greater distinction between the interface and bulk phase of the foam as oppose to the dry foam produced from Rsn-2 alone (Figure 5.2). Two samples of approximately 0.5 mL (total volume) of Tungara frog foam nest were used in this work. One was spiked with 5 µL 89 µM iLOV-Rsn-2, the other with 5 µL 89 µM iLOV as a control. The two samples were left at room temperature for 30 minutes to allow the protein to disperse throughout the foam.

Fluorescent imaging of the two samples was carried out with assistance from Dr. Volodymyr Nechyporuk-Zloy (Infection, Immunity and Inflammation Life Sciences, University of Glasgow), who performed all microscopy. Imaging was performed on a DeltaVision Core microscope (Applied Precision, USA), with 20x objective, under LED illumination. A stack of acquired images (65 slices over a 200 nm range) was recorded via a Fluorescein isothiocyanate (FITC) emission filter using a CoolSnap HQ2 camera

(Photometrics, USA). Images were deconvolved utilising softWoRx software using the conservative algorithm (Applied Precision, USA) to improve resolution.

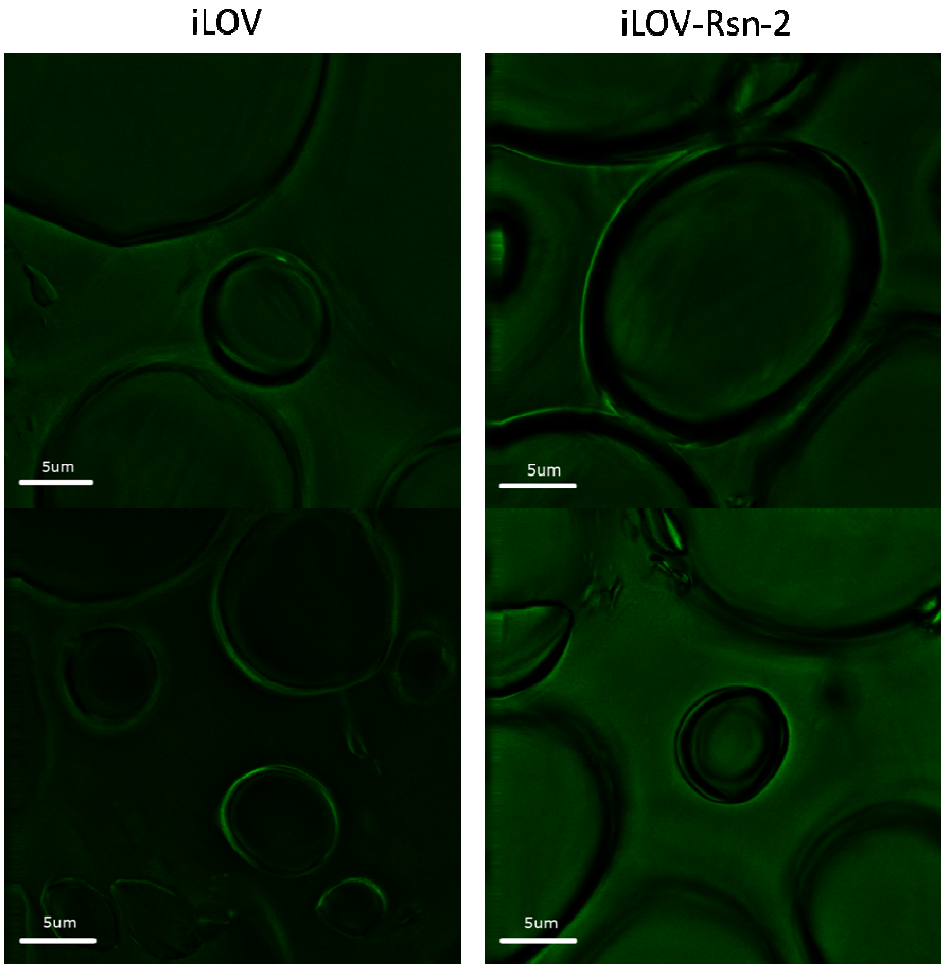


Figure 5.4- Fluorescence microscopy of iLOV-Rsn-2 and iLOV spiked frog foam. Two images of each sample are shown. Images represent a 3 nm slice of the foam. Lighter areas of the images indicate increased fluorescence emission.

The captured images (Figure 5.4) do show an increase in fluorescence at the area immediately surrounding the interface. However there is no clear discernible difference between the foam containing iLOV-Rsn-2 and the iLOV control. One possible explanation may be the surfactant activity shown by the iLOV protein alone. If the iLOV protein is indeed intrinsically surface active, it would be expected that it would, like Rsn-2 adsorb at the interface.

Inspection of the images, which should represent 3nm slices through the foam, would also suggest levels of fluorescence arising from within the air cavities within the bubbles. The slices are sufficiently thin to only sample a small cross section of the bubble and therefore any fluorescence from the surface of the bubbles should be removed leaving the centre of the bubbles dark. This may indicate that the imaging technique, and subsequent

deconvolution software utilised are not competent to handle the complex light refraction properties of the foams. These effects, rather than increased concentrations of iLOV protein, may cause the apparent fluorescence at the interface. Due to time constraints, further optimisation of the methodology was not possible. Alternative techniques such as two-photon excitation fluorescence may have improved the quality of the images producing more representative results.

5.6 Conclusions

A fusion of iLOV and Rsn-2 has been designed and successfully produced. The addition of the fusion partner, iLOV has no detrimental effect upon the surfactant activity of Rsn-2, confirming the potential of this approach for the creation of functional surfactants or biofoams. The iLOV-Rsn-2 protein is capable of producing fluorescent foams. Attempts to further understand the activity of Rsn-2 utilising fluorescent imaging techniques indicated increased fluorescence at the air-water interface for both iLOV-Rsn-2 and the iLOV control. Further work is required in order to improve the imaging technique in order to allow the study of Rsn-2 in foam.

6 Latherin Structure Determination

Latherin has, prior to the work described in this thesis, been identified as a surfactant protein but only limited biophysical data was available with regards to the structure of the protein. Without high resolution structural information, it is difficult to gain an insight into the molecular basis for the proteins activity. Latherin displays moderate levels of sequence similarity to the PLUNC family of proteins, found throughout the mammalian respiratory tract and postulated to have some role in the innate immune response. No structure has thus far been reported for any member of the PLUNC family, making latherin an attractive target for structure determination.

6.1 Previous Work

Attempts to crystallise both recombinant latherin as well as latherin purified directly from horse sweat were unsuccessful. This prevented any investigation into the structure of latherin by X-ray crystallography. NMR structure determination was identified as an alternative approach. NMR structural assignment of latherin had been initiated previously by Dr. R. McDonald. [187]. Dr. McDonald identified a buffer composition of 20 mM NaPi, 50 mM NaCl, pH 7.5 and temperature of 310 K as optimum conditions at which to obtain well resolved spectra and reached the stage of backbone assignment. However limiting sample concentrations impeded further progress. In this work higher yields and sample concentrations were achieved and a co-purified latherin aggregate identified and removed.

6.2 Sample Preparation

Uniformly (^{15}N) and double (^{15}N & ^{13}C) labelled samples of latherin for NMR were expressed in M9 minimal media enriched with the relevant isotope(s) and purified following the procedure described in Chapter 2. Pure, labelled latherin samples were exchanged into 20 mM NaPi, 50 mM NaCl, 1 mM NaN_3, pH 7.5, concentrated to 570 µL, mixed with 30 µL D_2O for field frequency locking and transferred into a 5 mm diameter NMR tube.

6.3 NMR Data Processing

6.3.1 DFT Based

Conventionally sampled spectra were processed using the *Azara* suite of programmes (Dr. Wayne Boucher, www.ccpn.ac.uk/azara). Initial processing scripts were created from the Bruker data files via the programme *ReferenceB* (Dr. Krystyna Bromek, University of Glasgow).

All data was fitted with an appropriate window function and zerofilled to optimise signal resolution and minimise truncation artefacts before the free induction decay (FID) was Fourier transformed from the time, into the frequency domain. Phase correction was then carried out on the transformed spectrum. Where required, baseline correction or convolution difference to remove residual water signal was applied.

6.3.2 Alternatives to the direct Fourier transformation

In multidimensional NMR, due to limitations on the amount of data that can be recorded, it is not always possible to record the FID until it has decayed to zero. Direct Fourier transform of this truncated data can lead to artefacts in the resultant spectrum. Alternatives to the direct Fourier transformation that generate spectra with the effects of these artefacts reduced have been developed. Popular methods include: maximum entropy [188], linear prediction [189], the filter diagonalization method [190], multidimensional Fourier transformation [191] and multidimensional decomposition [192].

In this work, maximum entropy (MaxEnt) processing was employed as an alternative to Direct Fourier transformation for processing truncated spectra. MaxEnt processing creates various trial frequency domains which it then inverse Fourier transforms and compares with the measured data. Of the possible simulated frequency domains that are consistent with the recorded FID the one with the highest entropy is selected [188].

6.3.3 Non-Uniform Sampling

Multidimensional NMR experiments have been established as key techniques in the study of protein structure in solution. A limitation of these experiments is the time required to record the number of points in the indirectly sampled dimensions required to produce well resolved spectra. As a consequence many weeks or even months can be devoted to obtaining all the multidimensional spectra required for a structure determination.

One approach that allows the multidimensional spectra to be recorded in an accelerated fashion is to employ non-uniform sampling. In conventional multidimensional NMR experiments, regularly spaced incrementation of delays is employed to frequency encode the signal in the indirect dimension to provide sufficient digital resolution in the indirect dimensions. With each additional dimension the number of data points sampled in the indirect dimensions increases rapidly as does the time taken for collection. The prospect of non-uniform sampling was first introduced by Barna *et al*. [193]. Rather than sampling every time point for both indirect dimensions, a reduced number of time-domain data points from the full set is collected. Due to the missing data points Direct Fourier transformation of the sampled data produces spectra with a high level of artefacts. However by utilising MaxEnt processing to reconstruct the data, these artefacts can be avoided. It has been demonstrated that via the selection of an appropriate sampling schedule and utilising MaxEnt reconstruction, non-uniform sampling can produce spectra of equal quality to that of conventional sampling within much shorter time scales [194].

6.4 Sequence Specific Resonance Assignment

The process of structure calculation from NMR data relies upon the ability to identify the resonance frequency of almost all the individual nuclei in a protein. As the amino acid sequence for latherin is known we effectively have a map of the protein sequence to which observed resonances must be assigned. All assignment was carried out using the CCPNMR Analysis suite of programmes [179].

6.4.1 Sequential Backbone Resonance Assignment

Root amide resonances were established from a ^{15}N-HSQC of latherin. The ^{15}N-HSQC correlates resonances of covalently bonded ^{1}H and ^{15}N nuclei. In a protein this includes: any side chain amides from glutamines and asparagines, side chain amines in arginines, histidine and tryptophan, and the backbone amides of which there is one for every amino acid in the protein, with the exception of prolines. Likely side chain amide peaks were identified based upon characteristic chemical shifts and the occurrence of pairs of peaks at identical ^{15}N resonances. The side chain amine belonging to the lone tryptophan was identified by its characteristic chemical shift and later confirmed through inspection of the NOESY spectra. The remaining anonymous amide peaks were picked and assigned as a spin system with associated ^{1}H and ^{15}N amide resonances.

Triple resonance heteronuclear experiments, adapted for non-uniform sampling by Dr Brian Smith (University of Glasgow), were then used to identify specific resonances correlated to the amide resonances. Complementary pairs of experiments allowed for identification of both resonances for the same, and the preceding residue as the amide group (intra and inter-residue (i and i-1)) resonances. For example, HNCACB [195] and CBCA(CO)NH [196] experiments were used to correlate intraresidue and inter-residue C_α and C_β resonances to the amide crosspeaks. In the HNCACB experiment magnetisation is transferred between the amide resonances and C_α/C_β resonances of both the same, and the preceding residue. The partner experiment, in this case a CBCA(CO)NH, discriminates between the intra and inter-residue correlations by directing magnetisation through the carbonyl group of the preceding residue, and thus only identifying the inter-residue C_α and C_β resonances. Thus, by using the pair of experiments, the identity of both the intra and inter (i-1) C_α and C_β resonances can be determined and added to the appropriate spin systems. In this work three pairs of experiments were used: HNCACB and CBCA(CO)NH discussed already, HNCO and HN(CA)CO to identify intra and inter (i-1) residue backbone carbonyl carbon correlations [197], and HBHA(CBCA)HN and HBHA(CBCACO)NH to indentify intra and inter (i-1) residue H_α and H_β correlations [196].

The three pairs of experiments provide three complementary routes for sequential assignment. This reduces the occurrences of ambiguity caused by chemical shift degeneracy which due to relatively large size of the protein and the high number of leucine residues was an issue in latherin. Using this combination of spectra, sequential assignment

was carried out with the aid of the *protein sequence assignment* popup within the *Analysis* programme [179]. The software identifies and ranks spin systems with intraresidue resonances that match the inter-residue resonances of a query spin system. In instances where an unambiguous match is identified, spin systems can be linked sequentially (Figure 6.1).

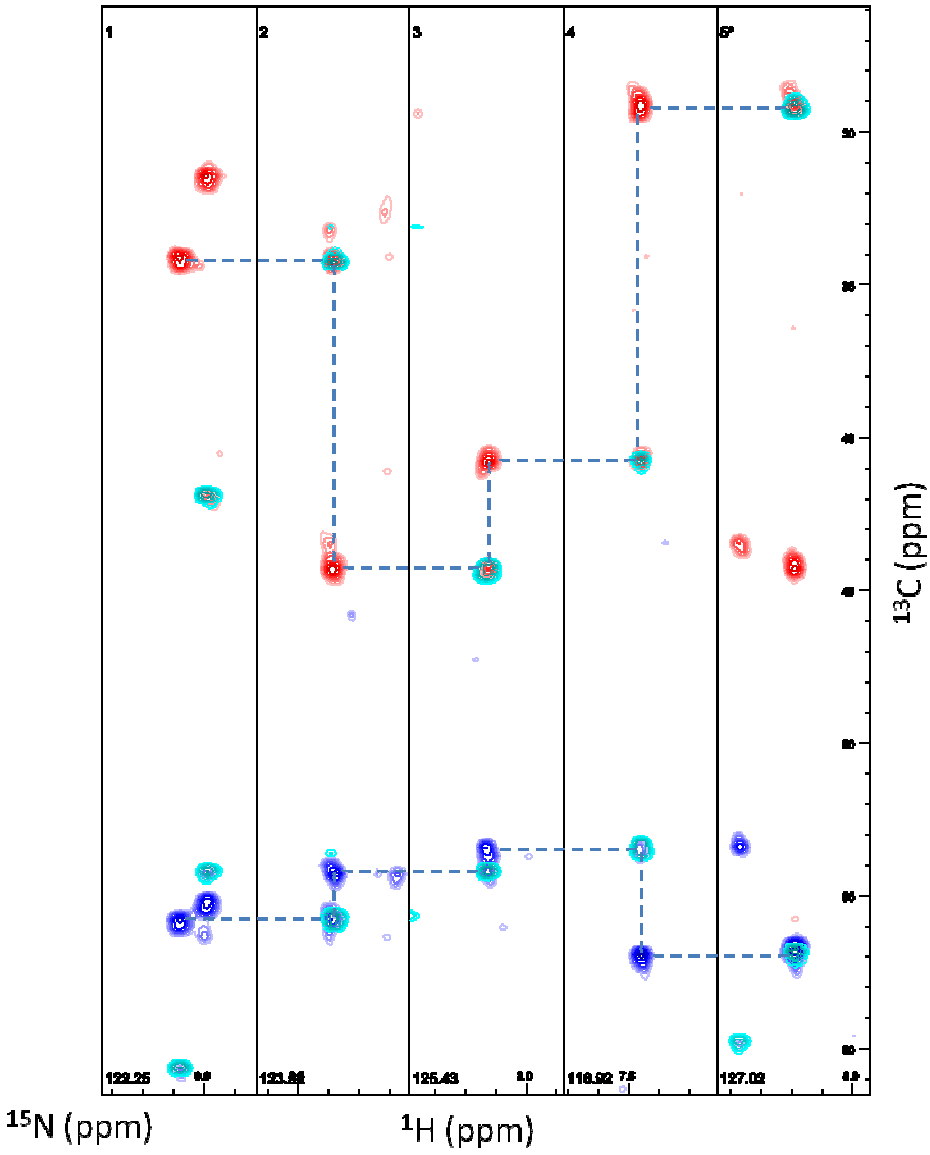


Figure 6.1- HNCACB and CBCA(CO)NH spectra used for linking sequential spin systems. Strips correspond to amide resonances of consecutive spin systems. The ^{15}N chemical shift is shown in the bottom left hand corner of the strips. The HNCACB experiment, which identifies both intra and inter (i-1) residue C_α and C_β resonances, is shown in blue for C_α and red for C_β. The CBCA(CO)NH experiment which identifies inter (i-1) C_α and C_β resonances is shown in cyan.

Characteristic C_α and C_β, and to a lesser extent H_α and H_β chemical shifts allowed for predictions of amino acid type. These predictions coupled with the protein's amino acid sequence, allowed sequentially linked spin systems to be assigned to specific sections of

the protein's backbone. This information was sufficient to identify over 90% of the protein backbone. The assignments obtained at this stage were compared to those made by Dr McDonald in the previous work [187]. The two sets of assignments were found to be in good agreement.

6.4.2 Side Chain Assignment

The assignment of side chain resonances carried out for this work can be separated on the basis of strategy used into two main sections: aliphatic and aromatic side chain assignment.

6.4.2.1 Aliphatic Side Chain Assignment

Conventionally, experiments such as HCCH-COSY and HCCH-TOCSY, which identify coupled aliphatic resonances based upon networks of J-coupling in an amino acid side chain, are utilised for the assignment of side chain resonances. Due to the high number of leucine residues present in latherin it was postulated that HCCH-COSY and HCCH-TOCSY spectra would be excessively overlapped in regions containing leucine resonances for complete resonance assignment to be achieved. In order to overcome the effects arising from latherin's unusual sequence, two sets of experiments, that are adaptations of conventional carbon TOCSY experiments were utilised for aliphatic side chain assignment in latherin: HCC(CO)NH-TOCSYs and $HCCH_3$-TOCSYs.

Two 3D HCC(CO)NH-TOCSY experiments: a (H)C(CCO)NH-TOCSY and H(C)(CCO)NH-TOCSY were used [198,199,200]. In both the (H)C(CCO)NH-TOCSY and H(C)(CCO)NH-TOCSY magnetisation is transferred between aliphatic side chain carbons under isotropic mixing, and then, via the carbonyl carbon, to the amide resonance of the following (i+1) residue. In the (H)C(CCO)NH-TOCSY experiment, magnetisation is allowed to evolve under influence of the aliphatic carbon resonances' chemical shift, whereas the in H(C)(CCO)NH-TOCSY, it evolves under the influence of the aliphatic proton resonances' shift. The HC(CCO)NH-TOCSY experiments therefore correlate backbone amide resonances to side chain aliphatic proton and carbon resonances of the preceding (i-1) residue. Specific assignment of side chain aliphatic resonances for each residue was carried out based upon reference chemical shifts.

As the sequence and resonance assignment of the majority of backbone amide resonances had already been determined, the side chain resonances were assigned to the relevant

residue (Figure 6.2). The HCC(CO)NH-TOCSY experiments were also useful to confirm the previous sequence and resonance assignments.

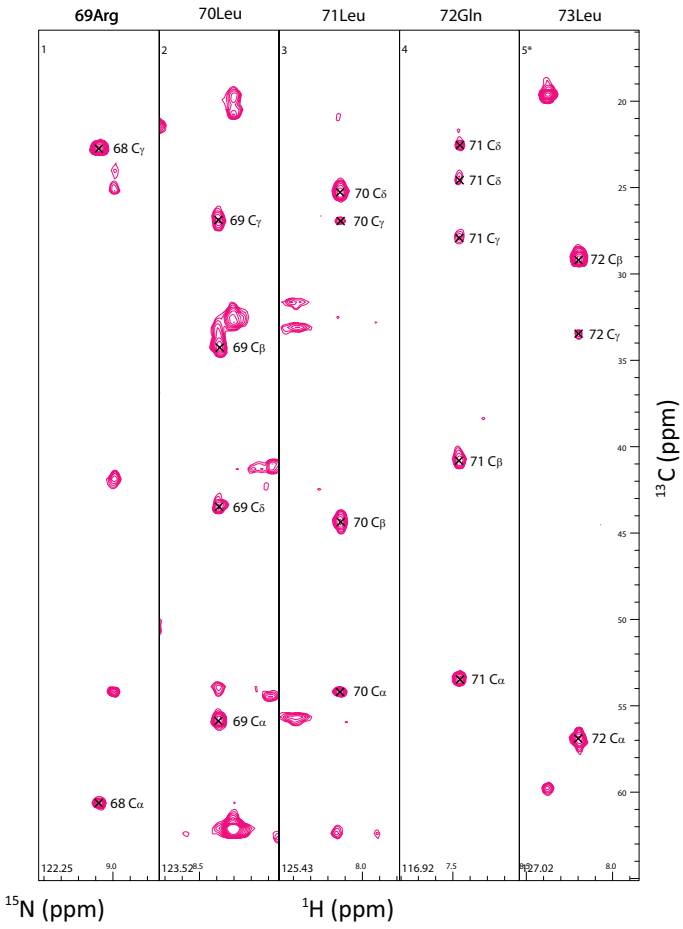


Figure 6.2- Spectra demonstrating the results of (H)CC(CO)NH-TOCSY experiment. Consecutive strips at the backbone amide resonances of 69Arg, 70Leu, 71Leu, 72Gln and 73Leu are shown. The spectra allows for the assignment of side chain aliphatic hydrogen resonances of the preceding (i-1) residue in the sequence.

Although the HCC(CO)NH-TOCSY experiments were useful for the assignment of the majority of side chain resonances, some issues were encountered that impeded full side chain assignment in particular residue types. Due to the oscillatory nature of magnetisation transfer during isotropic mixing [201], a single mixing time does not result in equal transfer to all carbons in a side chain. In this work a mixing time of 12 ms was used which yielded intense α and β crosspeaks, but crosspeaks to more distal coupled spins were less intense. This made it difficult to fully assign side chain resonances in leucines, valines, isoleucines, lysines and arginines. Another issue was poor resolution of similar chemical shifts in the side chains of leucines, which prevented full assignment. These issues were overcome through the use of methyl specific TOCSY experiments, $HCCH_3$-TOCSY [202].

The $HCCH_3$-TOCSY experiments are similar to the HCC(CO)NH-TOCSY experiments described, however rather than transfer to the amide resonances, magnetisation is selectively detected on CH_3 carbon and proton resonances. As in the HCC(CO)NH-TOCSY experiments, either proton or carbon resonances of the aliphatic side chains can be recorded in 3D H(C)CH3-TOCSY and (H)CCH3-TOCSY experiments respectively. The experiment can also be modified for the removal of CH_2 or both CH_2 and CH from the methyl proton–carbon planes [202].

The methyl specific experiments offer three main advantages over conventional HCCH-TOCSY experiments: reduced levels of overlap due to the removal of non-methyl peaks; highly resolved spectra due to the unique relaxation properties of methyl groups; and higher digital resolution is achieved by measuring a higher number of points over a narrower sweep width.

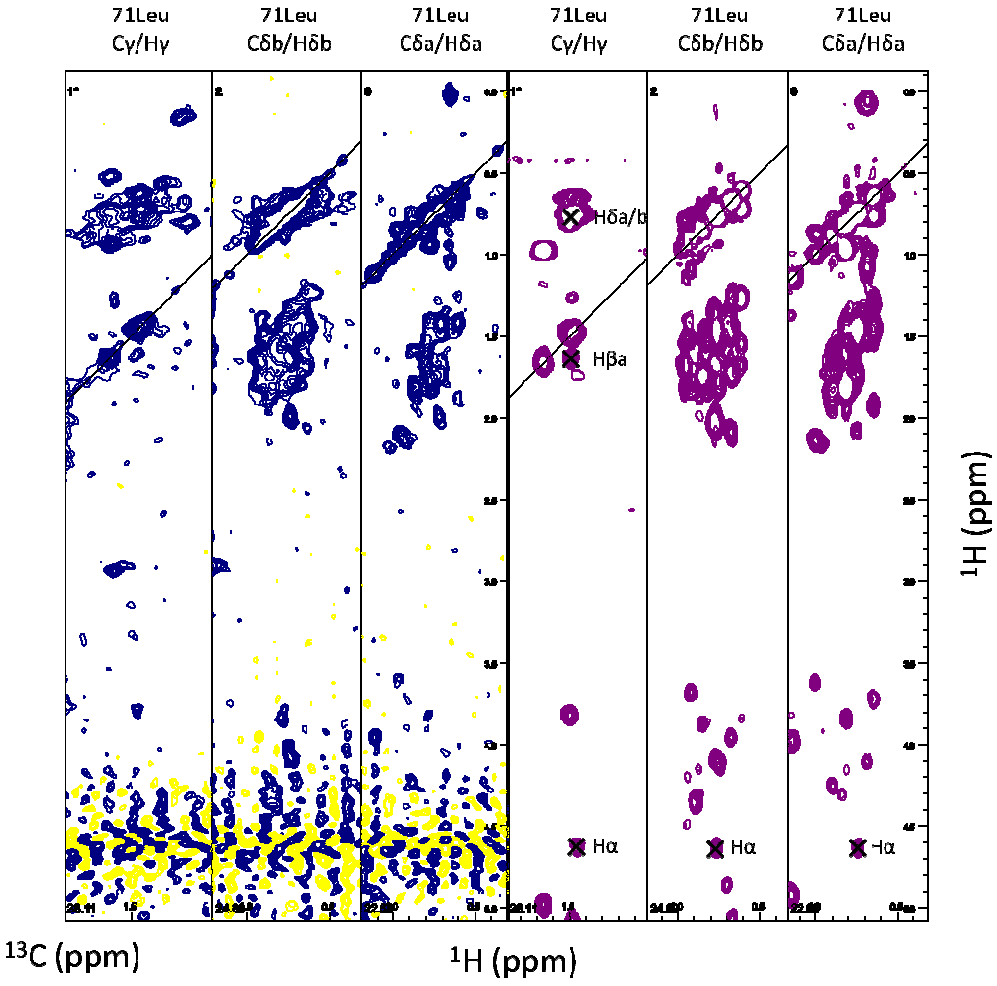


Figure 6.3- Comparison of conventional (left) and methyl-specific (right) TOCSY experiments. Strips correspond to the ^{1}H and ^{13}C resonances of the CH and CH_3 groups in 71Leu. The ^{13}C chemical shift is shown in the bottom left hand corner of the strips.

In this work, $HCCH_3$-TOCSY experiments with CH_2 resonances excluded were deemed most appropriate for the assignment of leucine. The high resolution spectra produced allowed full assignment of all leucine, isoleucine and valine residues and confirmed previous alanine and threonine side chain assignments. An unexpected feature of the $HCCH_3$-TOCSY spectra was the presence of C_ε and H_ε resonances of lysines residues, which escaped the multiple quantum filter and allowed for the assignment of the aliphatic portion of lysine side chains.

6.4.2.2 Aromatic Side Chain Assignment

Aromatic side chain resonances were assigned using 3D ^{13}C-NOESY spectra with the aid of 2D (HB)CB(CGCD)HD and 2D (HB)CB(CGCDCE)HE experiments, where the H_δ and H_ε resonances respectively, are correlated to intra residue C_β resonances by the transfer of magnetisation through scalar couplings [203]. Due to the relatively low level of aromatic residues in the latherin sequence (4 phenylalanines, 3 tyrosines and 1 tryptophan) the 2D spectra was sufficiently resolved for the assignment of H_δ and H_ε resonances to the appropriate residue. The remaining aromatic side chain assignments were carried out using NOESY spectra.

6.4.3 NOESY Spectra Assignment

In an attempt to decrease the ambiguity of NOE derived restraints used in subsequent structure calculations, any assignments that could be made with confidence in the NOESY spectra were completed manually. This was particularly useful for the leucine residues as some resonances that were within the set tolerances for creation of ambiguous restraints, could be separated by eye. The NOESY experiments also allowed the identification of the majority of remaining unassigned resonances from proline and arginine side chains as well as side chain amides.

6.4.4 Assignment Summary

From the experiments described, 206 of the 211 (97.63%) latherin residues were assigned at least partially. A full summary of the chemical shift assignments is shown in Appendix A. The majority of unassigned residues were concentrated between residues 143-155. No candidate amide resonances for many of these residues could be detected. Possible explanations could be spectral overlap, amide protons exchanging with water or broadened spectra due to chemical exchange. That backbone dynamics data (Chapter 7) indicates that

a combination of exchanging amide protons and chemical exchange are the most likely explanations. A summary of the assignment is shown in Table 6.1.

Resonance Type	Available	Assigned	% Assigned
Carbon atoms	1015	904	89
Hydrogen atoms	1238	1168	94
Nitrogen atoms	264	205	78
Amide	410	378	92
Backbone	832	778	94
Backbone non-H	633	589	93
Side chain H	1039	979	94
Side chain non-H	646	520	81

Table 6.1- The completeness of latherin assignment.

The vast majority of the assigned resonances displayed chemical shifts within the known distribution of shifts for that atom type [204]. However, there were a few exceptions (Table 6.2). The data contributing to the resonances were re-examined, and in each case the assignment was confirmed.

Resonance	δ (ppm)	Reference Chemical Shift Distribution (ppm)
85 Asp $H_{\beta a}$	1.618	2.2 – 3.2
113 Arg $H_{\beta a}$	0.708	1.2 – 2.2
113 Arg $H_{\gamma a}$	-0.269	1.0 – 2.0
113 Arg $H_{\gamma b}$	0.369	1.0 – 2.0
135 Leu C_{γ}	31.087	24.5 – 29.0
180 Asn H_{α}	3.246	4.0 – 5.3
180 Asn $H_{\beta a}$	1.570	2.2 – 3.3

Table 6.2- Latherin resonances with atypical chemical shifts. Reference chemical shift distributions were extracted from chemical shift distributions within CCPN Analysis based upon data from the RefDB database [204].

Subsequent analysis of the latherin structures allowed for the observed atypical shifts to be rationalised. 85Asp, 113Arg and 180Asn all display the effects of ring current shift due to their close proximity to aromatic residues: 85 Asp and 113Arg to 87Trp and 180Asn to 130Phe. 135Leu is however buried within the hydrophobic core of the protein surrounded by other aliphatic residues. Analysis of the stereochemical properties of the residue in calculated structures indicated φ, Ψ, χ_1 and χ_2 bond angles in favourable regions. No satisfactory explanation of the atypical chemical shift in 135Leu was encountered.

6.5 Restraints for Structure Calculation

Many of the parameters that can be measured by NMR spectroscopy are sensitive in some manner to molecular conformation. By quantifying these parameters individual restraints that together describe the three dimensional form of a protein can be derived. It is these restraints that form the basis of protein structure determination by NMR. In this work several types of restraints were used.

6.5.1 NOE Restraints

The most commonly used restraints in protein structure determination are NOE restraints as they contain long range internuclear connectivities with defined distances. NOE restraints are created as a consequence of the nuclear Overhauser effect (NOE) between spins. The nuclear Overhauser effect arises from the cross-relaxation of dipolar coupled nuclei that are close together in space. By disturbing the population distribution between states of one nucleus from equilibration, cross-relaxation processes induce a change in the population distribution of its dipolar coupled partner nucleus that can be observed as a change in signal intensity. The rate of cross relaxation (σ) between two nuclei A and B can be described by:

$$\sigma_{AB} = \zeta {r_{AB}}^{-6} \tag{6.1}$$

where r_{AB} is the internuclear distance and ζ is a correlation function, dependent upon the rotational correlation time, τ_c, of the internuclear vector [205]. After perturbation of a target nucleus from equilibrium and a subsequent delay τ_{mix} where cross-relaxation processes are allowed to occur, magnetisation can be transferred to neighbouring nuclei. Furthermore, as equation 6.1 describes, the efficiency of this transfer is proportional to the inverse sixth power of the internuclear distance. Therefore the intensity of the resultant peaks can be used to estimate the distance between the two nuclei and for ^{1}H to ^{1}H NOE, nuclei within 5 Å of the target nucleus can be identified.

A basic 2D ^{1}H NOESY experiment involves a series of three 90° proton pulses with two different delays between them (Figure 6.4). The first pulse transfers magnetisation into the xy-plane where incremented delay τ_1 is used to encode the chemical shift of the excited proton in the indirect dimension. The second pulse returns the magnetisation to the z-axis. During the NOE mixing time τ_{mix}, the cross-relaxation occurs between dipolar coupled

nuclei in a distance dependant manner. The final pulse transfers the magnetisation of the destination proton into the xy-plane for detection.

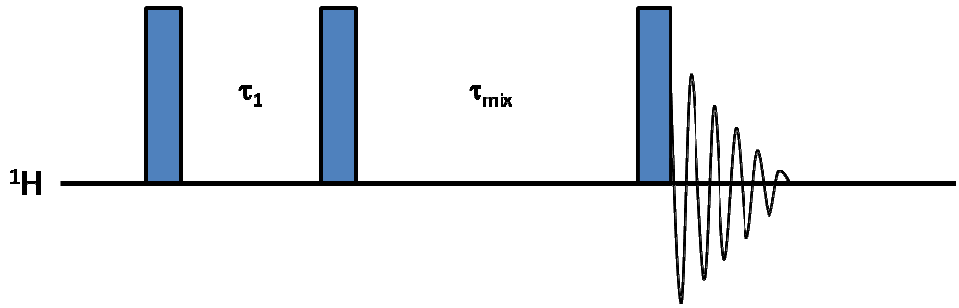


Figure 6.4- Pulse sequence of basic NOESY experiment. 90° RF pulses are represented by rectangles. The first two pulses are separated the evolution time τ_1. A delay of τ_{mix} then follows where cross-relaxation processes are allowed to occur. The final pulse transfers magnetisation to the xy-plane for detection.

In this work NOE restraints were collected from 3D ^{13}C-NOESY-HSQC and 3D ^{15}N-NOESY-HSQC experiments [206]. These experiments are essentially a NOESY pulse sequence (Figure 6.4) followed by an HSQC sequence to transfer the magnetisation to the relevant attached heteronuclei. Experiments were performed with a mixing time of 100 ms to provide adequate cross-peak intensities with tolerable spin diffusion for a molecule of this size [151].

NOE distance restraints were created within CCPNMR *Analysis* [179] by converting the observed peak height to a restrained internuclear distance using a calibrated distance function:

$$r_{AB} = r_{ref} \left(\frac{I_{ref}}{I_{AB}}\right)^{1/6} \tag{6.2}$$

The reference values were calculated using the intensities corresponding to a number of Hα(i) to HN(i+3) NOE peaks in an a region identified as α-helix based upon characteristic patterns of sequential and short range NOEs. In a regular α-helix these nuclei have a typical internuclear distance of 3.4 Å. These values are subject to further recalibration within the structure calculation process.

6.5.2 Dihedral Angle Restraints

Dihedral angle restraints, although adding little additional information in structures that are well defined by other data, can provide valuable structural information in areas where experimental data is sparse, where distance restraints are ambiguous and aid convergence of structure calculations [207].

Backbone dihedral angles were predicted using the Dihedral Angles from Global Likelihood Estimates (DANGLE) software [207], within CCPNMR *Analysis*. DANGLE predicts backbone φ and Ψ angles and secondary structure based upon sequence data and backbone chemical shifts [207]. Each residue is assessed in the context of its two preceding and two following amino acids. The sequence and chemical shift data (where available) for each five amino acid section, is analysed against a database of five residue fragments from solution protein structures to identify the ten closest matches. Analysis of the ten matches as well as predefined ranges based upon specific residue types (glycines, prolines, residues preceding prolines and the most commonly encountered, generic case which includes all other residues) are used to predict the query φ and Ψ angles. Prediction of secondary structure is carried out based upon secondary structures of the ten matches.

As DANGLE depends upon dihedral angles sampled from a database of known structures, the accuracy of the prediction therefore depends upon how representative the database is, rather than on direct experimental data. Consequently an under-represented backbone conformation in the database may result in inappropriate restraints being placed upon structure calculations. To prevent any such effects, dihedral restraints were used in high temperature steps of each simulated annealing procedure to aid convergence. However, they were omitted from final cooling steps (by setting the force constant to zero) to prevent biasing the calculated structures.

6.5.3 RDCs

NOE measurements provide semi-quantitative distance information for pairs of hydrogen nuclei in close proximity (< 5Å) in the protein. Errors in these short-distance restraints can lead to large accumulated errors over the length of a large molecule. To counter this, residual dipolar couplings (RDCs) can be used to define the relative orientation of specific bonds within the protein with regard to a molecular frame of reference. RDCs therefore act as long distance restraints that can improve the quality of the calculated structure.

The dipolar coupling between two nuclei can be described by:

$$D_{AB}(\theta) = \mathrm{D}_0\,(1 - 3cos^2\theta) \tag{6.3}$$

where θ, is the angle between the A-B internuclear bond vector and the magnetic field and D_0 depends on the internuclear distance between the two nuclei and the gyromagnetic ratios of each of the nuclei. The dipolar coupling contains information that could be potentially useful in the structure determination process. However, in solution the molecule tumbles rapidly, sampling all possible orientations of θ and as a result D_{AB} averages away to zero. This effect, although beneficial for the majority of NMR experiments as it increases the sensitivity and reduces the complexity of the resultant spectra, means that all the structural information contained in the dipolar couplings is lost. This averaging and loss of the dipolar coupling can be inhibited by influencing the molecule's tumbling so as to introduce a very slight preference for a particular orientation with respect to the magnetic field. This can be achieved using an alignment media such as liquid crystals or stretched gels where the molecule is orientated in a particular way due to electrostatic or steric interactions, but is still allowed to tumble at a sufficient rate to produce well resolved spectra.

For two dipole-coupled nuclei A and B within an anisotropically tumbling molecule the observed dipolar coupling, D_{AB}, can be described by [208]:

$$D_{AB}(\theta,\phi) = A_a^{AB}\left\{(3cos^2\theta - 1) + \frac{3}{2}R(sin^2\theta cos2\phi)\right\} \tag{6.4}$$

where A_a^{AB} and R are the axial and rhombic components, respectively, of the alignment tensor; θ is the angle between the A-B internuclear bond vector and the z-axis of the alignment tensor; and ϕ is the angle between the projection of the A-B internuclear bond vector onto the x-y plane and the x-axis. A_a^{AB} is equal to:

$$A_a^{AB} = -\left(\frac{\mu_0 h}{16\pi^3}\right) S\gamma_A\gamma_B\langle r_{AB}^{-3}\rangle A_a \tag{6.5}$$

where μ_0 is the permeability of a vacuum, h is Plank's constant, S is the order parameter, γ_A and γ_B are the gyromagnetic ratios of nuclei A and B respectively and A_a is the axial ratio of the alignment tensor. When observing a single type of dipolar coupling, in this case between N-H nuclei, almost all of these values remain constant and the only values that affect the RDC are the θ and ϕ angles i.e. the orientation of the internuclear vector with regards to the alignment tensor.

In this work Pf1 Filamentous phage (Profos, # 311079) was used as the alignment medium. In a magnetic field, the rod-like (60 Å wide by 20000 Å long) phage align parallel to the field, restricting the tumbling of surrounding molecules [209]. The extent of the alignment can be assessed by investigating the deuterium splitting. Quadrupolar nuclei such as deuterium have a non-spherical distribution of charge within the nucleus that is anisotropic with respect to the external field. Like dipolar couplings, in an isotropic environment the quadrupolar interaction is averaged away, however it can be observed in an aligned environment e.g. deuterium nuclei in HDO solvent molecules. The observed quadrupolar splitting in the 1D deuterium spectrum is approximately linearly correlated to the phage concentration. At a concentration of 5 mg/mL, pf1 phage, an HDO splitting of 10 Hz was observed.

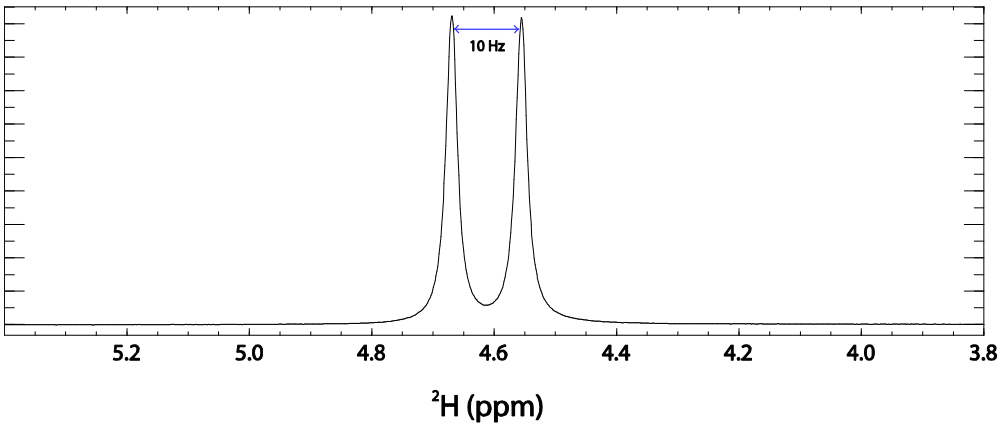


Figure 6.5- The quadrupolar deuterium splitting of HDO in 5 mg/mL pf1 phage at 310 K.

A 280 µM, 600 µL ^{15}N-labelled latherin sample was divided into two 300 µL samples. One was diluted up to 570 µL in 20 mM NaPi, 50 mM NaCl, 1 mM NaN_3, pH 7.5, the other up to 510 µL in the same buffer. 30 µL D_2O was added to each. 60 µL of pf1 phage was then added to the second sample to produce a final phage concentration of 5 mg/mL.

Spectra from which the ^{15}N scalar (J) and scalar + residual dipolar (J + D) couplings were measured was recorded from the unaligned and aligned samples respectively. Unaligned

and aligned experiments were carried out in duplicate to allow estimation of the measurement error.

Spectra were collected using the in-phase and anti-phase (IPAP) method and then combined by addition or subtraction to produce spectra containing only the upfield or downfield peak in the ^{15}N dimension [210]. The J and J+D values were then measured using the RDC calculator macro (Dr. Graham Ball, University of Edinburgh) within *Analysis* for adequately resolved peaks to calculate D for latherin backbone amide groups.

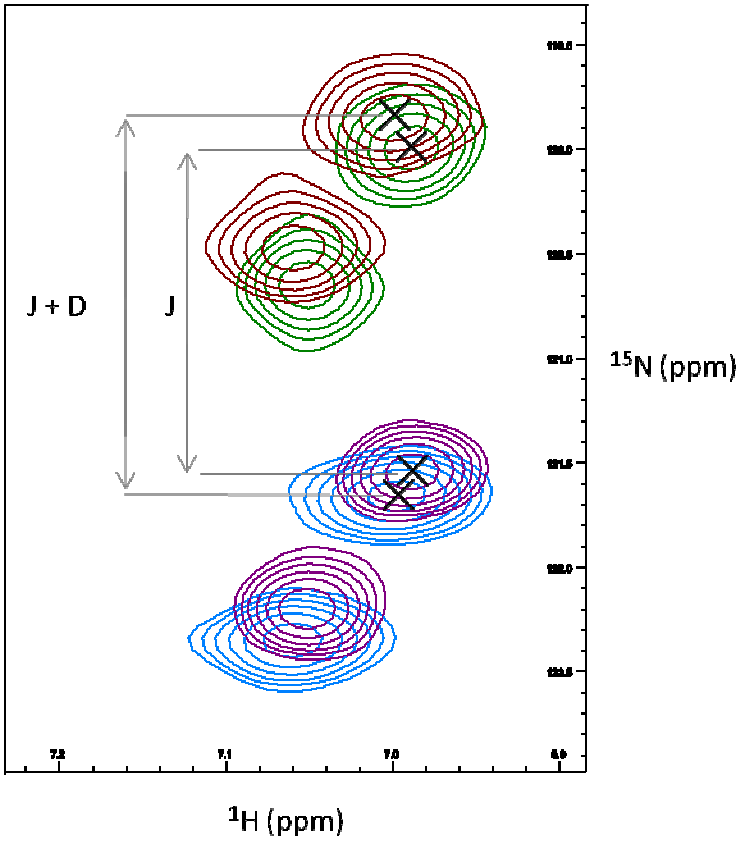


Figure 6.6- The scalar and RDC ^{15}N splitting of a backbone amide peak. The scalar splitting (J) was measured from a isotropically tumbling sample. The increased splitting (J+D) from an aligned sample was used to measure the RDC (D)

In order to make use of the calculated RDCs in protein structure determination, the magnitude of the axial, A_a, and rhombic, R, components of the alignment tensor must be extracted. These values were estimated using the method described by Clore *et al*. [211]. Extreme D_{AB} values correspond to orientations of vectors closest to the z ($\theta = 0°$) (D_{zz}) and y ($\theta = 90°$, $\phi = 90°$) (D_{yy}) axes of the alignment tensor. The highest probability dipolar coupling value, coincides with the magnitude of a bond vector aligned along the x ($\theta = 90°$, $\phi = 0°$) axis (D_{xx}). By plotting a histogram of the observed RDC frequencies these values

can be easily extracted and used to calculate the magnitude of the axial and rhombic components (Figure 6.7).

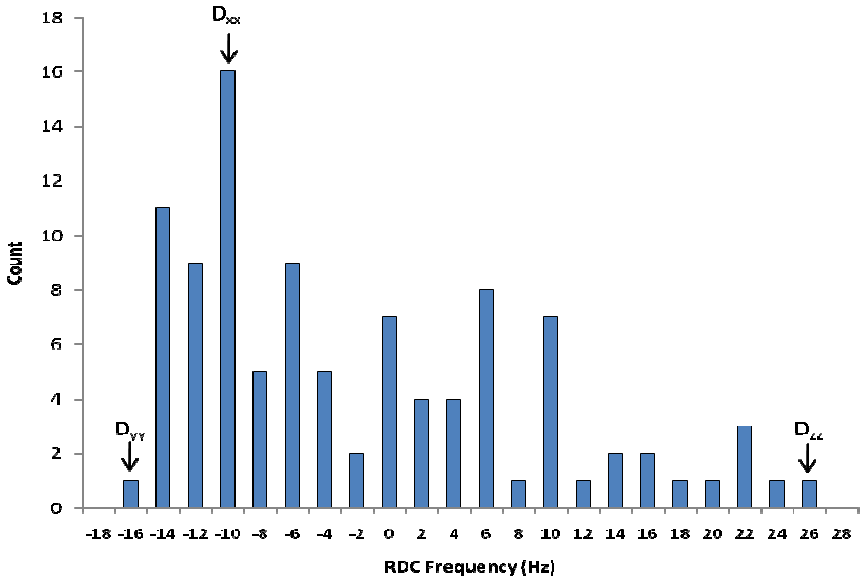


Figure 6.7- Histogram of RDC frequencies in latherin.

This technique is most accurate when a range of vector types (eg. C_α-H_α or N-C') are used as a more uniform dispersion of dipolar couplings is observed, however for an initial estimation, one internuclear vector type is sufficient [211]. The alignment tensor used in this work was calculated to have an A_a = 12.5 Hz and R = 0.1. 88 D_{NH} restraints with an error of ± 0.93 Hz were incorporated into the structure calculations via the SANI potential [212] in square-well mode.

6.5.4 Hydrogen Bonds

The backbone amide protons involved in hydrogen bonds were identified by their rate of exchange with the solvent when the sample was dissolved in D_2O. Solvent accessible hydrogen atoms are rapidly exchanged with deuterium from the solvent, whereas the exchange of hydrogen atoms that are involved in hydrogen bonding or are buried in the hydrophobic core of the protein is slow. The different magnetic properties of hydrogen and deuterium nuclei ensure that deuterium nuclei are not detected in experiments such as ^{15}N-HSQC. Therefore, the rate of exchange for backbone amide protons can be measured by a series of ^{15}N-HSQC experiments that monitor the exponential decay of the amide signal. This was carried out by dissolving 60 µL of 3mM ^{15}N labelled latherin in 540 µL of D_2O and recording ^{15}N-HSQC experiments at 20 minute intervals for the first 3 hours then at 1 hour intervals for a further 5 hours.

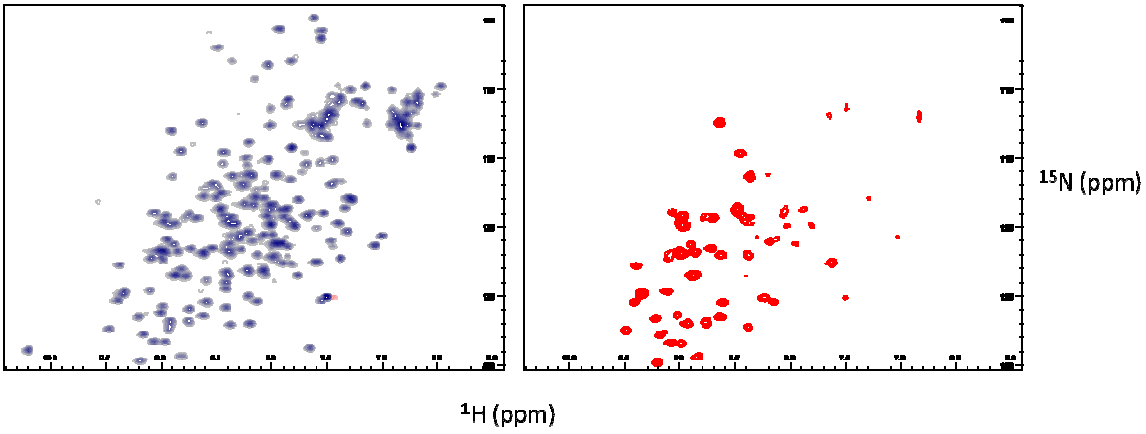


Figure 6.8- Latherin H/D exchange spectra. A ^{15}N HSQC of latherin in H_2O (blue) and after 20 minutes dissolution in D_2O (red) allowed for the identification of slowly exchanging amide protons involved in hydrogen bonding.

The majority of peaks reduced to 10% of the original intensity within the 20 minutes taken to record the first HSQC (Figure 6.8). Of the 42 peaks that persisted, 34 were sufficiently unambiguous for confident assignment and creation of hydrogen bond restraints. At this stage H/D exchange data was used as a qualitative approach to identify amide hydrogen's involved in hydrogen bonding and therefore the actual rates of exchange were not required. This data is however utilised in Chapter 7 to provide insight into the dynamics of the molecule. Hydrogen bond restraints were introduced in the later stages of structure calculation once bond acceptors could be confidently identified by inspection of the NOE-refined structures, supported by NOE data. The restraints were implemented as two distance restraints that enforce an appropriate bond angle between the donor and acceptor groups. One restraint between the proton on the donor amide group and the oxygen of the acceptor carbonyl group with lower and upper limits of 1.7 and 2.7 Å. And a second between the donor amide nitrogen and the oxygen of the acceptor carbonyl group with lower and upper limits of 2.5 and 3.9 Å.

Donor Amide	Acceptor Carbonyl	Donor Amide	Acceptor Carbonyl
15-Leu	11-Ile	107-Tyr	141-Glu
16-Thr	12-Thr	108-Val	92-Leu
31-Ile	27-Leu	109-Arg	139-Ala
37-Leu	33-Leu	110-Thr	90-Leu
39-Thr	35-Gly	111-Asp	136-Leu
40-Ile	36-Leu	112-Ile	88-Ile
41-Leu	37-Leu	113-Arg	134-Ser
62-Glu	97-Lys	114-Val	86-Ile
64-Lys	95-Tyr	115-Gln	132-His
69-Arg	91-Glu	116-Leu	84-Ile
71-Leu	89-Pro	118-Leu	82-Lys
74-Ser	87-Trp	119-Glu	127-Arg
76-Glu	85-Asp	127-Arg	119-Glu
84-Ile	116-Leu	129-Ala	117-Arg
85-Asp	76-Glu	134-Ser	113-Arg
87-Trp	74-Ser	136-Leu	111-Asp
88-Ile	112-Ile	141-Glu	107-Tyr
91-Glu	69-Arg	163-Leu	159-Ile
92-Leu	108-Val	167-Ile	163-Leu
94-Val	106-Leu	169-Glu	165-Asn
95-Tyr	64-Lys	171-Leu	167-Ile
96-Leu	104-Leu	179-Leu	175-Leu
97-Lys	62-Glu	183-Leu	179-Leu
104-Leu	96-Leu	195-Ile	191-Val
106-Leu	94-Val	200-Asp	196-Asn

Table 6.3- Identity of H-bond restraints. The donor amide groups were identified by analysis of the H/D exchange spectra. Acceptor carbonyl groups were identified by analysis of well defined structures and NOE spectra.

6.5.5 Disulphide Bond

Previously acquired biophysical indicates that the two cysteines, Cys133 and Cys175, in latherin form an intramolecular disulphide bond [70]. The observed chemical shifts of the cysteines, of which the C_β's in particular are sensitive to the oxidation state of the neighbouring S_γ, indicated that both cysteines were in the oxidised form [213]. However in order to prevent the biased linking of what could potentially be two intermolecular disulphides, a disulphide bond restraint was only included in structure calculations once the cysteines were found in close juxtaposition within well defined structures. This restraint was enforced as an unambiguous distance restraint between the two Sγ atoms with a distance of 2.02 ± 0.1 Å with the cysteines patched to remove the Hγ.

6.6 Structure Calculation by ARIA

The program ARIA 2.3 (Ambiguous Restraints for Iterative Assignment) was employed for structure calculation [214]. ARIA is a well established, automated protocol that can handle a large number of restraints.

6.6.1 Restrained Molecular Dynamics and Simulated Annealing

NMR restraints from experimental data alone do not contain sufficient information to completely describe protein structures. However, by supplementing this data with empirical restraints that enforce proper covalent structure of the protein, bond lengths, bond angles and other elements of standard covalent geometry, high quality structures can be produced. Therefore, the aim of structure calculations is to find conformations that satisfy the experimental restraints within the ranges of allowed covalent bonding geometries.

ARIA uses restrained molecular dynamics (RMD) and simulated annealing to produce multiple structures compatible with restraints and empirical information. ARIA uses CNS (Crystallography and NMR system) [215] as the structure calculation engine utilising PARALLHDG-5.3 force field with PROLSQ non-bonded energy terms [216] as well as torsion angle dynamics. The conditions of the simulated annealing procedure were adapted to improve convergence of calculated structures principally by increasing the number of steps at each stage to improve convergence (Table 6.4).

Step	Temperature (K)	Steps
High Temp	10000	10000
Refine	2000	4000
Cool 1	1000	20000
Cool 2	50	16000

Table 6.4- Summary of simulated annealing conditions.

6.6.2 Ambiguous Restraints

The main class of experimental restraints used in NMR structure determination are distance restraints created from NOESY spectra. For these distance restraints to be used effectively the identity of the nuclei involved must be known. However due to the high number of hydrogen nuclei within a protein, many of which are found at the same or very similar chemical shifts, it is often impossible to assign NOE crosspeaks in an unambiguous manner. In latherin this ambiguity was exacerbated due to the high population of leucine residues. As discussed previously, latherin NOESY crosspeaks were assigned manually in instances where this could be carried out with confidence, however in the vast majority of cases, where assignments were truly ambiguous or sufficient doubt existed, peaks were left unassigned to prevent biasing of the structure through mis-assignment.

ARIA is able to carry out structure determination utilising ambiguous distance restraints by combining all possible assignments for a specific NOE cross-peak within one restraint [214]. A list of possible assignments is achieved by setting chemical shift tolerance ranges, with any resonances that fall within then being considered possible assignments within for the crosspeak under consideration. If the tolerances are set too tight, the correct assignment may be omitted resulting in incorrect assignment. Alternatively, too wide a tolerance range will lead to excessive ambiguity of the assignment, making calculated structures less likely to converge. In this work the tolerances were selected based upon inspection of the NOE crosspeak line width. Values for the proton, carbon and nitrogen dimensions were set to 0.03 ppm, 0.25 ppm and 0.4 ppm respectively.

6.6.3 Prochiral Swapping

Prochiral groups in the protein structure, in particular the methyl groups of leucine and valine residues, can be difficult to assign stereospecifically and any misassignment can lead to errors in the structures calculated. ARIA is able to deal with these ambiguous prochiral groups by allowing non-degenerate chemical shifts to be assigned to either member of the prochiral pair and swapping assignments during the calculation [217]. The low energy conformation is then selected for each prochiral centre.

6.6.4 Iterative Structure Calculation Scheme

ARIA uses an iterative strategy for structure calculation [217]. In iteration 0, an initial simulated annealing calculation is carried out using all unambiguous (manually assigned) and ambiguous restraints starting from a random structure. Iteration 1 begins by analysis of a predefined number of structures (in this case 7) with the lowest total energy calculated in iteration 0. Distance restraints are analysed to identify violations, likely due to incorrect assignments and noise peaks by selecting restraints that fall outwith the set violation tolerance. The number of assignment possibilities in an ambiguous restraint is reduced by eliminating assignments that only contribute below a certain fraction of the total peak intensity by decreasing the partial assignment filter. This process is repeated for a defined number of iterations with increasingly strict limits of partial assignment filter threshold and violation tolerance.

It	Distance Restraints	No of Structures	Selection Criteria	Selected Structures	Partial Assignment Filter	Violation Tolerance (Å)	Other Restraints
0	All	20	Total Energy	7	1.0	1000	Di, H, S
1	Checked from it0	20	Total Energy	7	0.9999	5.0	Di, H, S
2	Checked from it1	20	Total Energy	7	0.999	3.0	Di, H, S, RDC
3	Checked from it2	20	Total Energy	7	0.99	1.0	Di, H, S, RDC
4	Checked from it3	20	Ex Res Energy	7	0.98	1.0	Di, H, S, RDC
5	Checked from it4	20	Ex Res Energy	7	0.96	1.0	Di, H, S, RDC
6	Checked from it5	20	Ex Res Energy	7	0.95	0.1	Di, H, S, RDC
7	Checked from it6	50	Ex Res Energy	10	0.95	0.1	H, S, RDC
8	Checked from it7	100	Ex Res Energy	20	0.95	0.1	H, S, RDC

Table 6.5- The iterative strategy used for Latherin structure calculation in ARIA. This table shows the strategy of later structure calculations once disulphide and hydrogen bond restraints had been included. Ex Res = experimental restraints, RDC= RDC restraints, Di = dihedral restraints, H= H-bond restraints and S= disulphide restraints.

6.6.5 Restraint Analysis

After each round of structure calculations the restraints rejected by ARIA as violations were analysed within *Analysis* software. Any obvious sources of error were removed: noise peaks were deleted and mis-assignments were corrected. In some cases violated restraints identified previously unassigned resonances. Violated restraints for which no valid reason for exclusion could be identified were retained as restraints in subsequent structure calculations. This process was repeated until calculated structures displayed good convergence and low total and experimental energies.

6.6.6 Water Refinement

For efficiency structures are calculated *in vacuo* with simplified van der Waals interactions and no electrostatic term in the forcefield. This approach can lead to artefacts in the final structure such as nonoptimal packing and unsatisfied hydrogen bond donors or acceptors [218]. A further refinement step was therefore carried out where a defined number of structures from the final iteration were encapsulated in a thin layer of explicit solvent (water) molecules and refined against an empirical forcefield including fuller van der Waals representation and electrostatic terms. To ensure that the water refinement process does not cause deviation from the experimental data, the force field used is parameterised to complement those used *in vacuo* structure calculations. Water refinement was carried out for the 20 lowest total energy latherin structures from iteration 8 in accordance with the published procedure [218].

6.7 Structure Validation

The calculated structures provide models of the protein generated to best represent the experimental data. The quality of these models depends upon the quality of the data and how the data is interpreted into a three dimensional structure [219]. It is therefore important to have some measure of the quality of the calculated structure, i.e. how representative it is likely to be of the molecule of interest.

6.7.1 Selection of the Ensemble

The nature of NMR structure determination means that restraints define a range of allowed distances, angles etc. rather than distinct values. Therefore multiple structures will be representative of the experimental data. NMR structures are described as ensembles, a group of several structures that all fulfil the experimental restraints approximately equally well.

There is no general consensus on the selection of models to be included in the ensemble. One approach is to include all models with energies below an arbitrary cut-off point. Another approach is to distinguish between properly folded and misfolded models, manifested in a sharp increase in energy [219]. In the final *in vacuo* iteration of latherin calculations no clear increase could be observed until after approximately 60 models (Figure 6.9). This was deemed too large a sample for selection of a manageable ensemble, so the 20 models that displayed the lowest experimental energy (NOE + RDC) were selected. The ensemble of structures were then subjected to water refinement and these 20 water refined models used as the latherin ensemble.

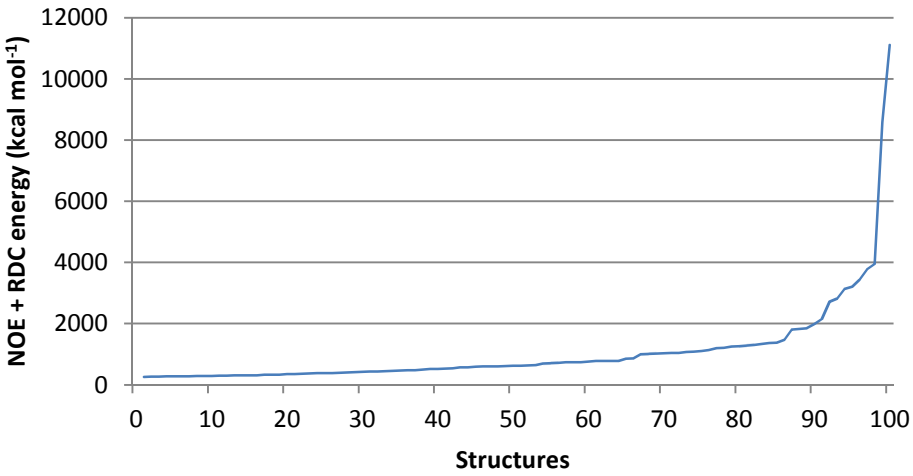


Figure 6.9- Graph showing the 100 latherin structures calculated in iteration 8 ranked by the sum of the NOE + RDC energies.

6.7.2 Fit of Latherin Ensemble Structures to Experimental Restraints

One method of assessing the quality of the calculated structures is to analyse the total number and quality of the restraints used in the structure calculation since the more restraints that are included the better the definition of the calculated structure [220]. The statistics of the experimental restraints, calculated from the 20 water refined structures in the ensemble are summarised in Table 6.6. No consistent violations over 0.3 Å were observed.

NOE distance restraints	
NOE restraints	6922
Ambiguous	2210
Unambiguous	4293
Intra-residue	1985
Inter-residue	2308
Sequential (i-j=1)	985
Medium-range (i-j<5)	518
Long-range (i-j>5)	805
Violations per structure > 0.5 Å	1.15
Violations per structure > 0.3 Å	8.60
Distance Restraint RMSD	0.036 Å
Other restraints	
RDCs	88
RDC Q factor	0.127
Hydrogen bonds	34
Dihedral angle restraints	369
Disulphide bond	1

Table 6.6- Statistics of the Experimental Restraints. Average statistics were calculated from the 20 water refined structures in the ensemble. The number of violations is shown as the average and standard deviation per structure.

The precision of the calculated structures can be used to gain an idea of how well defined the calculated structures are from the experimental restraints. It should be noted that precision is different from accuracy and that a more precise structure is not necessarily a more representative structure.

An estimation of the precision of the ensemble can be achieved by inspecting the variation between models visually. The 20 latherin models are very similar particularly in the structured regions, due to the high number of restraints within these areas (Figure 6.10). The loop and termini regions are less well defined, due to the low number of restraints in these areas. This in turn may be a consequence of dynamic processes in these areas.

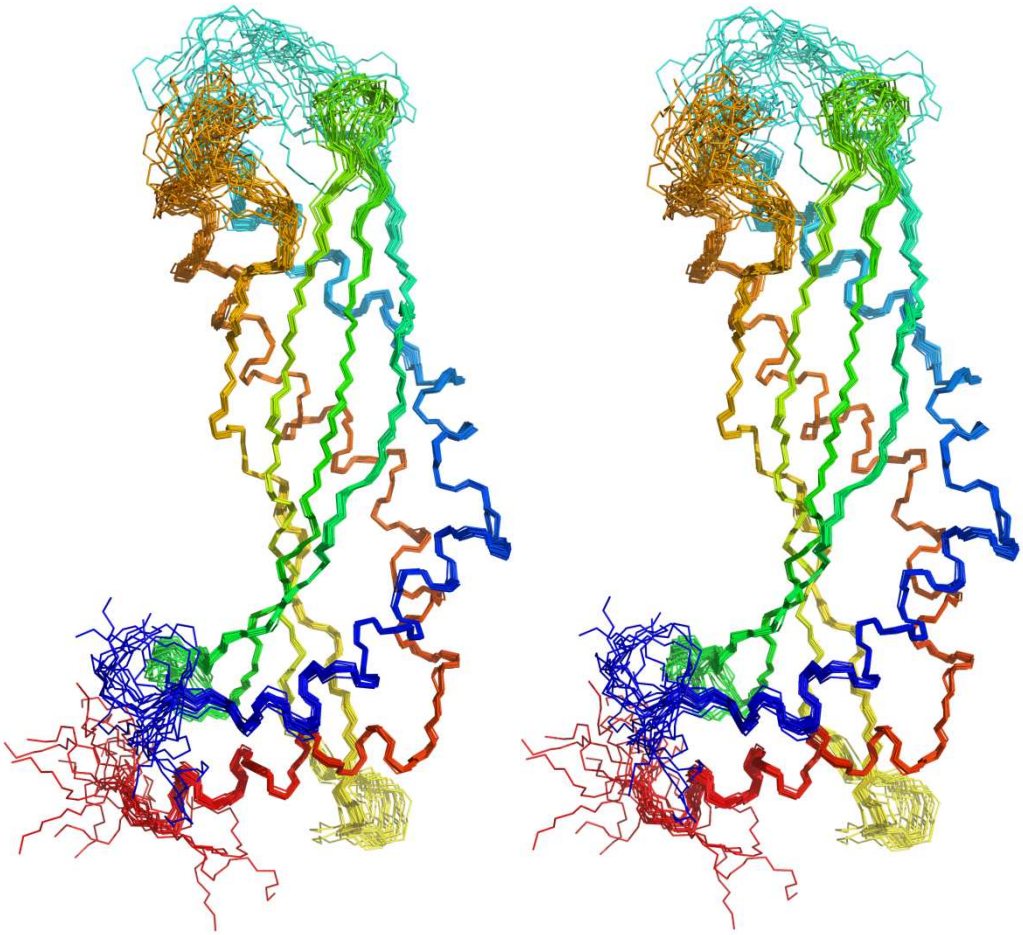


Figure 6.10- Stereo-view of the twenty latherin models in the ensemble superimposed upon one another. The backbone nitrogen and carbon atoms are shown and the molecule is coloured from blue (N-terminus) to red (C-terminus). Image was created using PyMOL software [47].

For a more quantitative comparison of the latherin models within the ensemble, the root mean square deviation (RMSD) of the atomic coordinates of each model can be calculated relative to an unbiased mean structure generated from the ensemble. The programme UWMN (M. Hartshorn and L. Caves, University of York) was utilised to create an average distance matrix between atoms from all models within the ensemble. This distance matrix is then projected back into three dimensional space to produce an unbiased mean structure.

The RMSDs for all heavy atoms, backbone atoms and all C_α atoms were calculated for all residues and for only the residues defined as being within the structured regions of the protein. The structured regions were determined based upon the relaxation data presented in Chapter 7. Series of sequential or nearby residues with S^2 values less than 0.75 were selected and analysed and the following areas deemed as sufficiently dynamic to be described as unstructured: 0-7; 31-33; 53-61; 70-73; 121-126; 143-157 and 202-208.

	RMSD to the unbiased mean (Å)	
	All	Structured
All heavy atoms	1.65	0.93
Backbone heavy atoms	1.40	0.66
C_α's	1.44	0.68

Table 6.7- RMSD's calculated using UWMN software for structured and unstructured regions. The structured regions were selected as residues 8-30; 34-52; 62-69; 74-120; 127-142 and 158-201.

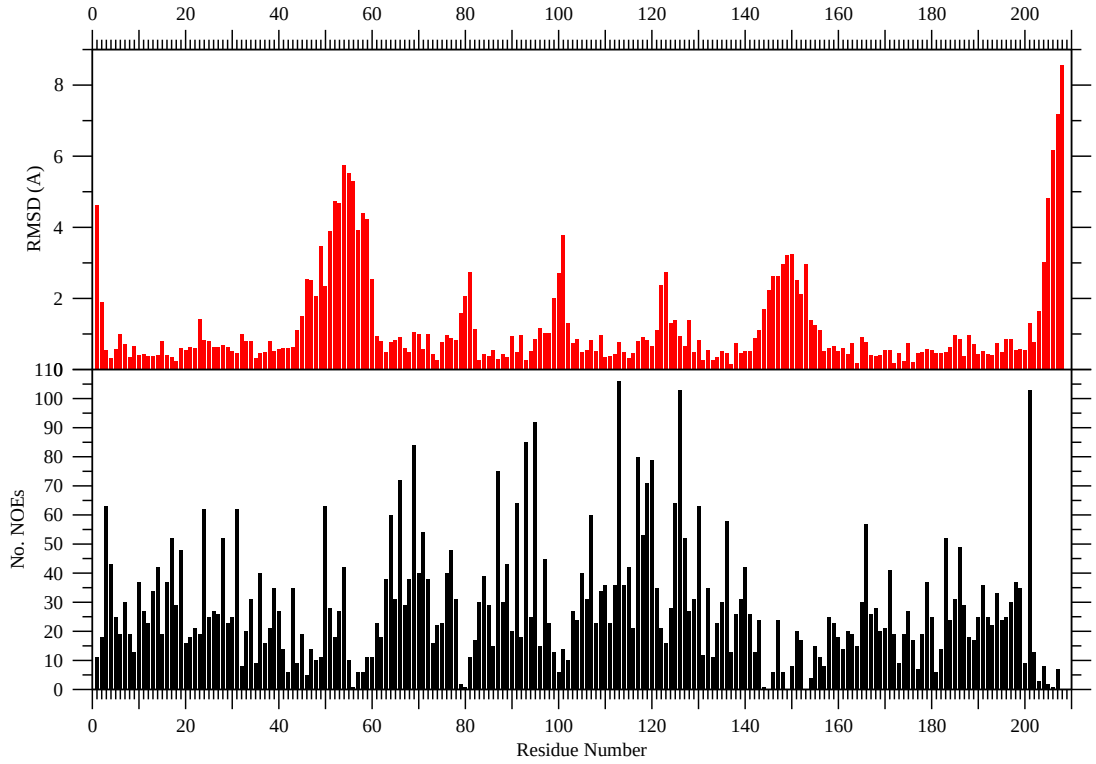


Figure 6.11 The correlation between distance restraints and RMSD in the latherin structure. The total number of NOE restraints used in the final *in vacuo* structure calculation and the RMSD of the 100 calculated structures are plotted per residue.

6.7.3 Geometric Quality of the Latherin Ensemble Structures

As discussed, structures are calculated based upon idealised values with regards to the covalent structure of the protein. Although these values are entirely dependent upon how the empirical values are weighted in the forcefield, deviations from the idealised value can provide an indication of disagreement between experimental restraints and empirical values.

RMSD from the ideal geometry	
Bond length (Å)	0.01 ± 0.00
Bond angle (°)	0.61 ± 0.01
Improper angle (°)	1.56 ± 0.07

Table 6.8- Summary of RMSD of the latherin ensemble from empirical ideal values.

In high resolution X-ray crystallography, the structure can be determined exclusively from experimental data and unlike NMR, does not rely on forcefields or refinement procedures. The stereochemical properties of residues in high quality crystal structures therefore present an accurate definition of available or likely geometries found within proteins. By comparing the results of protein structures calculated by NMR to standard stereochemical properties derived from high resolution crystal structures, abnormalities or errors in the geometry of the calculated structures can be assessed. For the latherin ensemble this was carried out using the PROCHECK_NMR program [221]. Results are shown in Appendix B. The Ramachandran statistics for all residues (excluding glycine and proline) are shown in Table 6.9.

Procheck Group	% residues
Residues in most favoured regions	82.3
Residues in additional allowed regions	15.4
Residues in generously allowed regions	1.7
Residues in disallowed regions	0.6

Table 6.9- A summary of the Ramachandran statistics for the ensemble of latherin structures as determined by PROCHECK_NMR.

More than 97% of the residues fall into the favoured and additionally allowed regions. This is consistent with high quality structures with good covalent geometry. Through analysis of the residue by residue Ramachandran plots (Appendix B), the residues located in the α-helices and β-strands lie within, or close to, the favoured regions of the Ramachandran plot. The residues that are found in the termini or loops show much greater variability in φ and Ψ angles, likely due to a lack of restraints to define these regions.

Another criterion that can be assessed to determine the geometric quality of the side chain torsion angle is the χ_1 value. The PROCHECK-NMR results (Appendix B) indicate well-defined χ_1 values for the majority of residues within the latherin structures. No residues are assessed to fall outside favourable χ_1 regions.

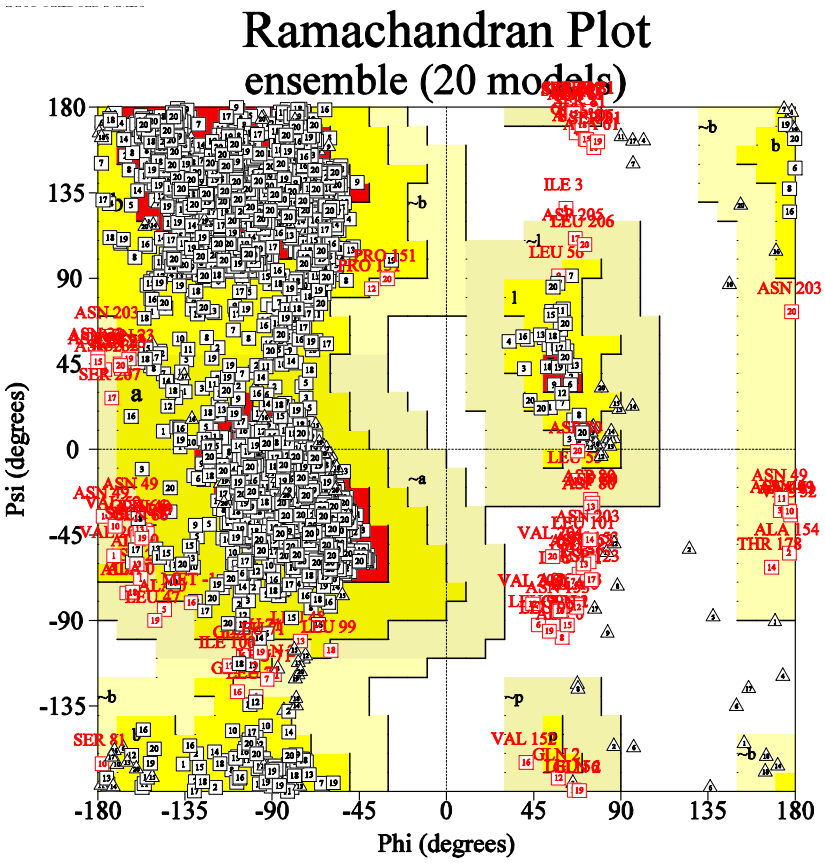


Figure 6.12- Ramachandran plot of the latherin ensemble. White boxes indicate residues in favoured regions. Triangles represent glycine residues. Red squares indicate residues in unfavourable regions. Diagram created by PROCHECK_NMR.

6.8 Latherin Structure

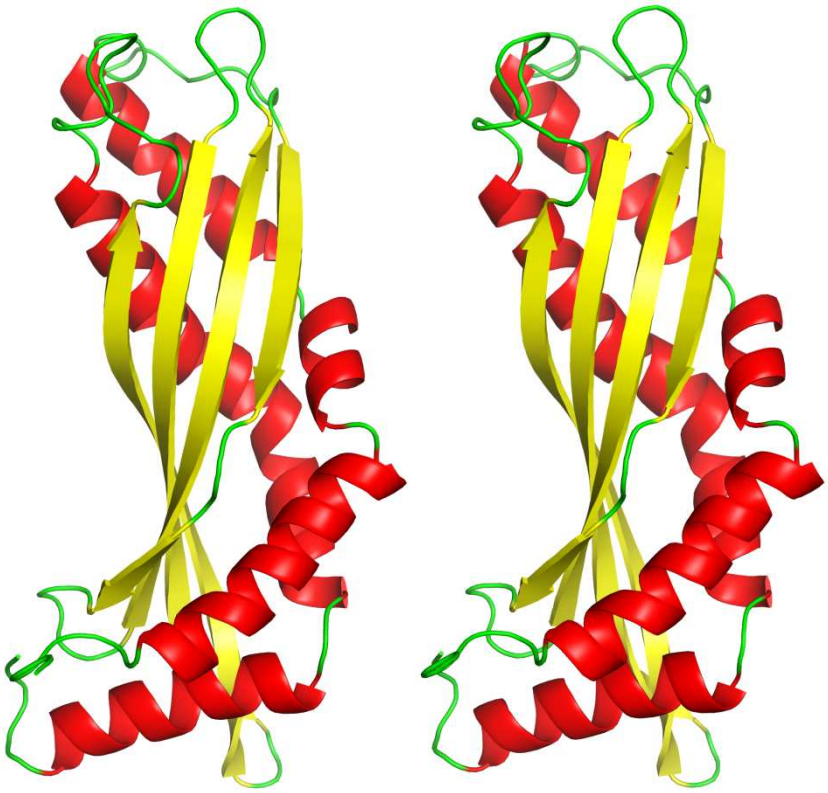


Figure 6.13- Stereo image of calculated latherin structure. Coloured by secondary structure: α-helix = red, β-strand = yellow and other = green. Image was created from the latherin model closest to the average of the ensemble, using PyMOL software [47].

Latherin has an elongated structure composed of a central four stranded, anti-parallel β-sheet, flanked by two long sections of kinked α-helix. The molecule forms a near-cylindrical shape with a length of 65 Å and 25 Å diameter. One end of the molecule is comprised of the two termini and two inter-strand loops, one unstructured and the other in a tight β-hairpin. The other end contains one short inter-strand loop and two long poorly defined loops that connect the β-strands to the α-helices.

The distribution of regular secondary structure in latherin was assessed using the PROCHECK-NMR predictions, based upon the method described by Kabsch and Sander [222], supported by analysis of the patterns of characteristic H-bond contacts in the calculated structures and is summarised in Figure 6.14. The PROCHECK-NMR results are shown in Appendix B.

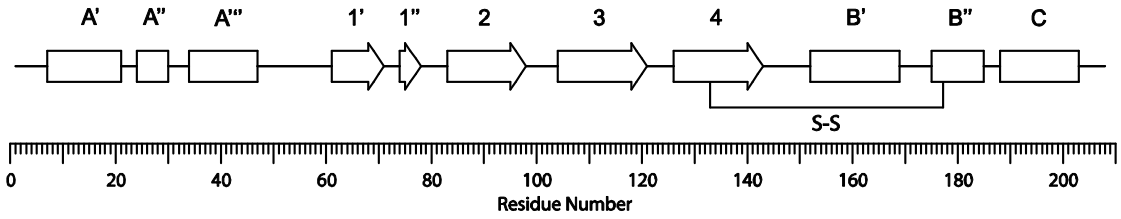


Figure 6.14- Diagram showing distribution of regular secondary structure elements of latherin. α-helices are represented by boxes, β-sheet by arrows. The location of the intramolecular disulphide bond is shown, labelled S-S.

The N-terminal helix, labelled α_A, stretches from residues 7-47 with two breaks in the regular secondary structure at residues 22-23 and 31-33. Helix α_A can therefore be subdivided into three sections: α_A' (7-21), α_A'' (24-30), and α_A''' (34-47). The C-terminal helix (152-203) is also interrupted by two breaks at positions 170-174 and 186-187. The length of helix from residues 152-185 has therefore been defined as one helix, labelled α_B with a break and therefore consisting of two sections: α_B' (152-169) and α_B'' (175-185). The final section (188-203) was determined to be a separate helix, α_C due to its orientation relative to the preceding helix. The four β-strands were labelled from 1-4: β_1 (61-77), β_2 (83-97), β_3 (104-120), and β_4 (126-142). β_1 is interrupted by a 3 residue break and therefore can be divided into two sections β_1' (61-70) and β_1'' (74-77). The disulphide bond (Cys133-Cys176) connects β_4 to α_B''. A topology model summarising the nomenclature of the regular secondary structure in latherin is shown in Figure 6.15.

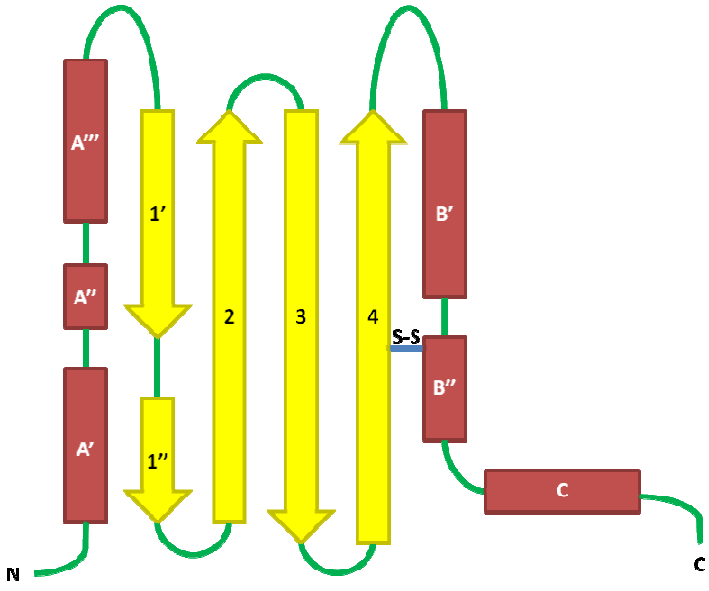


Figure 6.15- Topology model of latherin.

Analysis of the structure is continued with reference to the dynamic data (described in Chapter 7) in Chapter 8.

7 Latherin Backbone Dynamics

Proteins are highly dynamic molecules that undergo a range of internal motions, often integral to biological function. Furthermore, the importance of structural motion in the surfactant protein Rsn-2 has been proposed and discussed elsewhere in this work and the presence of similar dynamic processes within latherin must be considered. Protein motions occur on a range of timescales, from picosecond scale local motions up to seconds for concerted motions of large domains. NMR measurements are sensitive to motion on many of these timescales and NMR studies present a powerful technique to characterise functionally relevant dynamics within protein structures.

7.1 Relaxation Mechanisms

In NMR, the equilibrium state is defined by a complete lack of coherence, a slight excess of population in the low energy state and a slight population deficiency in the high energy state as defined by the Boltzmann distribution. Anything that perturbs this equilibrium will be followed by a process of relaxation back to the equilibrium state.

The process whereby magnetisation returns to these Boltzmann populations is called longitudinal relaxation. The energy evolved due to the return of the magnetisation to the low energy, equilibrium state, is dissipated into the nuclei's surrounding environment and is therefore also referred to as spin-lattice relaxation. Longitudinal relaxation is characterised by the time T_1 and its associated rate R_1 ($R_1 = 1/T_1$). Relaxation can also occur through the loss of coherence in the x y plane. This process is described as transverse relaxation. Transverse relaxation is also known as spin-spin relaxation as it occurs through interactions with other spins and is characterised by the relaxation time T_2 and rate R_2. Another mechanism arising due to relaxation is the NOE. The NOE relaxation occurs between two dipolar coupled nuclei. For a pair of dipolar coupled nuclei there are six relaxation pathways. Four are single spin flip processes corresponding to longitudinal relaxation processes discussed already. The other two are cross relaxation processes whereby both nuclei relax together: zero-quantum and double quantum transfer. It is these cross relaxation processes that give rise to the NOE.

The only mechanism whereby a nucleus interacts with its environment is through its magnetic interactions. Therefore, in order for a nucleus to undergo a transition between states, it must experience a magnetic field oscillating at an appropriate frequency. These

fields can be induced by interactions that modify the effects of the external magnetic field, primarily dipole-dipole interactions, but can also involve other interactions such as chemical shift anisotropy. As the molecule tumbles, the magnetic field experienced by each nucleus will vary depending upon the orientation of these interactions to the external magnetic field. The result is a localised magnetic field oscillating at a frequency modulated by the thermal motions of the nuclei involved. These motions mostly involve the molecular correlation time, τ_m, but may also include any internal motions that the nuclei involved undergo.

NOTE- the terms molecular rotational correlation time, τ_m and rotational correlation time, τ_c, are often used interchangeably, which is valid for isotropically tumbling molecules. However in this work, τ_m will be used to describe the rotational correlation time of entire molecules whereas τ_c will be used for individual internuclear vectors.

Relaxation is most efficient when the frequencies of the oscillations match the energy differences between the spin states. For longitudinal relaxation this corresponds to a frequency close to the Larmor frequencies of the nuclei involved. For the NOE, the energy differences between the double quantum and zero quantum transfers correspond to the sum and the difference, respectively between the two Larmor frequencies. Transverse relaxation, like longitudinal, is affected by the frequencies near the Larmor frequencies, however, is also dependent upon a second interaction: the secular part of transverse relaxation. In a rapidly tumbling molecule the field inhomogeneity caused by dipolar interactions is averaged away to zero. However if the rate of tumbling is reduced the averaging of the dipolar coupling is also reduced giving rise to high variations in the local field dependent upon the orientation of the individual dipolar couplings. The result is a loss of coherence with a maximal effect when the molecule is static. The effects of T_2 relaxation therefore increase as τ_c decreases (Figure 7.1).

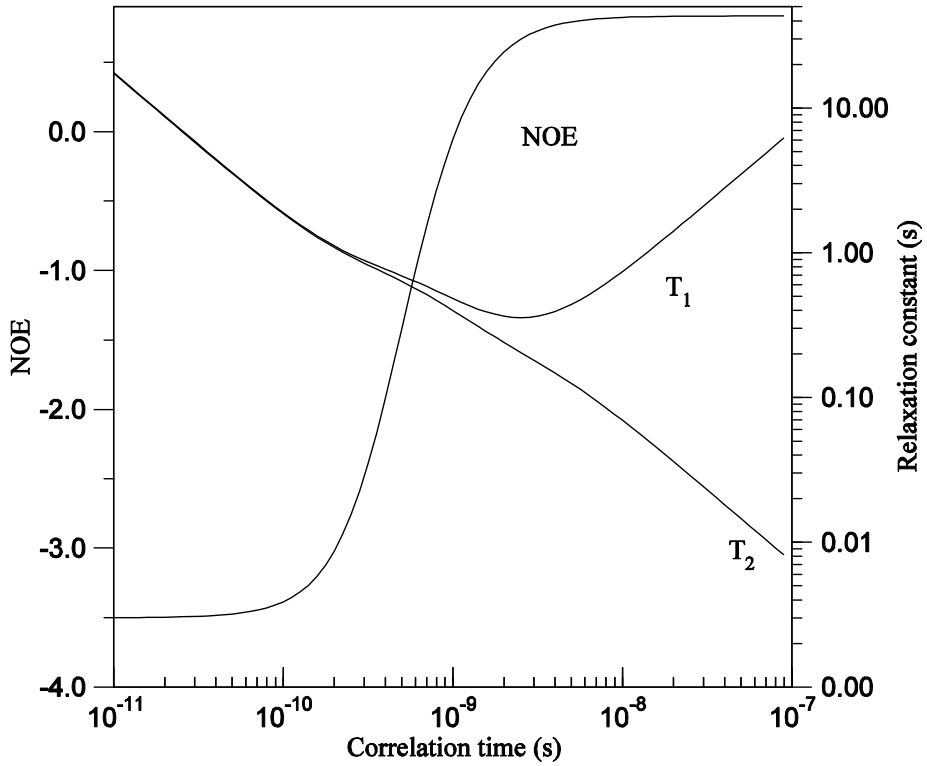


Figure 7.1- The relationship between ^{15}N T1, T2, NOE and correlation time at 14.1 T. Data supplied by Dr Brian Smith (University of Glasgow).

This relationship between correlation time and relaxation can be utilised to probe the dynamics of the molecules involved. Nuclei undergoing internal motions will have decreased effective correlation times. Deviations in the relaxation properties from the expected values can be analysed to extract information with regards to the form of any internal motions.

7.2 Dynamics from Relaxation

In this work only the relaxation properties of backbone amides were analysed due to the simplicity of the calculation and the ability to extract information concerning the dynamics of almost every residue in a polypeptide backbone.

The probability of the occurrence of an component oscillating at a particular frequency (ω) within a dipole tumbling at a specific rate, as defined by τ_c, can be defined by the spectral density function J(ω):

$$J(\omega) = \frac{2\tau_c}{1 + \omega^2\tau_c^2} \tag{7.1}$$

Relaxation of backbone amides can be defined using these spectral density functions by the following parameters [223]:

$$R_{1=}(d^2/4)[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + c^2J(\omega_N) \tag{7.2}$$

$$R_2 = (d^2/8)[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N) + (c^2/6)[4J(0) + 3J(\omega_N)] + R_{ex} \tag{7.3}$$

$$NOE = 1 + (d^2/4\,R_1)(\gamma_N/\gamma_H)[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N) \tag{7.4}$$

where $d = \mu_o h\gamma_N\gamma_H\langle r_{NH}^{-3}\rangle/(8\pi^2)$, $c = \omega_N\Delta\sigma/\sqrt{3}$, μ_0 is the permeability of free space, h is Planck's constant, ω_H and ω_N are the Larmor frequencies of ^{1}H and ^{15}N spins, respectively, $\gamma_H$and γ_N are the gyromagnetic ratios of the ^{1}H and ^{15}N spins; and $\Delta\sigma$ is the chemical shift anisotropy of the ^{15}N spin collinear with the N-H bond vector. The R_{ex} term is included to account for chemical exchange processes that contribute to the decay of transverse magnetisation.

The model-free formalisation, known as such because it makes minimum prior assumptions with regards to the motion of the molecule, described by Lipari and Szabo and extended by Clore *et al*., can be used to analyse the data to determine the amplitudes and timescales of the intramolecular motions [224,225,226]. The spectral density function is adapted to include a number of parameters to account for any deviation from the expected correlation time. In the simplest case, an isotopically tumbling molecule, the spectral density function can be described by:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(S_f^2 - S^2)\tau}{1+(\omega\tau)^2}\right] \quad (7.5)$$

In which $\tau = \tau_e\tau_m/(\tau_e + \tau_m)$, τ_m is the isotropic rotational correlation time of the molecule, τ_e is the effective correlation of the internal motions, $S^2 = S_f^2 S_s^2$ is the square of the generalized order parameter characterising the amplitude of the internal motions, defined by values between 1.0 (absolute restriction) and 0.0 (free rotation). S_f^2 and S_s^2 are the squares of the order parameters for the internal motions on the fast and slow timescales, respectively.

If the molecule experiences rotational anisotropy, the correlation times for individual ^{15}N-^{1}H vectors will differ depending upon their position relative to the molecular diffusion tensor. Therefore experimental knowledge of the proportions of the diffusion tensor and each vector's relationship to it is required for detailed analysis of intramolecular motions in non-spherical proteins. An alternative to equation 7.5 that accounts for these parameters for the analysis of molecules with axially symmetric diffusion tensors of proportions $D_{\parallel}$ x $D_{\perp}$ has been proposed [227,228,229]:

$$J(\omega) = \frac{2}{5}S_f^2\sum_{j=1}^{3} A_j\left[\frac{S_s^2\tau_j}{1+(\omega\tau_j)^2} + \frac{(1-S_s^2)\tau'_j}{1+(\omega\tau'_j)^2}\right] \quad (7.6)$$

where $\tau'_j = \tau_j\tau_e/(\tau_j + \tau_e)$, $\tau_1^{-1} = 6\,D_{\perp}$, $\tau_2^{-1} = 5\,D_{\perp} + D_{\parallel}$, $\tau_3^{-1} = 2\,D_{\perp} + 4\,D_{\parallel}$, $A_1 = (3\cos^2\theta - 1)^2/4$, $A_2 = 3\sin^2\theta\cos^2\theta$, $A_3 = (3/4)\sin^4\theta$, and θ is the angle between the N-H bond vector and the unique axis of the principle frame of the diffusion tensor.

7.3 Measurement of T_1, T_2 relaxation rates and heteronuclear NOE

7.3.1 T_1

In the experiment magnetisation is generated and detected on amide protons for maximum sensitivity. Magnetisation is transferred between proton and ^{15}N nuclei using refocused INEPT and reverse INEPT sequences. Transverse magnetisation produced by the initial INEPT sequence is converted to longitudinal magnetisation via 90° pulse and is then allowed to relax during a delay. The magnetisation is then returned to the transverse plane via a second 90° pulse for ^{15}N chemical shift labelling before the reverse INEPT transfer. The peak height in the resultant spectrum is proportional to the sampled z magnetisation immediately prior to the 90° pulse. By repeating the experiment with different time delays between the two 90° pulses, the rate of the return of the z magnetisation to equilibrium can be observed and the T_1 calculated by fitting an exponential function to the data curve.

Delay times used were 1.2 s, 1.6 s, 2.1 s, and 2.6 s with experiments with delays of 1.2 s and 2.1 s repeated to aid error estimation.

7.3.2 T_2

The T_2 value was also measured utilising a pseudo 3D ^{15}N HSQC format experiment with selected time points as the third dimension [230]. In order to measure T_2 the contribution from sample inhomogeneity and chemical shift evolution must be removed. This is achieved through a series of CPMG pulses resulting in signal decay proportional to the pure rate of T_2 relaxation. Again the initial step of this experiment is a refocused INEPT sequence to produce in-phase amide ^{15}N magnetisation in the transverse plane. The delay with CPMG pulsing is then followed by ^{15}N chemical shift labelling before the reverse INEPT converts magnetisation back to the protons for acquisition.

Delay times used were 16.96 ms, 33.92 ms, 67.84 ms, 101.76 ms, and 135.68 ms with experiments with delays of 16.96 ms and 67.84 ms repeated to aid error estimation.

7.3.3 NOE

The ^{1}H-^{15}N steady state NOE values were obtained by recording ^{15}N HSQC format spectra with magnetisation starting on the ^{15}N with (NOE_{sat}) and without (NOE_{unsat}) ^{1}H saturation applied before the start of the experiment. The differences in intensity of the resultant

peaks allow an estimation of the contribution from NOE relaxation. Both experiments were repeated to allow for error estimation.

7.3.4 Data Analysis

Each of the data sets was processed using the *Azara* suite of programs (W Boucher, www.ccpn.ac.uk/azara). Initial processing scripts were created from the Bruker data files via the program *ReferenceB* written by Dr. Krystyna Bromek. The processed spectra were loaded into CCPNMR *Analysis v.2* [179]. The T_1 and T_2 data was each described as an experimental series with corresponding relaxation delays. Assignment of the spectra was carried out by copying peak assignments from existing ^{15}N-HSQC experiments for every backbone amide peak that were deemed sufficiently resolved to produce reliable data. For the T_1 and T_2 experiments, a relaxation time and associated error were calculated for each peak by applying an exponential fitting function ($I = A.\,\mathrm{e}^{-\mathrm{Bt}}$) to peak intensity by height (I) plotted against delay time (t). This was carried out using the *follow intensity changes* popup within the *Analysis* programme [179]. The error of the fit was calculated using the covariance error method [231]. The peak heights from the NOE_{sat} and NOE_{unsat} spectra were exported from *Analysis* and a measurement error was estimated from the signal to noise in duplicate experiments using the *snratio_nmr* script (A. G. Palmer III, www.palmer.hs.columbia.edu/software.html). The peak heights and associated errors were then used to calculate NOE ratios using the *noecalc_nmr* script (A. G. Palmer III, www.palmer.hs.columbia.edu/software.html).

7.3.5 Results

The calculated T_1, T_2 and NOE values were plotted as a function of amino acid sequence (Figure 7.2).

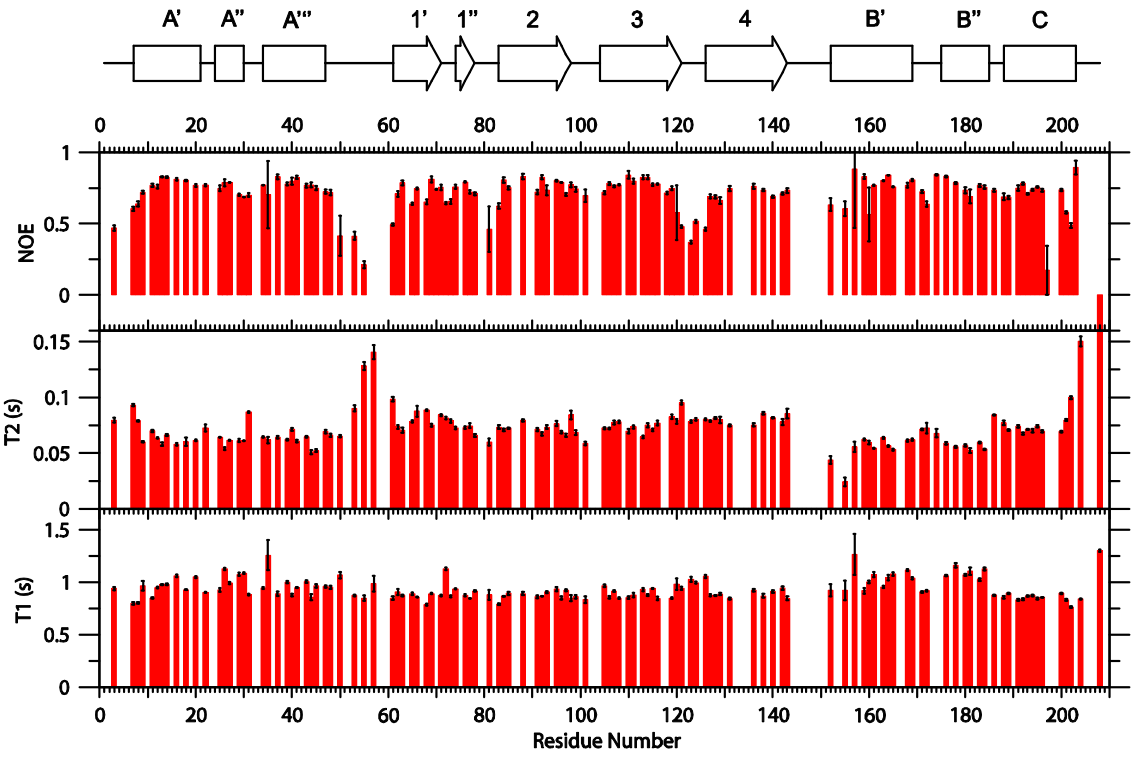


Figure 7.2- ^{15}N T_1, T_2 and heteronuclear NOE values for each resolved backbone amide in latherin at 60.8MHz (^{15}N) and 310K. The T_2 and NOE values of the C-terminal residue 208Val have been truncated for clarity. These values are T_2 = 0.68 s and NOE = -1.39.

There are trends in the data relative to various structured regions of latherin. The four β-strand regions and helix C all show decreased T_1 and elevated T_2 values relative to the remaining α-helical sections (Table 7.1).

Secondary Structure	Avg. T_1 (ms)	Avg. T_2 (ms)
Helix A	965 ± 101	65 ± 9
Strand 1	891 ± 79	80 ± 8
Strand 2	878 ± 41	72 ± 4
Strand 3	895 ± 50	74 ± 5
Strand 4	909 ± 63	80 ± 3
Helix B	1038 ± 97	57 ± 10
Helix C	850 ± 36	75 ± 9

Table 7.1- Average T_1 and T_2 times for residues in the regular secondary structure features of latherin.

It would appear that this is a consequence of anisotropic molecular tumbling. N-H bond vectors aligned with the long axis of the diffusion tensor will experience longer rotational correlation times, and thus longer T_1s and shorter T_2s, than those orientated perpendicular to it. These results are consistent with the orientation of the secondary structure features relative to the long axis of the molecule in the calculated structures.

7.4 Estimation of Correlation Times and the Rotational Diffusion Tensor

In order to determine the contribution to relaxation from internal motions, an assessment of the correlation times of the individual backbone amide vectors as well as that of the entire molecule must be determined. The equation introduced by Kay *et al*. identifies the relationship between the ratio of T_1/T_2 and the rotational correlation time τ_c for each residue [232]. Therefore τ_c can be estimated on a residue by residue basis from the T_1/T_2 ratio. A plot of T_1 vs. T_2 for each residue allows for an estimation of the rotational correlation time of the entire molecule (Figure 7.3).

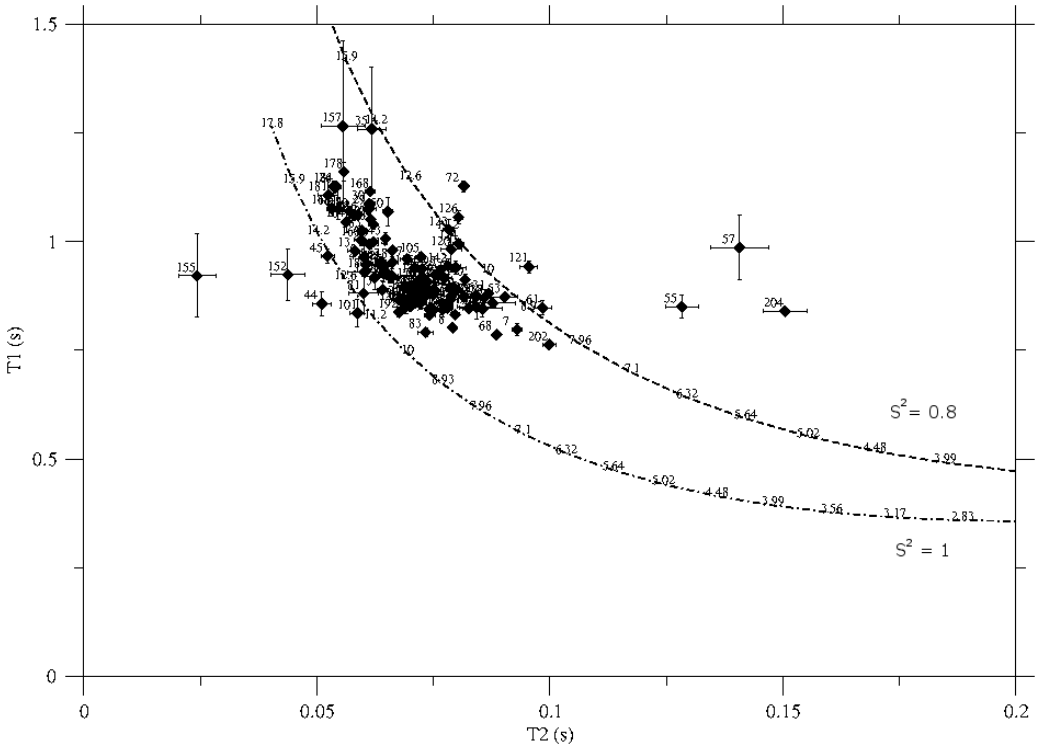


Figure 7.3- ^{15}N T_1 against T_2 for latherin at 60.8MHz (^{15}N) and 310K. All resolved residues are represented by a data point apart from 208, which had a T1= 1.30 s and a T2= 0.68 s. The dashed curve nearest to the origin represents a simulated S^2 of 1.0 and the other dashed curve a simulated S^2 of 0.8. The numbers along these curves indicate the correlation time for isotropic tumbling corresponding to expected T_1/T_2 values.

The majority of residues are spread along a curve parallel to the S^2 contours in the correlation time range from 10 to 14.2 ns. This dispersion pattern of the points is indicative of a molecule exhibiting anisotropic rotational diffusion as would be expected from the calculated structure. The majority of the points are clustered between the two lines and correspond to residues in relatively well-defined regions in the 3D structure. The points lying to the right of the S^2= 0.8 line are generally contributed by residues found in loops or at the C-terminus of the protein. Theoretically, no residue should have an S^2 greater than 1.0, however residues found to the left of S^2= 1.0 boundary can be rationalised by the presence of chemical exchange, R_{ex}, which would decrease T_2 and produce points shifted to the left of the graph. The residues found at this position primarily consist of those found on the large loop between residues 145 and 155. The majority of the amides of residues on this loop remained unassigned, perhaps as a consequence of this conformational exchange on the spectra.

Kay's equation can be used to produce an estimation of τ_c for each residue [232]. This was carried out using the *r2r1_tm* script (A. G. Palmer III, www.palmer.hs.columbia.edu/software.html), which calculates τ_c from the ratio of R_2/R_1. In order to gain an accurate estimate of τ_c any residues with a NOE value < 0.6 and any remaining residues with an R_2/R_1 ratio more than one standard deviation from the mean were removed from the calculation. The τ_c values calculated for the remaining residues are shown in Figure 7.4.

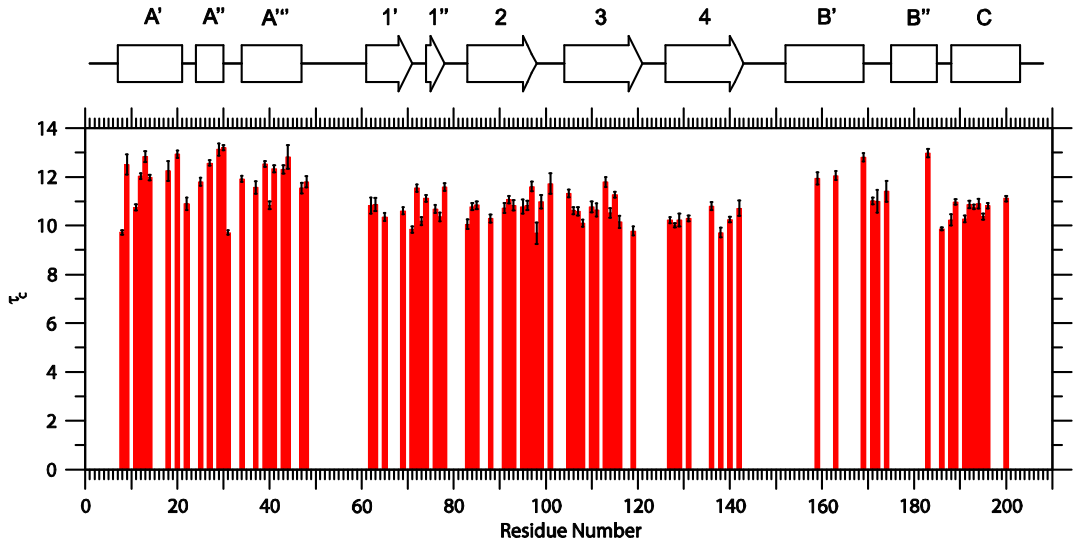


Figure 7.4- τ_c estimated from R_1 and R_2. Residues with NOE less than 0.6 or with an R_2/R_1 ratio more than one standard deviation from the mean were omitted.

A τ_m of 11.1 ± 0.9 ns of the entire molecule was then estimated based upon the average τ_c of the individual amide residues.

Analysis of the relaxation data together with the elongated nature of the calculated solution structure of latherin suggested that the molecule would either undergo axially symmetric or rhombic rotational diffusion rather than isotropic. The likelihoods of each of these models and the parameters of the subsequent rotational diffusion tensor of latherin were determined using the quadric representation approach proposed by Brushweiler and Lee [233,234]. The trimmed τ_c values calculated above and the calculated latherin structure that was closest to the mean of the ensemble, were used as input to the *quadric_diffusion* program (A. G. Palmer III, www.palmer.hs.columbia.edu/software.html). The program indicated that an axially symmetric diffusion tensor with a $D_{\parallel}/D_{\perp}$ ratio of 1.665 was the most appropriate fit for the calculated latherin structure.

7.5 Model Free Analysis

The internal motion of the latherin backbone was modelled using the FASTModelFree program (P. Loria, Yale University, http://xbeams.chem.yale.edu/~loria/software.php) [235] which interfaces with the ModelFree program (A. G. Palmer III, http://www.palmer.hs.columbia.edu/software/modelfree.html). The ModelFree program optimises “Lipari-Szabo model free” parameters to heteronuclear relaxation data. The FASTModelFree program fully automates the process of model selection and tensor optimisation. The three relaxation parameters; R_1, R_2 and heteronuclear NOE, their associated errors, the estimated τ_m and the proportions of the calculated diffusion tensor were modelled to five possible models with no more than three model-free parameters. These models were: Model 1 (S^2), 2 (S^2, τ_e), 3 (S^2, R_{ex}), 4 (S^2, τ_e, R_{ex}) and 5 (S^2, τ_e, S^2_f). The selection criteria for the models was performed based upon Mandel’s protocol [236] with small modifications suggested by Chen [237]. Once residues have been assigned to a model, the motional parameters and the rotational diffusion tensor are further optimised. The process is then repeated with the optimised diffusion tensor in order to improve the fit of the parameters. The final optimisation was carried out with a τ_m of 11.3 ns and a rotational diffusion tensor with a $D_{\parallel}/D_{\perp}$ ratio of 1.68. A summary of the assignment of the 124 input residues to the five models, as well as the residues that could not be fitted to any of the models and residues that were not analysed due to lack of data or spectral overlap is shown in Table 7.2.

Model	#	Residues
1 (S^2)	20	13, 14, 35, 37, 50, 69, 81, 84, 88, 93, 98, 101, 110, 111, 113, 114, 120, 136, 159, 174,
2 (S^2, τ_e)	34	11, 18, 22, 25, 30, 39, 62, 63, 65, 74, 76, 77, 78, 83, 85, 91, 95, 106, 107, 115, 116, 123, 126, 129, 140, 142, 163, 171, 172, 183, 188, 191, 195, 200
3 (S^2, R_{ex})	6	26, 41, 92, 157, 160, 176
4 (S^2, τ_e, R_{ex})	29	12, 20, 29, 34, 43, 44, 45, 48, 57, 72, 96, 97, 99, 105, 118, 152, 155, 161, 165, 168, 180, 181, 184, 189, 192, 193, 194, 196, 204
5 (S^2, τ_e, S^2_f)	27	3, 7, 8, 9, 27, 31, 47, 53, 55, 61, 66, 68, 70, 71, 73, 108, 119, 121, 124, 127, 128, 131, 138, 143, 186, 201, 202
Not fitted	8	16, 40, 164, 169, 178, 197, 203, 208
Not assessed	84	1, 2, 4, 5, 6, 10, 15, 17, 19, 21, 23, 24, 28, 32, 33, 36, 38, 42, 46, 49, 51, 52, 54, 56, 58, 59, 60, 64, 67, 75, 79, 80, 82, 86, 87, 89, 90, 94, 100, 102, 103, 104, 109, 112, 117, 122, 125, 130, 132, 133, 134, 135, 137, 139, 141, 144, 145, 146, 147, 148, 149, 150 ,151, 153, 154, 156, 158, 162, 166, 167, 170, 173, 175, 177, 179, 182, 185, 187, 190, 198, 199, 205, 206, 207

Table 7.2- Model Free assignment of the latherin backbone amide resonances and the identity of the resonances not included in the analysis. The 84 unanalysed residues consist of the 11 prolines, the 8 amide resonances that could not be assigned and a further 65 residues where the amide resonance was not deemed sufficiently resolved to produce reliable data.

The optimised model free parameters of S^2, τ_e, R_{ex} and S^2_f for the backbone amide vector of each residue are listed in Appendix C together with the sum square error (SSE). The magnitude of the various model free parameters is shown for each fitted residue in Figure 7.5.

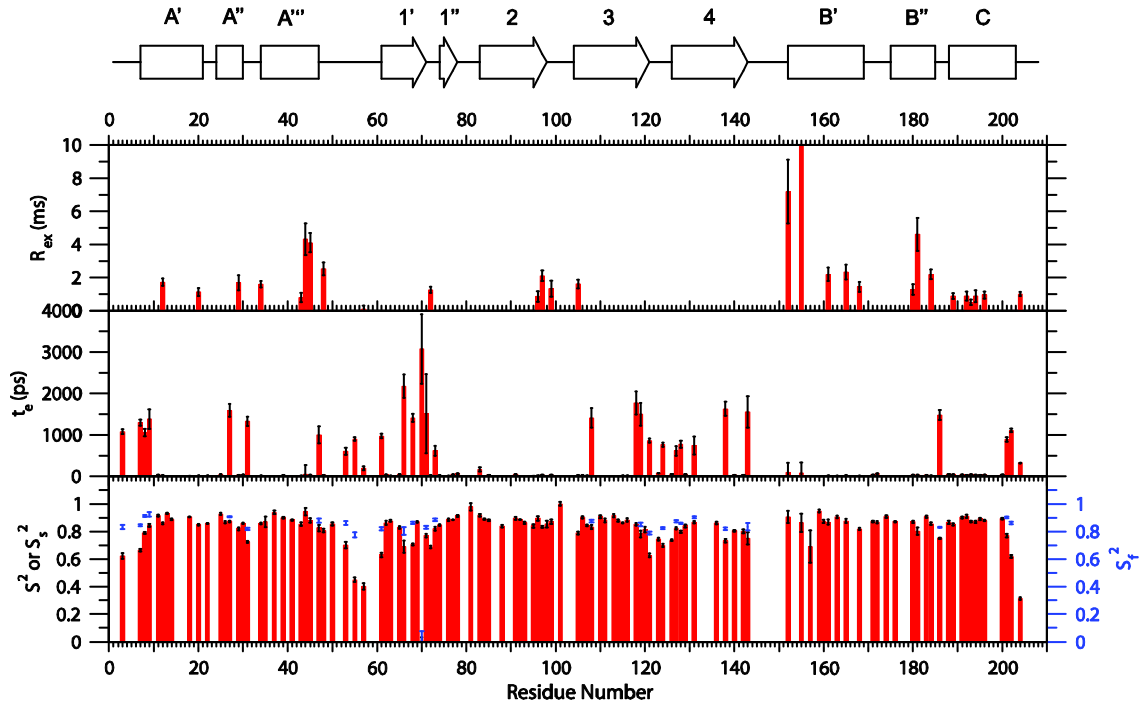


Figure 7.5- Model free parameters plotted as a function of residue number.

The results indicate that the two termini are dynamic. At the N-terminus residue 3Ile, as well as the first two residues of helix A', 7-Val and 8-Ser, show low values of S^2 and internal motion on the nanosecond timescale. At the unstructured C-terminus (204-208) the only residue assessed 204-Val has a very low S^2 value of 0.31. This disorder appears to extend into the C-terminus of helix C, where both 201-Arg and 202-Ala have low S^2s of 0.77 and 0.62, respectively.

The residues within the structured regions of helix α_A all show S^2 values greater than 0.75 indicative of restricted motion. There does appear to be some degree of chemical exchange occurring in residues 44-Val, 45-Thr and 48-Leu at the C-terminus of α_A'''. Of the two breaks in helix α_A, residue 22-Gly is structured as defined by a S^2 of 0.86, whereas 31-Ile shows a high rate of internal motion with S^2 0.73 and τ_e 1.33 ns. Helix α_B has S^2 values all greater than 0.75, indicating limited internal motion. However, the R_{ex} parameters of the first few residues at the N-terminal end of the helix α_B' indicates chemical exchange. The five residue break between sections α_B' and α_B'' appears to be highly structured as demonstrated by an average S^2 of 0.88, whereas, the break between α_B'' and α_C appears to be more dynamic with 186-Leu showing nanosecond timescale internal motion. Helix α_C itself is limited in its motion according to the S^2 parameter, apart from the final few.

The four β-stands show low levels of internal motion. There are however a few exceptions. The residues preceding (66-Gln, 68-Thr and 70-Leu) and within (71-Leu 72-Gln and 73-Leu) the break in the first strand, all have low S^2 values, that can accounted for by nanosecond timescale internal motion. Another region of the β-stands with elevated mobility is the residues preceding and following the loop between β_3 and β_4 (121-125). Residues 115-119 in β_3 and various residues located in β_4 experience internal motion on the nanosecond timescale.

Of the five loops that connect the regular secondary structures only loop 2 (78-82) represented by residues 78-Ser and 81-Ser, shows little or no evidence of internal motion. The first loop between α_A and β_1 (48-60) is highly dynamic with depressed S^2. Loop 3, between β_2 and β_3 (98-103), although having high values of S^2 also requires R_{ex} parameters indicating chemical exchange. Loop 4, between strands β_3 and β_4 (121-125) like the region surrounding it discussed above, is undergoing internal motion. There is very little data for the long loop between β_4 and $\alpha_{B'}$ (143-151) as a large proportion of this area was not assigned. Consequently, only one residue, 143-Gln at the β_3 end of the loop was analysed. 143-Gln appears to be quite dynamic with a τ_e of 1.56 ns. As discussed already the area of α_B' immediately following this loop shows high levels of chemical exchange. This evidence supports the hypothesis that the reason this loop was not observable in assignment spectra was as a consequence of broadened peaks due to chemical exchange.

7.6 Hydrogen-Deuterium Exchange

The hydrogen-deuterium exchange experiment carried out in the structure determination process can also provide insight into the dynamics of regions of regular secondary structure. Hydrogen bonded regions that are restricted in their motion will be less likely to exchange hydrogen with the solvent, whereas those that experience internal motion or chemical exchange will exchange with the solvent rapidly.

As discussed, after dilution of a ^{15}N latherin sample in 90% (v/v) D_2O the vast majority of peaks decayed to 10% of their original intensity within the time required to run the first experiment. These residues were described as under fast exchange. The 42 peaks that persisted after 20 minutes dissolution in D_2O were monitored during the remainder of the experiment. Those which displayed no change in intensity over the duration of the experiment were described as under slow exchange and those that decayed within the

lifetime of the experiment as under medium rates of exchange. The identity of the residues within the medium and slow rates of exchange groups are shown in Table 7.3.

Group	Residues
Medium exchange	16, 40, 41, 62, 69, 71, 74, 76, 84, 91, 93, 97, 107, 109, 111, 113, 114, 115, 119, 127, 129, 136, 163, 169, 179, (15/167), (104/187), (134/135)
Slow exchange	85, 88, 92, 95, 96, 106, 108, 110, 116, (21/86/90/117), (33/87), (64/112), (94/141), (166/182)

Table 7.3- Groupings of rates of H/D exchange for residues in latherin. All other residues were determined to be undergoing fast exchange. Brackets indicate ambiguous peaks with possible assignment options.

The rates of exchange for amides within the medium rate of exchange group were calculated by applying an exponential fitting function ($I = A.\mathrm{e}^{-\mathrm{Bt}} + \mathrm{c}$) to peak intensity by height (I) plotted against time after dilution in D_2O (t). This was carried out using the *follow intensity changes* popup within the *Analysis* programme [179]. The error of the fit was calculated using the covariance error method [231]. A plot of the exchange rates as well as an overview of the distribution of the residues in each group against the secondary structure is shown in Figure 7.6.

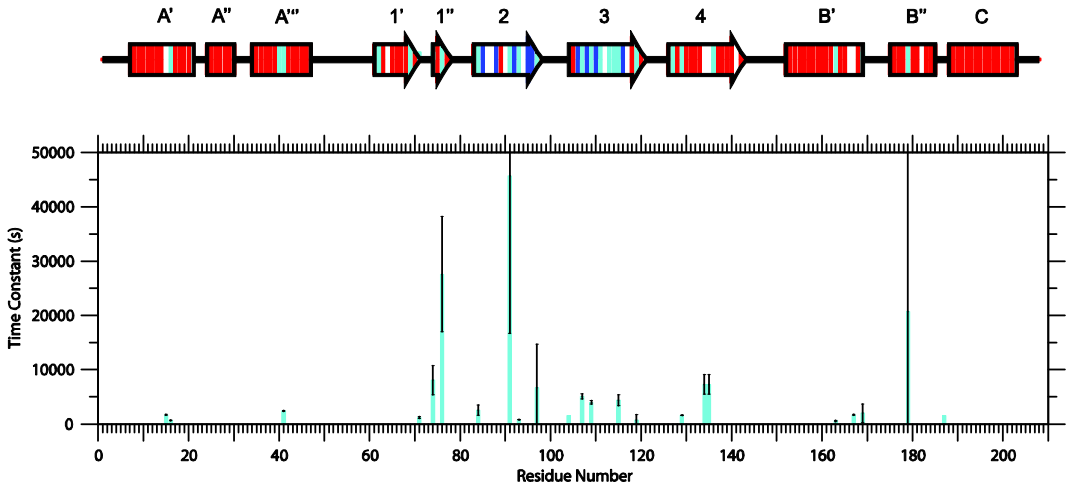


Figure 7.6- Latherin hydrogen exchange rates plotted by residue. Residues undergoing fast exchange (< 20 minutes) are coloured red, medium exchange (20 – 480 minutes) cyan and slow exchange (> 480 minutes) blue in the secondary structure cartoon. The time constants for residues undergoing medium rates of exchange are shown in the graph.

The results indicate that all of the slow and the majority of the medium rate exchanging residues are located in the two central strands β_2 and β_3. This agrees with the model free analysis where the majority of residues within these strands are fitted to model 1. The two

outer strands have some residues undergoing medium rate exchange. The remainder of the protein appears to be undergoing fast exchange with a few isolated medium exchanging peaks found in the middle of α-helical sections.

7.7 Conclusions

The measured relaxation parameters are indicative of a molecule undergoing anisotropic motion. The individual parameters of each of the sections of regular secondary structure fit well into their orientation relative to the overall rotational diffusion tensor of the protein. This data correlates very well to the calculated structures validating the determined structure. An optimised τ_m of 11.30 ns and a rotational diffusion tensor with a $D_{\parallel}/D_{\perp}$ ratio of 1.683 were calculated for the latherin molecule.

Analysis of the internal motions of the molecule identified that the majority of the regular secondary structure elements of the molecule are relatively rigid. Regions where dynamics were observed include the two termini and the loop between helix α_A‘ and strand β_1. Chemical exchange occurs at the ends of each of the long sections of α-helix furthest from the termini, as well as the loop between strands β_2 and β_3. The loop between residues β_4 and α_B’, which is conspicuous by the absence of amide resonances in the assignment spectra, immediately precedes one of these regions supporting the possibility that the loop is also undergoing chemical exchange.

The H/D-exchange results agree well with the model free analysis, further demonstrating the stability of the two central strands of β-sheet.

In short, latherin is a molecule with a rigid core and two dynamic ends; in particular the end containing the three loops which shows high levels of chemical exchange.

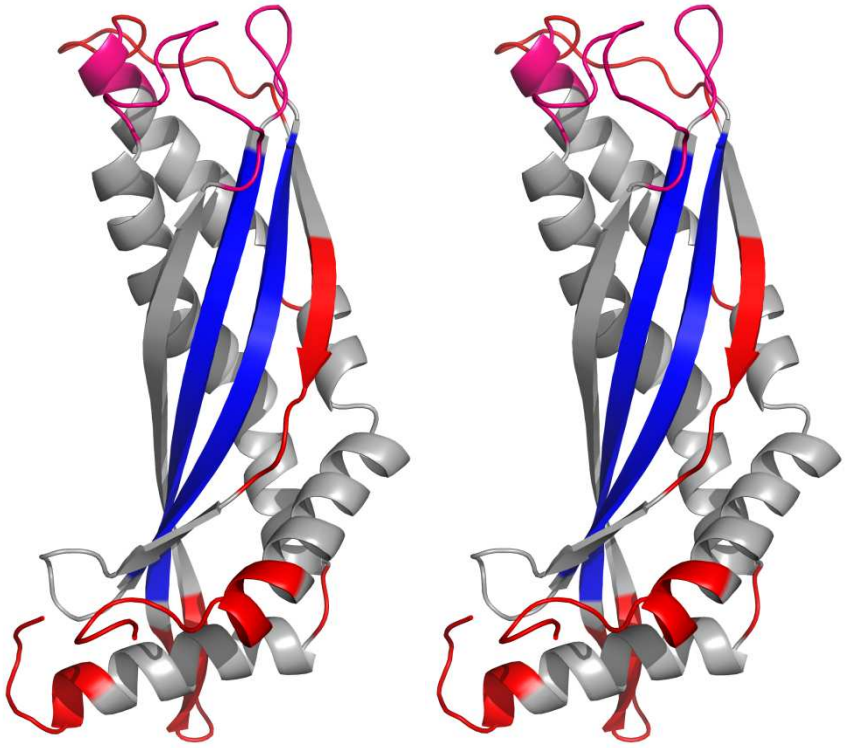


Figure 7.7- Model summarising areas of high and low dynamics in the latherin structure. Areas of high internal motion are coloured red, areas of chemical exchange are coloured pink and areas that show low levels of dynamics are coloured blue. Image created using PyMol [47].

8 Analysis of the Latherin Structure

The high resolution solution structure of latherin presents an opportunity to investigate and rationalise specific properties of the molecule. In this chapter the structure will be probed to gain insight into its character, its implications and how it functions as a surfactant.

8.1 General

8.1.1 Molecular Shape

Latherin has an elongated shape with a long axis of 65 Å and 25 Å diameter. This equates to a molecule with an axial ratio of approximately 2.6:1. Comparison to the results of analytical size-exclusion chromatography carried out in Chapter 3, indicate this value to be in good agreement with the estimations results based upon the Stokes radius calculated by Beeley *et al*. [70], where a monomer with an axial ratio of 3:1 was predicted.

That latherin is monomeric in solution is further reinforced by the solution structure and more directly by the relaxation data which indicated a rotational correlation time, τ_m, of 11.30 ns. This value, although slower than the estimated correlation time for a perfect sphere of latherin's mass at 310K: 7.31 ns, is sufficiently fast that the possibility of a dimer, which would have a theoretical minimum rotational correlation time of 13.34 ns, is unlikely. The disparity between the measured and calculated rotational correlation times can be rationalised by the molecule's elongated shape. This conclusion is reinforced by estimation of the rotational correlation time carried out based upon the calculated structure utilising the programme HYDROPRO [238] which indicates a rotational correlation time of 12.69 ns at 310K.

8.1.2 Secondary Structure

The distribution of secondary structure in the solution structure shows greater proportions of regular secondary structure than that predicted from CD spectra carried out in previous studies [73], but agrees relatively well with predictions based upon sequence data alone.

Secondary Structure	CD	Sequence	NMR
α-helix	30 %	34 %	39 %
β-sheet	18 %	30 %	25 %
Other	52 %	36 %	36 %

Table 8.1- Comparison of secondary structure predictions to NMR structure. The distribution of secondary structure in the NMR structure was calculated based upon the assignment of secondary structure set out in Chapter 6. Predictions from CD spectra were taken from previous work [73]. Predictions based upon amino acid sequence were carried out using the PSIPRED program [59].

In this work a co-purified latherin aggregate was eliminated from sample preparations. If this misfolded form was present in the sample previously analysed by CD spectroscopy it would likely have caused an underestimation of regular secondary structure elements, which may account for the observed disparity. Alternative explanations may include inaccurate sample concentration determination or insufficient representation in the reference database of CD spectra.

8.1.2.1 α-helices

The relative orientation and quality of the six sections of α-helix were assessed using QHELIX software [239].

Irregularity of Helical Shape	
Helix A’	2.28 Å
Helix A’’	2.01 Å
Helix A’’’	2.26 Å
Helix B’	2.35 Å
Helix B’’	2.25 Å
Helix C	2.33 Å
Inter-helical Angle	
Helix A’ - Helix A’’	35.0°
Helix A’’ - Helix A’’’	66.4°
Helix A’’’ - Helix B’	149.8°
Helix B’ - Helix B’’	41.9°
Helix B’’ - Helix C	93.9

Table 8.2- QHELIX analysis of the α-helix within the latherin structure.

All six helices are determined to deviate from a linear ideal. This irregularity can be observed in the structures as the α-helices kink and curve round the β-sheet. The largest sequential inter helix angle is observed between B'' and C where an angle of over 90° is observed. This reflects helix C's perpendicular orientation with regards to the majority of regular secondary structure features, reinforcing the proposed rationalization of the contrasting relaxation properties between these two sections of helix, observed in Chapter 7.

Helix B' is terminated by a short section of π-helix, defined by i to i + 5 hydrogen bonding between the amides of residues 171 and 172 and the carbonyl groups of 166 and 167 respectively. π-helices are believed to arise in α-helices as a consequence of a single residue insertion, without subsequent accommodation by the original secondary structure [240]. Sections of π-helix cause destabilisation of the regular secondary structure and are often associated with functional sites.

8.1.2.2 β-sheet

The four stranded β-sheet displays a classic right-handed twist which allows it to pack efficiently into the kinked α-helices. Strand 1 of the β-sheet (residues 61-77) is disrupted between residues 71-73 producing a small irregular section containing an extra amino acid relative to the adjacent region of strand two. The cause appears to be the presence of 89Pro in strand two, which does not present a hydrogen bonding partner for strand one. Furthermore, the steric interactions from the cyclic proline prevent this region of strand one from packing effectively next to strand two. The extra amino acid allows the preceding and following sections of strand 1 to form regular β-strand hydrogen bonded to strand 2.

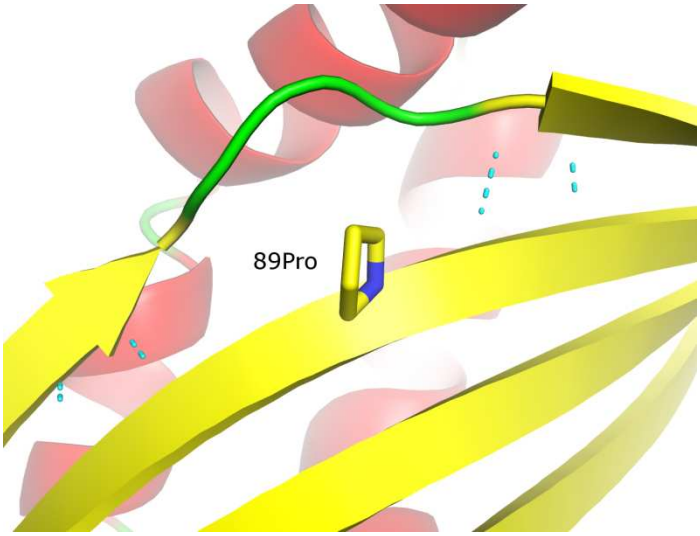


Figure 8.1- Image showing β-bulge and location of 89Pro. The backbone nitrogen is coloured blue. Hydrogen bonding between strands 1 and 2 are shown in cyan. Image was created using PyMOL [47]**.**

8.1.2.3 Overall

The overall distribution of secondary structure in latherin is summarised in a topology model in Figure 8.2.

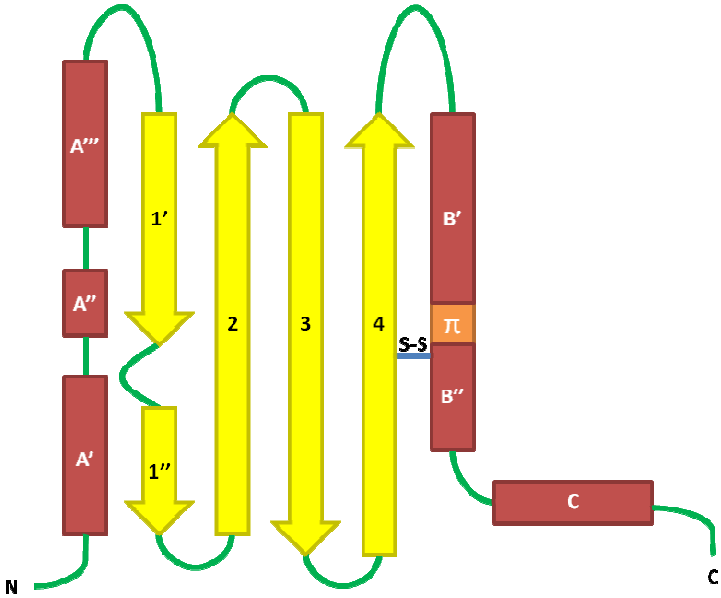


Figure 8.2- Topology model of latherin. α-helices are represented by red rectangles; β-strands by yellow arrows; and non-regular secondary structure as green lines. The intramolecular disulphide bond,133Cys-175Cys, is represented as a blue line. The short section of π-helix is represented as an orange square. The β-bulge by a curved green line between strands 1' and 1''.

8.1.3 Distribution of Residues

For latherin to function as a surfactant it must display some degree of amphiphilicity to adsorb effectively into the air-water interface. In a protein, since the backbone is hydrophilic and side chains are the only source of hydrophobicity, any such property is likely to be defined by the distribution of amino acids through the structure.

8.1.3.1 Aliphatic

The most striking feature of latherin's amino acid composition is the high proportion of leucines, 49 of 208 residues. This unusually high population of hydrophobic leucine residues has been proposed to have some role in the protein's surface activity. The leucines are distributed fairly evenly throughout the latherin structure, showing only a slight preference for the regular secondary structure (Table 8.3). In the structured regions the vast majority of leucines are buried in the interior while in the loops approximately one third are exposed to the solvent.

Secondary Structure	# of residues	# of leucines	% leucines
α-helix	81	20 (4)	24.7 (4.9)
β-strand	53	14 (1)	26.4 (1.9)
Other	74	15 (5)	20.3 (6.8)
Overall	208	49 (10)	23.6 (4.8)

Table 8.3- Distribution of leucines in latherin by secondary structure. The values for leucines exposed to the solvent are shown in brackets. Solvent exposed leucines were identified through manual inspection and from PROCHECK analysis of the calculated structures.

The five leucines in structured regions of the protein that are exposed to the solvent are distributed throughout the structure and represent single occurrences of hydrophobic residues at the surface rather than areas of hydrophobic nature on the protein's surface.

There is a large disparity between the proportions of leucines in the unstructured regions at the two extremes of the molecule. At the end of the protein that includes the two termini, the unstructured region consists of 21 residues, only one of which is a leucine. Whereas the three loops at the opposite end of the protein together contain 28 residues, 9 of which are leucines, corresponding to almost a third of the residues in these features and almost one fifth of the leucines in the structure. The majority of the other aliphatic residues display an even distribution, with the vast majority found within the hydrophobic core of the protein. The isoleucines, like the leucines are more frequently found in the loops at the non-termini

end of the protein. 2 of the 16 isoleucines are found on these loops with the other 14 buried in the interior of the protein.

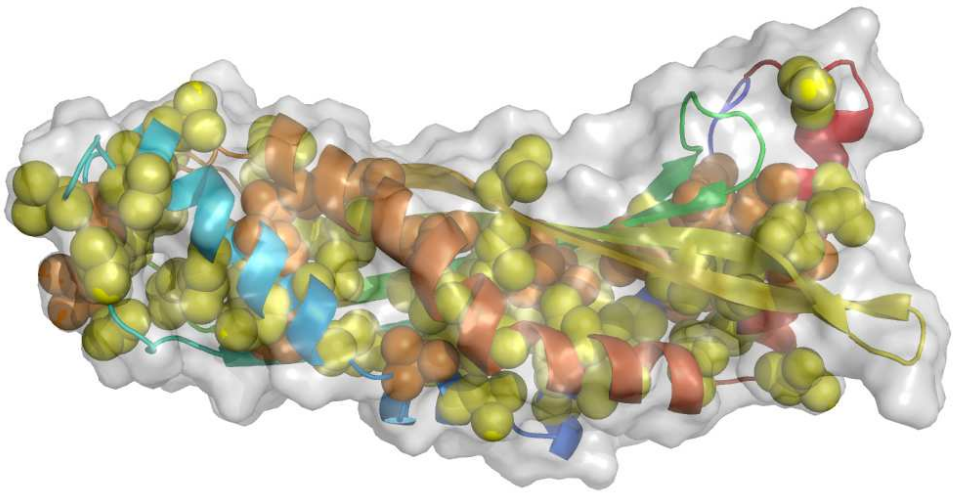


Figure 8.3- Distribution of leucine and isoleucine residues in the latherin structure. Leucine side chains are displayed as yellow spheres. Isoleucine side chains as orange spheres. Latherin is displayed in cartoon representation, coloured from blue at the N-terminus to red at the C-terminus. Image created using PyMOL [47]

8.1.3.2 Polar

Initially the distribution of polar residues appears to be entirely typical of a globular protein: polar or charged residues are located upon the surface of the protein, resulting in a hydrophilic exterior. The exterior of the protein is primarily anionic due to the higher proportion of aspartic and glutamic acids over arginines, histidines and lysines (Figure 8.4).

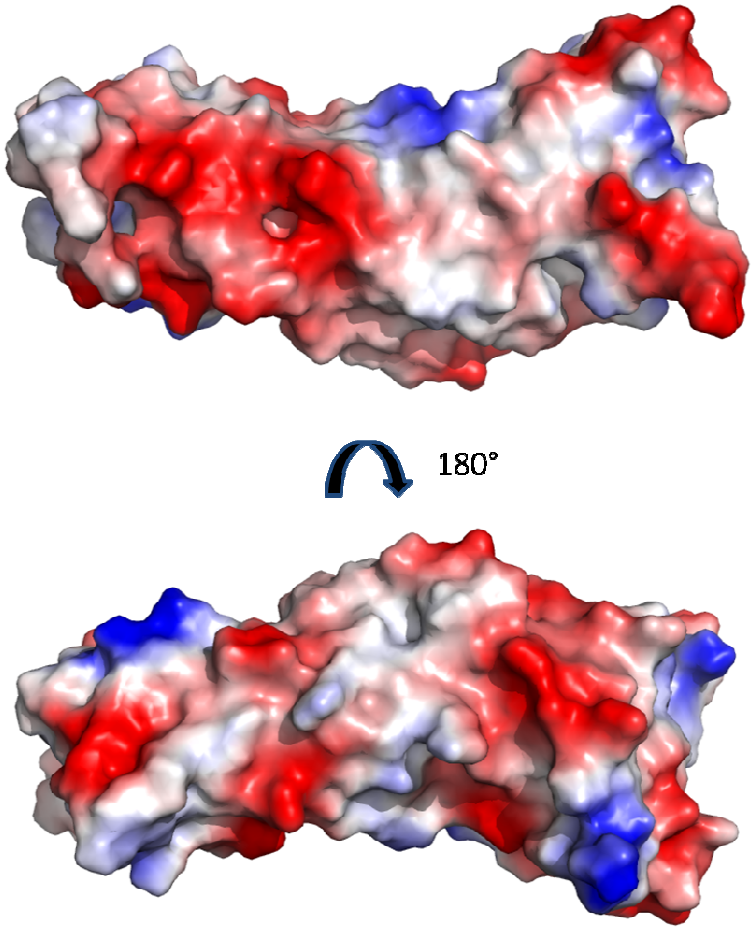


Figure 8.4- Surface contact potential of latherin. Two opposing faces of the molecule are coloured by charge: red = anionic, blue = cationic and grey = uncharged. Image created using PyMOL [47].

What is remarkable is the almost complete exclusion of polar residues from the interior of the protein. Only one polar residue, 110Thr is observed within the hydrophobic core of the protein where it is completely surrounded by hydrophobic aliphatic side chains. However its position may be stabilised by the side chain hydroxyl forming a hydrogen bond to the main chain carbonyl group of leucine 136. This clear segregation of polar and non-polar residues within the latherin molecule may have some role in its surfactant activity.

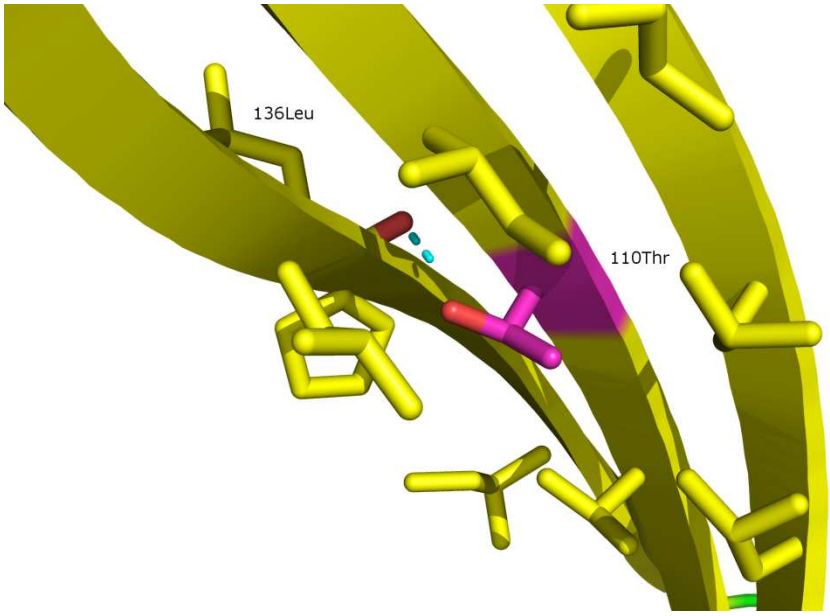


Figure 8.5- The environment of the buried threonine residue at position 110. 110Thr is coloured purple. Surrounding aliphatic residues are coloured yellow. The potential polar contact to the backbone carbonyl of 136Leu is depicted as a green dotted line. Image created using PyMOL [47].

8.1.3.3 Aromatic

The aromatic residues also show a striking distribution as discussed already with the four phenylalanines all found buried within the interior of the protein, whereas the three tyrosines are solvent exposed on the exterior of regular secondary structures.

In previous work the single tryptophan, 87Trp in latherin was found to produce an unusually high fluorescence emission at 350 nm that was not shifted by denaturation [73]. This led to the proposition that the residue is solvent exposed. Analysis of the solution structure shows that the tryptophan is located on the second strand of the β-sheet orientated towards the solvent. Further abnormalities were encountered during quenching studies: iodide, normally an effective quencher of tryptophan fluorescence was found to be relatively ineffective on latherin [73]. This was proposed to be a consequence of an anionic environment around the tryptophan that produced repulsive electrostatic effects with the iodide. The solution structure confirms this hypothesis, showing two aspartic acids: 85Asp and 111Asp, and one glutamic acid: 76Glu, within the immediate environment of the tryptophan (Figure 8.6).

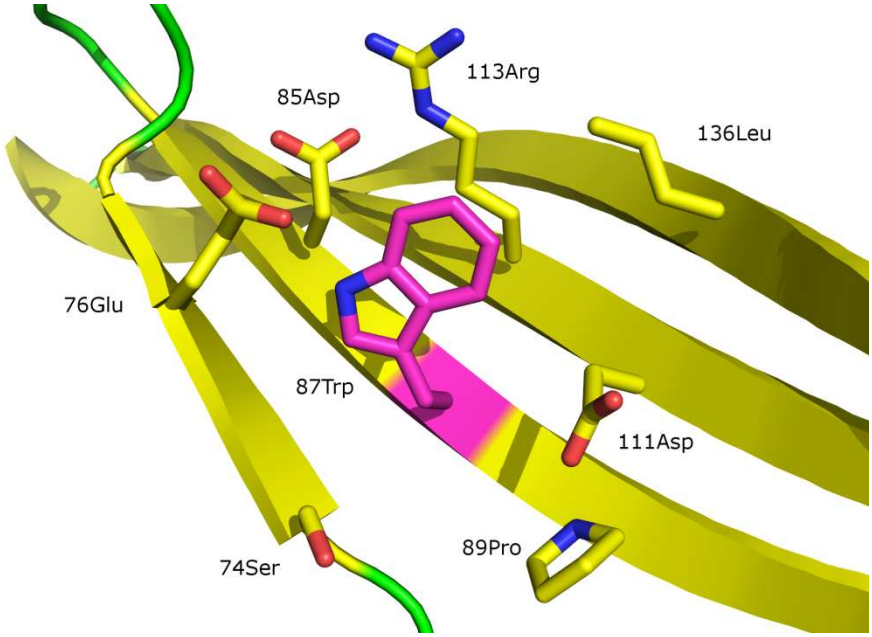


Figure 8.6- The anionic environment of solvent exposed tryptophan 87. Tryptophan 87 is coloured pink. Surrounding side chains are coloured yellow. Side chain oxygens are coloured red and nitrogens blue. Image created using PyMOL [47].

8.2 Comparison with Similar Structures

Comparison of a protein with other similar folds can aid the process of rationalising the relationship between a protein's structure and its function. The comparison can enable the identification of features that are common to the fold, perhaps as a consequence of a common ancestor, and subsequently features that are unique to the protein of interest that may have a role in the observed function.

8.2.1 PLUNC/BPI Superfamily

Visual inspection of the structure reveals similarity between the latherin molecule and single domains of the BPI and CETP structures discussed in Chapter 1. Visual analysis of the structures suggests that latherin is more similar to the N-terminal domains of both (Figure 8.7). As expected, the highly conserved disulphide found in the N-terminal domains of the BPI-like family is found in a very similar position in latherin (linking the final strand of the β-sheet to the C-terminal α-helix).

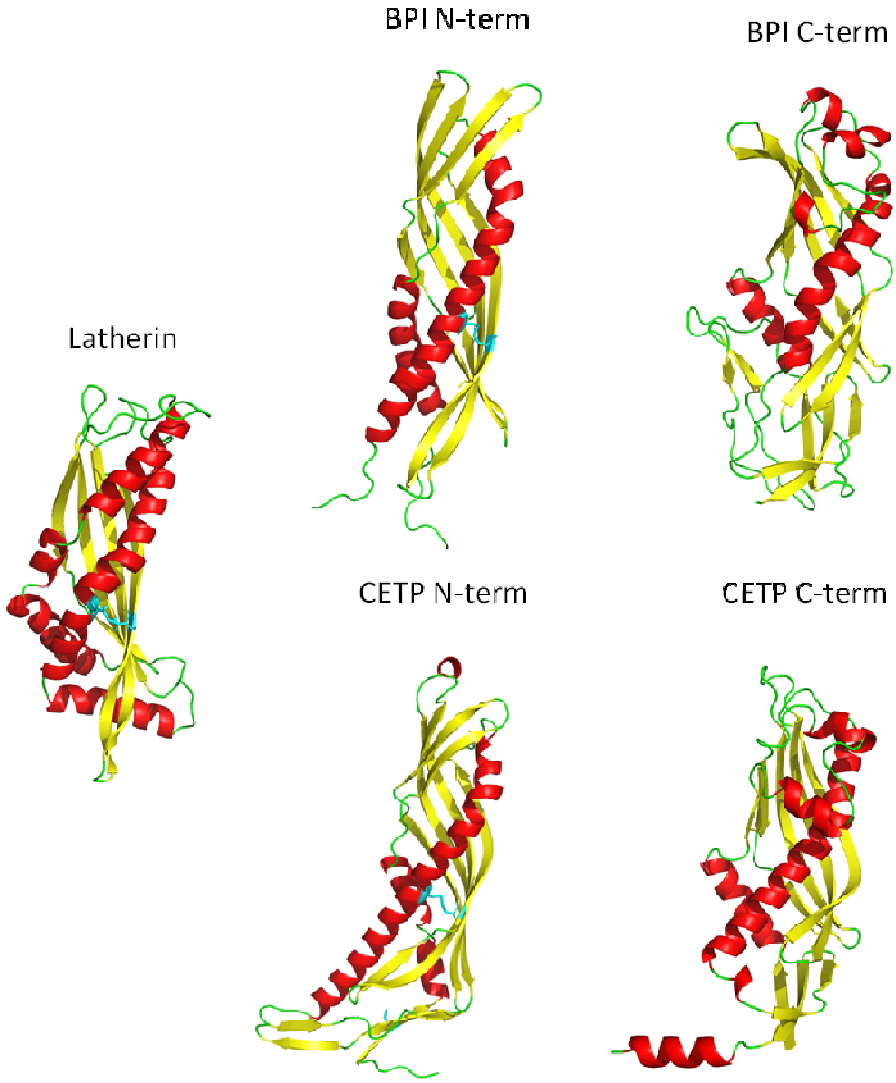


Figure 8.7- Structures of latherin and BPI-like proteins. Molecules are coloured by secondary structure: α-helix = red; β-strand = yellow; other = green. Disulphide bonds are shown in cyan. The latherin structure shown is the structure closest to the mean of the ensemble. BPI N and C-term were created from PDB file 1BP1, CETP N and C-term from 2OBD. Images created using PyMOL [47].

Latherin distinguishes itself from all four domains of BPI and CETP in its distribution of secondary structure. In all the other structures an initial section of N-terminal helix is followed by a region of β-strand that acts as the first strand of a five stranded β-sheet (Figure 8.8). In latherin this feature is replaced by a long section of α-helix, leaving a four stranded β-sheet.

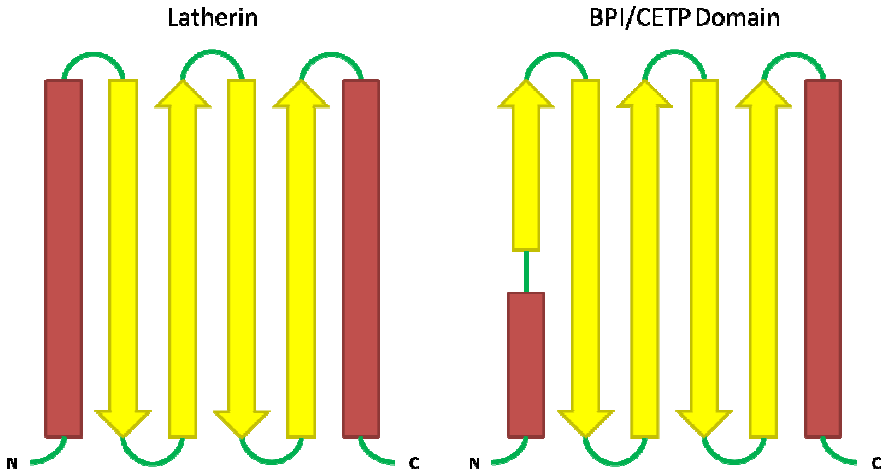


Figure 8.8- Simplified topology models of latherin and a BPI/CETP domains. α-helices are represented by red rectangles, β-strands by yellow arrows and non-regular secondary structure as green lines. For simplicity features such as short breaks in regular secondary structure have been omitted from the models.

Further analysis of the BPI and CETP structures indicate that many of the unusual secondary structures found in latherin (discussed above) are also found within other members of the BPI-like family [241]. Short sections of π-helix are found in the C-terminal helices of both domains of BPI and CETP. The π-helix in latherin in particular appears to align very well with the π-helix in the C-terminal domains of the two proteins (Figure 8.9).

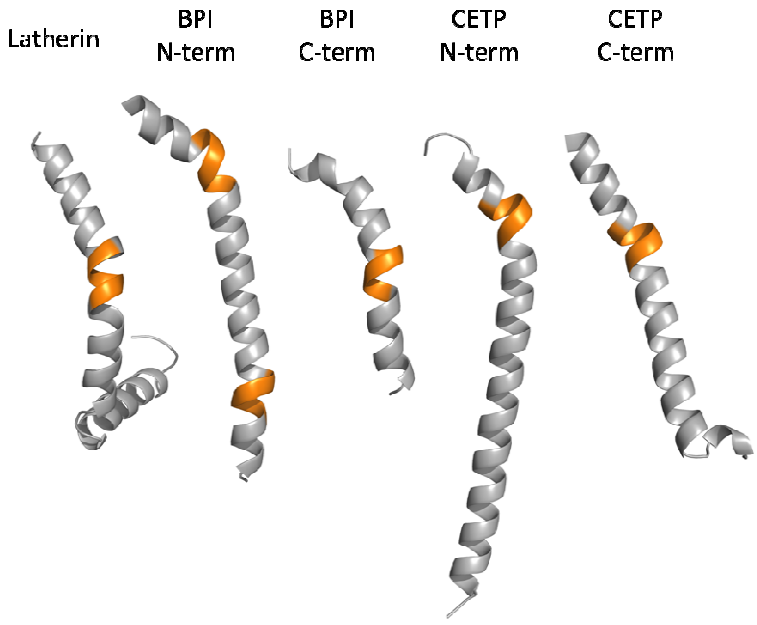


Figure 8.9- Location of π-helix in latherin, BPI and CETP. The C-terminal helical regions from latherin (Helix C:152-203) and the two domains of BPI (N-term 152-192; C-term 398-423) and CETP (N-term 162-206; C-term 403-436) are shown in grey. The areas defined as π-helix are coloured orange. The latherin structure shown is the structure closest to the mean of the ensemble. BPI N- and C-term were created from PDB file 1BP1 and CETP N and C-term from 2OBD. Images created using PyMOL [47].

The BPI and CETP structures also indicate that the β-bulge in the first strand of latherin is a consistent feature between the structures. Only in the N-terminal domain of CETP is a β-bulge absent from the equivalent position. In the C-terminal domain the bulge appears to also be caused by a proline (329Pro). However, in this case the proline is located upon the first strand rather than the second as in latherin. In BPI, the N-terminal domain has two adjacent β-bulges on the first two strands, again probably as a consequence of a proline in the first strand (64Pro) and two adjacent polar residues in the second strand (81Ser and 82Asn). In the C-terminal domain of BPI a single β-bulge is observed in the first strand, caused by two consecutive proline residues (323Pro and 324Pro) in the same strand and another in the adjacent strand (337Pro).

The role of these features, if any, is not known. In CETP the presence of the π-helices and β-bulges have been postulated to aid structural rearrangements of the molecule involved in activity, either through destabilisation of the secondary structure or peristalsis-like shifts of the π-helices along the helices [241].

In Figure 8.9, it can be observed that the π-helices cause the observed kinks in the helices and therefore a possible alternative explanation of the occurrence of these abnormalities in regular secondary structure may be to allow the long sections of helix and β-sheet to curve round one another, allowing maximum efficiency of packing. The increased stability gained, may offset the reduced stability of the secondary structure.

8.2.2 Similarity Searches

Database comparisons identify the fold that latherin adopts as a Aha1/BPI domain-like (SCOP [242], code 55393) or as a super-roll (CATH [243], code 3.15).

Searches for other proteins with related folds were carried out using DALI (L. Holm and P. Rosenstrom, http://ekhidna.biocenter.helsinki.fi/dali_server/) [244]. The search returned over 1000 results. Structures with better than marginal match scores (z score > 5) are shown in Table 8.4. As well as the structures of BPI and CETP, the search identified several similar structures from outwith the cognate PLUNC/BPI superfamily of proteins.

Molecule	PDB Code	Z-score	RMSD Å	Length of Alignment	No Residues
BPI	1BP1	10.4	4.1	166	456
Der P 7	3H4Z	8.4	4.2	158	563
JHBP	2RCK	7.8	4.6	157	208
Top1	3E8T	6.4	4.6	149	216
Aha1	1USV	6.4	4.0	119	143
CETP	2OBD	5.9	5.4	140	472
Yceb	3EYR	5.7	3.8	121	169

Table 8.4- DALI results. The results are ranked by Z-score. RMSD is calculated between C_α atoms within the aligned regions of the structures.

Der p 7 is a 198 amino acid dust mite allergen that elicits strong IgE antibody and T-cell responses in patients with mite allergy through an unknown action [245]. The similarities of Der p 7 to BPI led Mueller *et al*. to highlight the connections between the allergen and the BPI and LBP proteins involved in the innate immune response. It was also suggested that although Der p 7 did not bind LPS, it was likely to bind some unknown bacterially derived lipid.

Juvenile hormone binding protein (JHBP) transports the highly hydrophobic and chemically labile juvenile hormone (JH), integral to the insect life-cycle. The structure of JHBP from *Galleria mellonella* (245 amino acids) indicates two cavities at either end of the molecule [246]. The cavity at the non-termini end of the molecule is believed to be the site of JH binding, whereas the role of the cavity at the termini-end is unknown.

Takeout protein 1 (Top1) is another insect protein associated with lipid binding activity and shows low levels of sequence similarity to the JHBP proteins. The crystal structure of *Epiphyas postvittana* takeout 1 reveals a continuous hydrophobic tunnel analogous to that found in CETP, containing an *E.coli* derived, ubiquinone-8 molecule [247].

Aha1 is a co-chaperone of the molecular chaperone, heat shock protein, Hsp90. Aha1 stimulates Hsp90's ATPase activity and is critical for its function [248]. The crystal structure of the complex between the N-terminal domain of Aha1 (1-153) and the middle segment of Hsp90 reveals an interaction via a series of hydrophobic and polar residues on the surface of Aha1's two α-helices to a corresponding region upon Hsp90. This interaction causes a structural rearrangement of Hsp90 into its active form. No role for the C-terminal domain which does not form a super roll has thus far been proposed.

The structure Yceb corresponds to an as yet uncharacterised lipoprotein from *E.coli*.

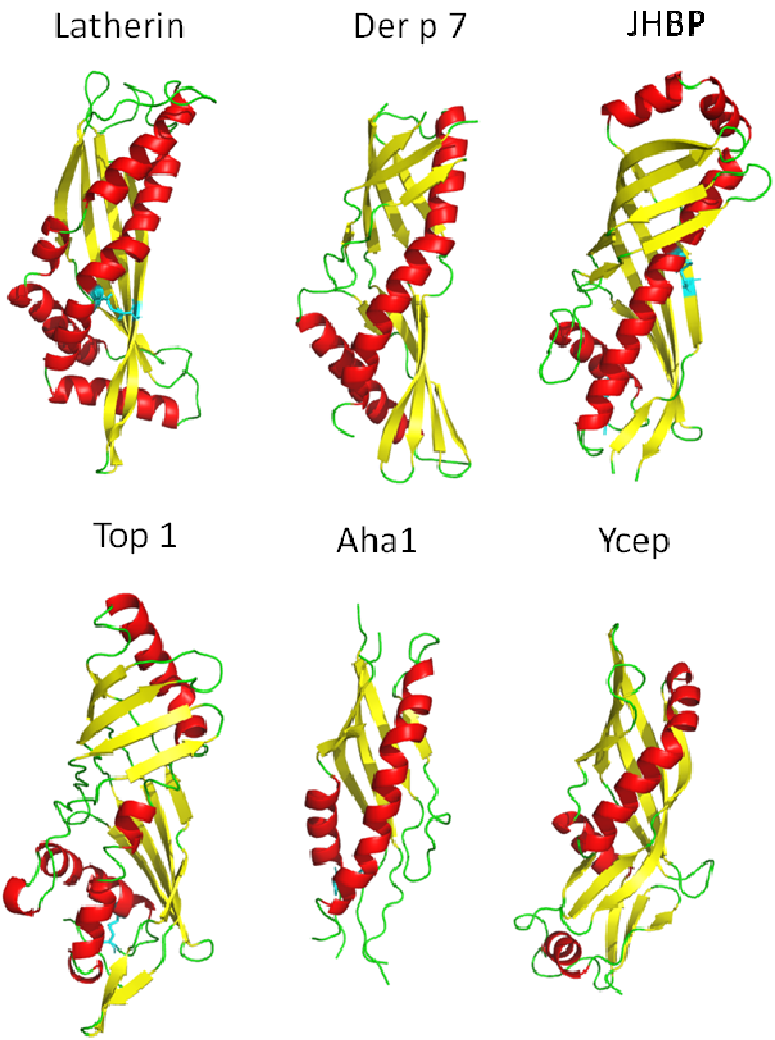


Figure 8.10- Structural Comparison of latherin with other similar proteins identified by DALI. The structure cartoons are coloured by secondary structure: α-helix = red; β-strand = yellow; other = green. Disulphide bonds are shown in cyan. Images were created from the following PDB files: Der P 7 from 3H4Z; JHBP from 2RCK; Top1 from 3E8T; Aha1 from 1USV; and Ycep from 3EYR. Images created using PyMOL [47].

All the structures show a regular secondary structure distribution similar to that of the domains of BPI and CETP, where the N-terminal α-helix is followed by a section of β-strand that contributes to a five-stranded β-sheet. This suggests that this is a feature common to super roll folds with latherin being the exception.

The single disulphide bond in latherin that is highly conserved in the PLUNCs and the N-terminal domains of the BPI-like proteins, is not found in any of the other similar structures. JHBP does however contain a single disulphide between the two corresponding regular secondary structure features at a position closer to the N-terminal of the final section of α-helix.

Short sections of π-helix can be observed in the C-terminal helices of Der P 7 and Aha1, at positions very similar to the π-helix within latherin. None of the structures display a β-bulge corresponding to that seen on latherin, although many do have breaks in the regular secondary structure at a similar location and in some cases upon the adjacent stands of the β-sheet.

Almost all of the proteins that have a similar fold to latherin have been shown to, or are believed, to bind lipids or hydrophobic molecules within hydrophobic cavities. No evidence of the presence of any bound ligands was observed in this or any previous work investigating latherin, although it should be noted that no specific search has ever been carried out.

The solution structure of latherin indicates that the majority of the hydrophobic core of the protein is tightly packed. This feature of the calculated structure is reinforced through analysis of the NOESY spectra, where numerous NOE peaks between residues upon opposite sides of the core, indicate their close proximity. The structure also suggests the presence of two cavities, consistent through all the structures within the ensemble. One small enclosed cavity occurs at the termini end of the protein, completely surrounded by aliphatic side chains. The other appears between the dynamic loops at the other end of the molecule and extends down into the structured region of the protein. This cavity is exposed to the solvent and whilst the majority of its surface is surrounded by aliphatic side chains, two aspartic acids (42Asp and 60Asp) and a threonine (45Thr) surround parts of its entrance. The cavity has a length of approximately 16 Å from the tips of the loops to the furthest point within the core of the protein. Unless significant structural rearrangement occurs on ligand binding, the properties of the cavity would therefore suggest that any binding lipids would be short with a cationic head group.

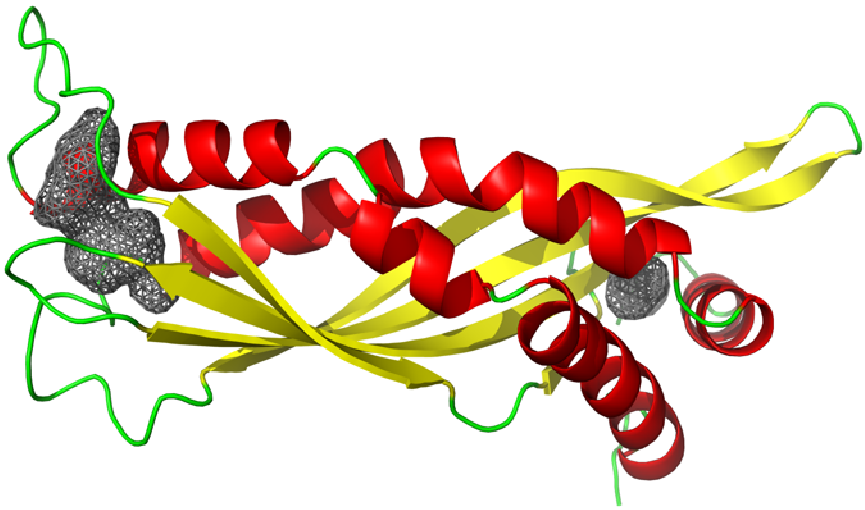


Figure 8.11- Cavities within the latherin solution structure. Cavities were detected with a cavity detection radius and cut off of 4.2 Å. The two detected cavities are represented as grey mesh. Image created using PyMOL [47].

Inspection of the NOE spectra of the residues surrounding the cavities failed to identify any consistent, unassignable, peaks that may indicate the presence of a bound molecule, or any inter-residue peaks that would suggest the cavity is an artefact of the structure calculation. However, due to the high proportion of methyl or methylene resonances (which are likely to overlap any signals from a hydrophobic chain) in this region, and thus the high ambiguity of the NOE spectra, it is difficult to either confirm that the cavity is real or identify any resonances from a bound species.

Analysis of the cavities within the other super roll structures identifies a large range of sizes and location of the cavities. The largest cavities are observed in the C-terminal domain of CETP and in Top1. These structures as well as sections of JHBP appear to have a slightly looser fold than many of the other super roll structures in order to incorporate these features. However, many of the other proteins have cavities significantly larger than that found within latherin with what appear to be equally compact tertiary structures. In some cases the absence of an equivalent cavity in latherin can be rationalised by the variations in the orientation of the regular secondary structure. For example, in the N-terminal domains of BPI and CETP the cavities are located at the termini ends of the proteins. In latherin, this region is occupied by helix C which turns along the end of the molecule. In many of the other structures the presence of a large cavity appears to be a consequence of the identity of the residues that make up the core of the protein. A good comparison is between latherin and the C-terminal domain of BPI: both structures have

similar folds, almost identical diameters, but the BPI C-terminal domain has a large cavity within its hydrophobic core (Figure 8.12).

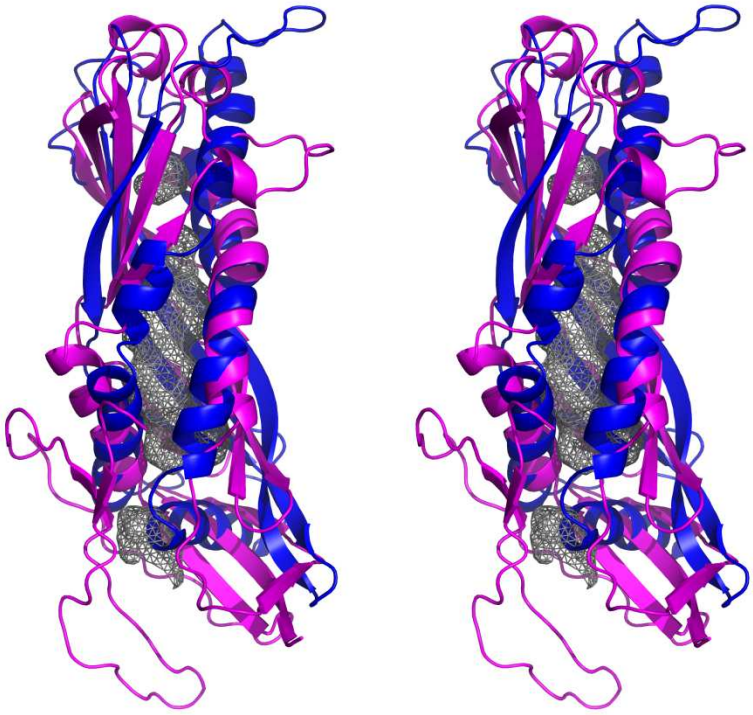


Figure 8.12- Superposition of latherin and BPI C-terminal domain. A stereo view of cartoon representations of latherin (blue) and the C-terminal domain of BPI (magenta) superimposed upon one another, demonstrating the similarity between the two structures. The internal cavity of the C-terminal domain of BPI is shown as grey mesh.

Analysis of the residues that surround this cavity identifies a number of glycine, alanine, proline, valine, methionine, leucine, isoleucine and phenylalanine residues (Figure 8.13), whereas the corresponding region in latherin is inhabited almost exclusively by leucines, isoleucines and phenylalanine residues (Figure 8.14). To accommodate the preponderance of bulky, aliphatic side chains, latherin's core is tightly packed without cavities, whereas the presence of the shorter, less bulky side chains of residues such as glycine, alanine, proline and valine contribute to the formation of a significant cavity within the C-terminal domain of BPI. Therefore, a possible reason for the high proportion of leucine residues found in latherin is to fill redundant cavities that occur within the super roll fold, providing greater stability of the molecule in solution due to increased hydrophobic interactions.

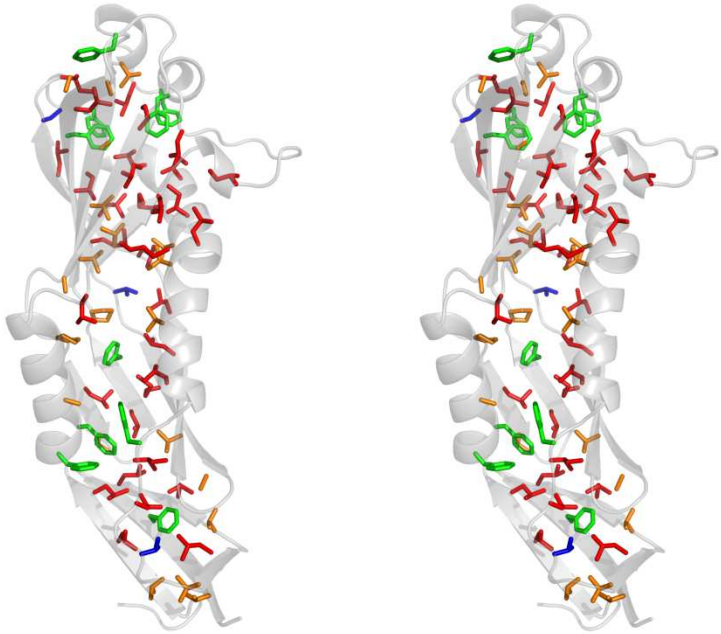


Figure 8.13- Internal residues in the C-terminal domain of BPI. Stereo views of a cartoon representation of the C-terminal domain of BPI. The side chains of buried residues are shown as sticks. Bulky, aliphatic side chains (Leu, Ile, Met) are coloured red, other aliphatic side chains (Val, Ala, Pro) are coloured orange, aromatic side chains are coloured green and polar residues are coloured blue.

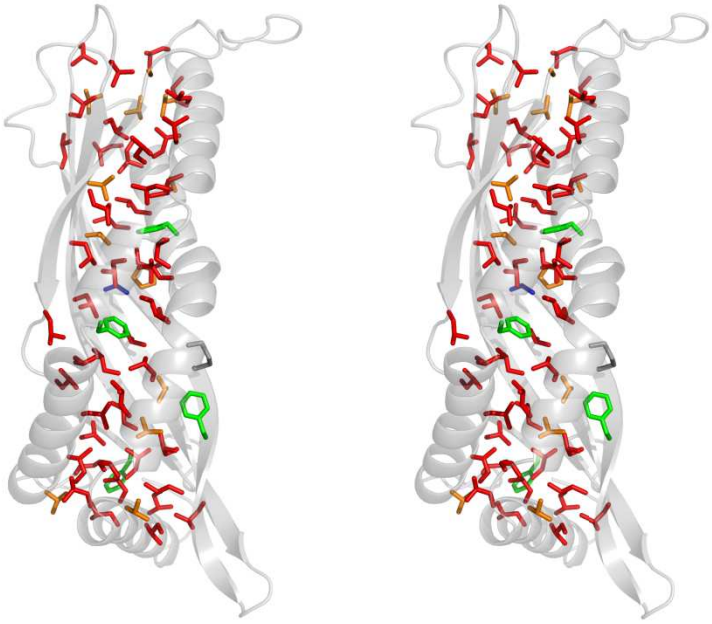


Figure 8.14- Internal residues in latherin. Stereo views of a cartoon representation of latherin. The side chains of buried residues are shown as sticks. Bulky, aliphatic side chains (Leu, Ile, Met) are coloured red, other aliphatic side chains (Val, Ala, Pro) are coloured orange, aromatic side chains are coloured green and polar residues are coloured blue.

The structural similarity between latherin and dust mite allergen Der p 7 is also interesting in that latherin has been identified as the source of horse dander allergen peptides Equ c 4

and Equ c 5 [73]. Equ c 4 which is derived from latherin residues 56-106, corresponds to part of the inter $\alpha_{A''}$ β_1 loop, the entirety of β-strands 1 and 2 and the first 6 residues of strand 3. Equ c 5 consists of residues (127-138) found in β-strand 4. The exact molecular mechanisms that cause allergenicity are poorly understood and extend beyond the scope of this thesis [249]. However, the common structural features between the two molecules may indicate some shared property that incites a subsequent immune response. The presence of bound lipids, which may have a role in stimulating an immune response, cannot be confirmed or ruled out in either case.

8.3 Implications for the PLUNCs

The high degree of sequence similarity between latherin and members of the PLUNC family, as well as biophysical data, suggests that the distribution of secondary structure in the PLUNCs is closer to that of latherin than either of the domains of BPI or CETP (discussed in Chapter 1). The latherin structure presents an opportunity to gain insight into the structure and character of the PLUNC proteins. In particular, human SPLUNC1, is of additional interest. Like latherin it contains a remarkably high proportion of leucines and displays similar levels of surfactant activity to that observed for latherin [100]. One would assume that the surfactant activity observed in each protein is a consequence of some shared molecular feature(s) that are not present in BPI, which shows no such activity. Insight into the PLUNC structure may aid identification of these features.

8.3.1 Latherin-PLUNC Alignment

An alignment was created using the amino acid sequence of human PLUNC (NP_057667) with the proposed signal sequence removed and the sequence of latherin [100]. The majority of the two sequences align very well (Figure 8.15) due to the reasonable degree of sequence identity (28%). Most of the secondary structure features appear to be highly conserved, as demonstrated by the high degree of sequence identity within these features. In particular, helix A' and strands 1, 3 and 4 show good levels of identity. The conserved disulphide also aids the alignment.

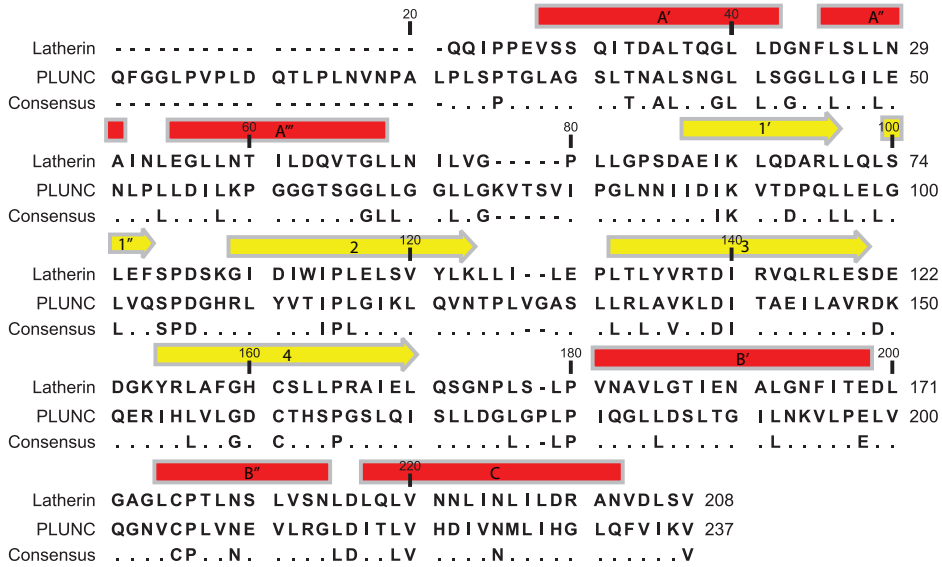


Figure 8.15- Sequence alignment of Latherin and human PLUNC. The regular secondary structure in latherin is shown as red rectangles for α-helix and yellow arrows for β-strand.

The alignment identifies several inequivalent features between the two proteins. The largest section of disparity is the first 21 residues in the PLUNC sequence which, secondary structure prediction software (PSIPRED, [59]) predicts to be unstructured. PLUNC also has three short insertions that fall between conserved regions of regular secondary structure in the latherin structure. Each of these sections occur in the loops at the non-termini end of the protein. The additional residues may correspond to extended sections of the regular secondary structure preceding or following these loops, or simply as longer sections of loop. Secondary structure prediction software (PSIPRED, [59]) predicts that all the additional residues form coil rather than regular secondary structure, suggesting that extended loops are the more likely. There is also an apparent inequivalence in the regular secondary structure features between the latherin template and the PLUNC sequence. In the alignment, the sequence GGGTSGG is aligned alongside the latter half of helix A’’’. This glycine rich sequence is highly unlikely to form α-helix, as occurs in latherin, or even β-strand, as occurs in the majority of super rolls. It appears probable that a short section of helix is terminated by 60Pro and the following sequence, up to the start of the first strand of β-sheet is disordered. This proposal agrees well with PSIPRED secondary structure prediction.

8.3.2 PLUNC Model

In order to aid visualisation of the human PLUNC structure a model was created. The MODELLER programme (Andrej Sali, University of California San Francisco, http://salilab.org/modeller/) was utilised to create five models based upon the sequence alignment in Figure 8.15 using the latherin structure determined in this work as a template. The approach used in MODELLER is analogous to the process utilised in the determination of protein structures from NMR-derived restraints, discussed in Chapter 6 [250]. Two sets of spatial restraints are used to define the molecule: homology-derived restraints, which include distances and dihedral angles, derived from the alignment of the target sequence with the template structure, and empirical restraints which define statistical preferences of bond angles, bond lengths and other elements of standard covalent geometry. Optimisation is then carried out to minimize the violations of these restraints and produce a structural model. In order to create a model as representative of the likely PLUNC structure as possible, the extended N-terminus which is not present in latherin was omitted and the extended loops and glycine rich sequence were left unrestrained in the alignment. The models were analysed and the model with the lowest Discrete Optimised Potential Energy (DOPE) score [251] was selected.

8.3.3 Analysis of the PLUNC Model

The alignment between latherin and PLUNC, suggests that the majority of the secondary structure is conserved between the two molecules. The exception is a shortened section of α-helix in the PLUNC model that corresponds to the helix A''' in the latherin structure. In the latherin structure, at this end of the protein, helix A''' extends out from the other regular secondary structure features. In the PLUNC model, the end of the helix aligns with the end of the β-sheet and helix B and there is no reason to suspect the shortened helix affects the packing of PLUNC's core. This does however result in what appears to be a very long unstructured loop (26 residues) between the N-terminal helix and strand 1. The exact role and form of this loop is difficult to predict based upon sequence information alone.

Our model predicts that the β-bulge observed in latherin is conserved in PLUNC. This can be determined by the presence of a conserved proline in the middle of the second strand of the β-sheet. Interestingly the residues that appear in the bulge are also well conserved. In latherin these residues are: 70Leu, 71Leu, 72Gln, 73Leu and in PLUNC: 96Leu, 97Leu, 98Glu, 99Leu.

The PLUNC model shows a similar distribution of residues to that observed in latherin. The same high occurrence of leucine residues in the loops at the non-termini end of the protein is observed: 11 of the 50 leucines within the PLUNC structure are found within these loops. A similar distribution is observed for the isoleucines. Other aliphatic residues appear to be distributed evenly throughout the molecule.

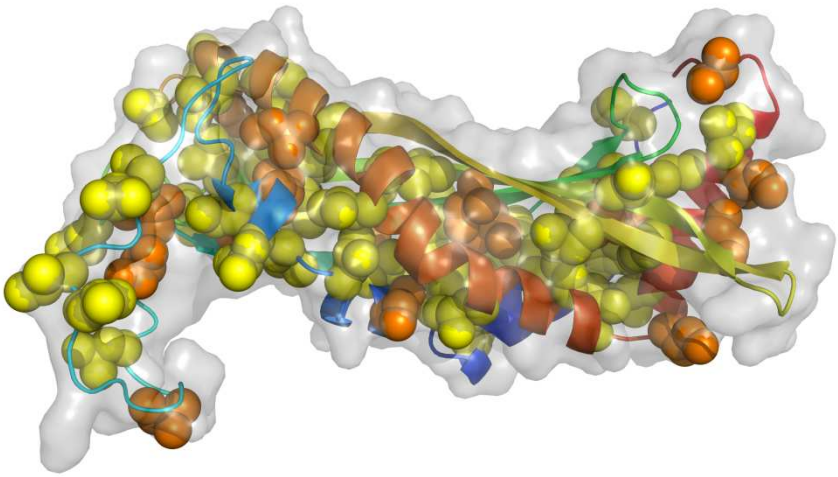


Figure 8.16- Distribution of leucine and isoleucine residues within the human PLUNC model. Leucine side chains are displayed as yellow spheres. Isoleucine side chains as orange spheres. Human PLUNC is displayed in cartoon representation, coloured from blue at the N-terminus to red at the C-terminus. The leucines are distributed evenly throughout the protein apart from a high number found on the loops at the non-termini end of the protein (to the left of the figure). Image created using PyMOL [47]

Like latherin the exterior of the model is highly polar and there is no indication of a hydrophobic patch on the surface (Figure 8.17). A distinct absence of any polar residues from the core of the protein is also observed in the PLUNC model. The exception in this case is a single histidine (163His). Any role for this histidine is unclear. In the model it is predicted to form hydrogen bonds with the main chain carbonyl of 200Val, upon the adjacent α-helix.

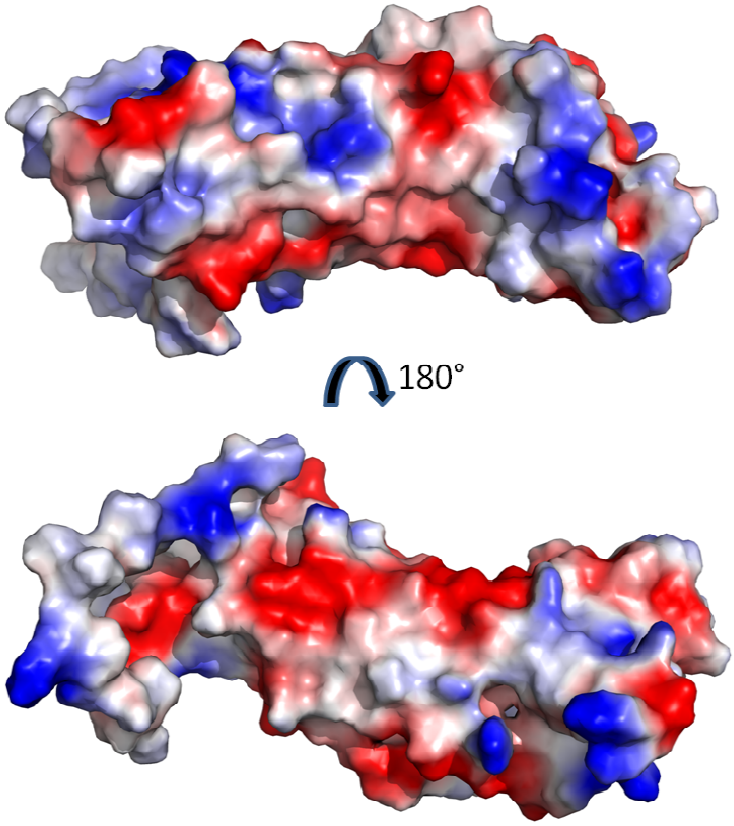


Figure 8.17- The surface contact potential of the human PLUNC model. Shown for two opposite faces of the molecule. Red = anionic, blue = cationic and grey = uncharged. The model suggests no evidence of any large uncharged areas that may correspond to a hydrophobic patch. Image created using PyMOL [47].

There are very few aromatic residues within the PLUNC sequence: no tryptophans, a single tyrosine and two phenylalanines at the two extremes of the sequence. The N-terminal phenylalanine (2Phe) lies within the N-terminal section not included in the model. The other (233Phe) lies in the unstructured C-terminus. The single tyrosine (111Tyr) lies in strand 2, directed out into the solvent.

9 Molecular Basis of Surfactant Activity

The solution structures of the two surfactant proteins provide little indication of the surface amphiphilicity that would be expected of a surface active molecule. They do however, along with biophysical data gathered in this work and elsewhere, provide a starting point for speculation as to how these molecules may function. In this chapter various aspects of both proteins' activity will be discussed, leading to the creation of hypothetical models of activity and proposals as to how these models may be probed further.

9.1 How are the Proteins Adsorbed into the Air-Water Interface?

For a molecule to behave as a surfactant it must possess some degree of amphiphilicity that drives adsorption into the air-water interface. As discussed in Chapter 1, in the hydrophobins this feature arises in the form of a large hydrophobic patch shielded from the solvent by the formation of dimers. Figure 1.10 and Figure 8.4 demonstrate that the exterior surfaces of Rsn-2 and latherin (and according to the model, human PLUNC (Figure 8.17)) are covered in polar residues with no indication of any large hydrophobic patch. Furthermore, both proteins have been identified as monomers in solution, ruling out the possibility of dimerisation to shelter hydrophobic features from the solvent. Rsn-2 and latherin have therefore distinguished themselves from the hydrophobins through the molecular basis of their activity.

In the case of Rsn-2 the only apparent concentration of aliphatic residues, outwith those that constitute the hydrophobic core of the protein, lie upon the dynamic N-terminus. Of the seventeen residues preceding the α-helix, 6 are leucines, 2 are isoleucines and 1 is a valine. The hydrophobicity of the N-terminus is demonstrated by a GRAVY (grand average hydropathy) score, the average hydropathy value of a peptide or protein sequence [60], of 0.694 for the N-terminus, compared to -0.830 for the entire protein. Although the terminus is predominately hydrophobic, many of the remaining residues within this region (3 Asp, 3 Lys, 1 Pro and 1Gly) possess distinctly polar characteristics and therefore it is more accurate to describe the N-terminus as amphiphilic, rather than hydrophobic.

The contribution of the dynamic N-terminus to surfactant activity was determined in Chapter 3. Mutants with the N-terminus removed, displayed both reduced rates and total effect upon surface tension of buffer solutions. It was concluded that the N-terminus must

not only be involved in the adsorption of the protein into the interface, but also the final amphiphilic form of the molecule. The reduction in rate of onset of activity observed for the mutant is far more significant at lower concentrations, suggesting that at high concentrations the lack of the amphiphilic N-terminus can be overcome through naturally occurring, high surface concentrations. Whereas, at low concentrations the amphiphilic N-terminus is required to drive protein adsorption from the bulk solution into the interface. That the activity of the mutant, although significantly reduced, is not completely inhibited at low concentrations suggests that the terminus is a catalyst for the adsorption of Rsn-2 into the interface, as oppose to the sole feature responsible. The necessity for rapid adsorption to the interface may be a consequence of biological pressures. Perhaps, to minimise energy exerted by the frog in the process of “foam whipping”, or to prevent dispersion of the protein into the bulk solution before foam production.

The latherin structure presents two structural motifs that stand out as unusual surface features. One is the surface exposed tryptophan (87Trp), suggested in previous work to be a potential hydrophobic element for orientating the molecule into the interface [73]. The anionic residues surrounding the tryptophan (Figure 8.6) would demonstrate that the tryptophan is not part of any sort of hydrophobic patch. Furthermore, tryptophans are not particularly hydrophobic relative to residues such as isoleucine, valine and leucine and, according to the hydropathicity index created by Kyte and Doolittle, show little preference for either hydrophobic or hydrophilic environments [60]. The surface active human PLUNC contains no tryptophans, nor is there any evidence of any surface exposed aromatic or aliphatic residues at the corresponding region on the PLUNC model. It is therefore difficult to imagine, that this lone surface exposed tryptophan is responsible for, or even involved in the proteins surfactant activity.

The other motif that stands out upon the surface of latherin is the three loops at the non-termini end of the molecule which contain a high proportion of aliphatic residues. The high occurrences of aliphatic residues in these three loops make it very difficult to propose a manner in which these loops could pack to shield this hydrophobic character from the solvent. Furthermore, relaxation data identifies these loops, and the residues surrounding them, as the main areas of chemical exchange in the protein. These hydrophobic loop features are also observed in the PLUNC model, a protein which has also been shown to be surface active. In the BPI structure, which has shown no surfactant activity [100], these loops are populated by polar residues, in particular basic lysine, arginine and histidines which have been shown to be involved in the proteins binding to anionic LPS [81].

Loop	Latherin	PLUNC	BPI N-Term
Inter α_A-β_1	LNILVGPLLGPSD	LGKVTSVIPGLNNIID	KH
Inter β_2-β_3	LLILEP	TPLVGASL	RF
Inter β_4-α_B	QSGNPLSLP	SLLDGLGPLPI	ISKSKV

Table 9.1- Sequences of the loops at the non-termini end of latherin, human PLUNC and the N-terminal domain of BPI. Hydrophobic residues are underlined. The loop between α_A and β_1 in the PLUNC model is not well defined due to the preceding glycine rich sequence, therefore only the section which aligns with the loop in the latherin structure is shown.

The same loops have been shown to be critical in the activity of both LBP and PLTP. In LBP, as in BPI, the loops contain cationic residues believed to mediate interaction with LPS [91], whereas in PLTP the same loops, which show a distinct hydrophobic character due to a number of aromatic residues, have been found to be critical in the molecule's ability to bind HDL [114]. Therefore in several members of the PLUNC/BPI superfamily, the loops at this end of the domain are critical to activity. It seems likely that the same could be true for latherin and human PLUNC. These surface active proteins have evolved hydrophobic loops, which rather than show affinity for a specific lipid, drive the adsorption of the protein into the air-water interface, where the molecule is then able to function as a surfactant.

That the features identified as the most prominent hydrophobic features in both proteins are located upon highly mobile regions may provide insight into another area of the proteins' activity. The dynamic character of these motifs may allow them to reorient and pack independent of the main structure in such a way that the molecule does not exhibit low water solubility or is required to form multimers to shield hydrophobic character from the solvent as in the hydrophobins, but still permits the features to be available to recognise the air-water interface when it presents itself.

9.2 What is the Form of Protein at the Interface?

Once at the interface the molecules must display some degree of amphiphilicity in order to function as a surfactant. Two possible models for activity present themselves. In the most simple, the protein molecules cause a reduction in surface tension via the hydrophobic features identified in Section 9.1 alone and the protein maintains its solution structure (Figure 9.1).

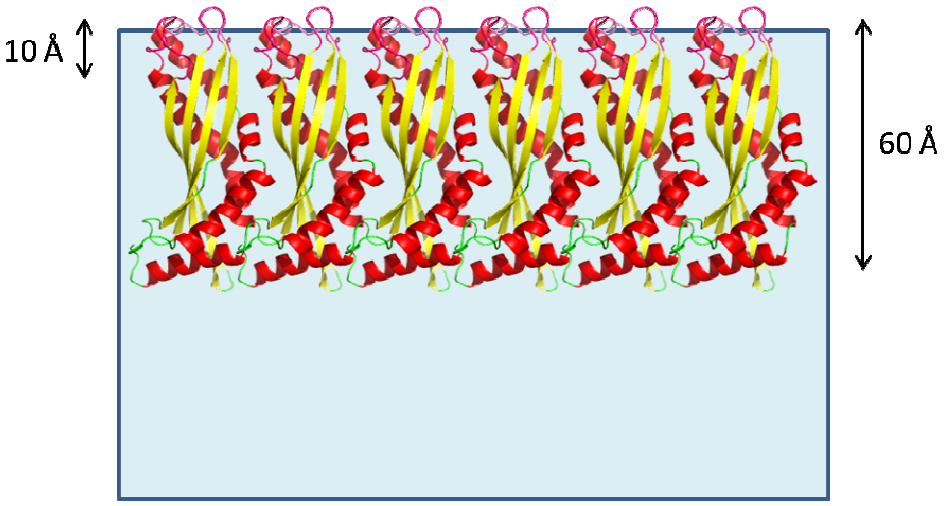


Figure 9.1- Schematic representation of a proposed mode of action for latherin where structure is retained. The dynamic, hydrophobic loops are shown in pink. The remainder of the protein is coloured by secondary structure: red = α-helix, yellow = β-sheet, green = unstructured.

For both Rsn-2 and latherin, this model would be inconsistent with neutron reflectivity data that describes a 10 Å thick layer of protein at the interface in both cases [62,73]. Even in their narrowest dimension, both proteins have a diameter greater than 20 Å. That the N-terminus of Rsn-2 is not solely responsible for activity is further demonstrated by the results of the mutagenesis in Chapter 4, where although the mutant with the N-terminus removed did display reduced activity vs. the wild-type, significant levels of activity were retained.

The experimental data gathered thus far indicates another potential mode of action: conformational change upon reaching the interface. A model for the potential structure of this alternate conformation of Rsn-2 has been proposed [62]. In this work attempts were made to probe and confirm the presence of this alternate Rsn-2 structure. Investigations using both NMR and mutagenesis provided insufficient evidence to prove or disprove Rsn-2 conformational exchange.

A similar model, whereby the molecule, once at the interface, unfolds to expose its hydrophobic core to the air can be proposed for latherin. An unfolded, planar, latherin molecule would be consistent with a layer 10 Å thick. The distinct separation of hydrophobic and polar residues within the proteins solution structure would result in distinctly amphiphilic character of the resultant surface structure. The requirement for the molecule to unfold may also explain some of the unusual secondary structural features observed in the latherin structure, such as the π-helix, the β-bulge and that latherin is the only super roll structure that contains a four-stranded rather than five-stranded β-sheet. These features may reduce the stability of the secondary structure and therefore reduce any energy barriers towards unfolding.

The proposition that the molecule unfolds upon interaction with the interface again has potential parallels to the proposed method of interaction of BPI with LPS-containing lipid membranes. As discussed above, the loops in the N-terminal domain of BPI that correspond to those that have been proposed here to facilitate latherin's adsorption into the air-water interface are believed to mediate BPI's interaction with LPS [89,90]. Furthermore, biophysical data that suggests that the N-terminal domain of BPI unfolds upon interaction with LPS-containing membranes has been reported [92]. This observation corresponds well with the mechanism proposed here for latherin.

The latherin structure was analysed for any indication of how the protein might unfold to achieve an amphiphilic form. Two simplified classes of unfolding can be proposed involving retention of secondary structure or its loss. In reality, any intermediate between these two extremes may arise, however for simplicity, these two opposing models were considered. If all secondary structure were relinquished, any number of unfolding processes may occur, with residues only restrained by their covalent bonds to their neighbours and their preference for the hydrophilic or hydrophobic environments. One could imagine unfolding spreading down the protein from the hydrophobic loops as the molecule is drawn into the interface.

Alternatively, if we assume that the secondary structure (or the majority of it, at least) is retained at the interface, the molecule would require some mode of unfolding that did not disrupt the regular structure elements. From the calculated structure, one would assume the four stranded β-sheet to be relatively stable as a consequence of the cooperative hydrogen bonding between its strands. This assumption is reinforced by the H/D-exchange data that identifies the inter-strand H-bonds, in particular those between strands 2 and 3, as the most

stable in the molecule. This leaves three obvious “seams” between the secondary structure features where the molecule could “unstitch”: between the two helices; between helix A and strand 1; and between helix B and strand 4. Unstitching between helix B and strand 4 seems unlikely as these two features are covalently constrained by a disulphide bond, whereas in the other two proposed seams only hydrophobic interactions maintain the packing between the secondary structure features. An analysis of the interfacial surface area in these two areas identifies the interface between the two helices as the greatest with a solvent excluded surface area of approximately 2000 $Å^2$ compared to the seam between helix A and strand B which has a surface area of approximately 1300 $Å^2$. The two interfaces appear to be well packed showing no signs of potential weakness apart from one small cavity between helix A and strand 1, in the vicinity of the conserved β-bulge. These observations would seem to identify the intersection between helix A and strand 1 as the more likely break in the structure. However, this hypothesis is based upon assumptions with regards to the effect of hydrophobic interactions at the interface where, in the absence of polar solvent, the effect is known to be significantly reduced [252,253].

Based upon the hypothesis that the hydrophobic loops are the feature that recognises the interface, any unfolding seems likely to initiate from this end of the protein. Strand 1 and helix A at this end are linked by one of the hydrophobic loops that would prevent the two features reorienting independently of one another, whereas the two sections of α-helix, which are at opposing ends of the protein sequence would be free to unfold away from one another relatively freely. It is the two loops that connect the sections of helix to the β-sheet that display chemical exchange and these two loops are by far the longest in the protein. These observations may indicate that these loops allow the two α-helices an additional degree of freedom relative to the remaining structured part of the molecule.

Therefore, based upon the calculated latherin structure, and the experimental data accrued thus far, the most likely process of unfolding at the air water interface is along the seam between the two helices. A model of the hypothesised process is shown in Figure 9.2.

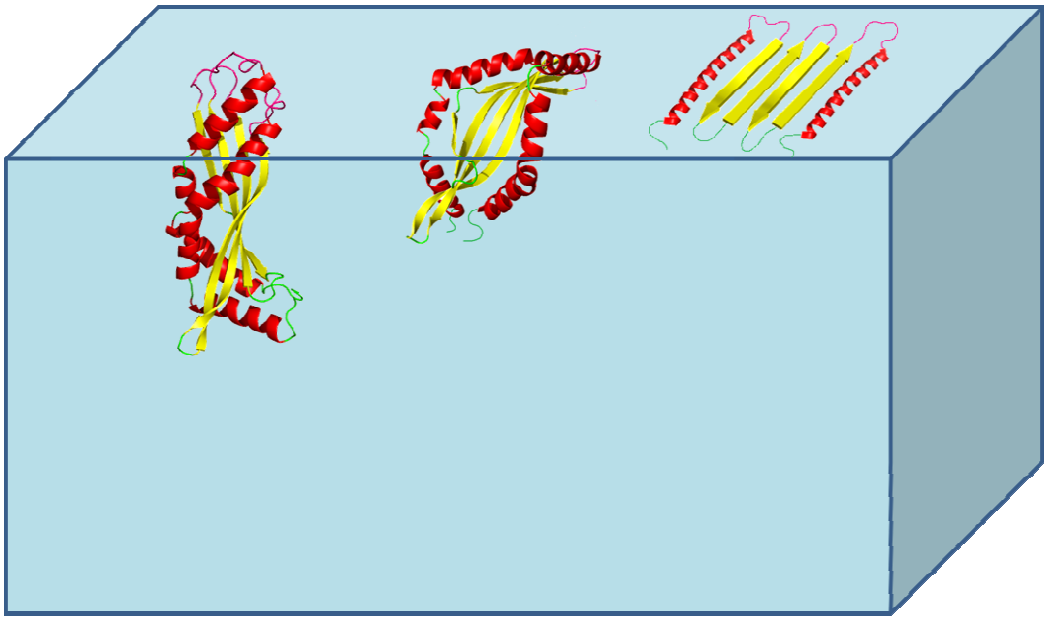


Figure 9.2- Model of latherin unfolding at the air-water interface. The latherin structure is coloured by secondary structure: red = α-helix, yellow = β-sheet, green = other and the hydrophobic loops are shown in pink. The model shows three stages from left to right: recognition of the interface via the hydrophobic loops, initial unfolding between the two sections of α-helix, initiated from the loops, and an open, planar, conformation of latherin, with its hydrophobic core exposed at the interface.

The models outlined in this and previous work [62], remain purely hypothetical extrapolations of the data accumulated thus far. To advance understanding of the behaviour of these surfactant proteins, more experimental insight into the form of these molecules at the interface is required

9.3 Probing the Molecular Basis of Surfactant Activity

Ideally, high resolution structural data of the unfolded, active forms of the proteins would be able to confirm the validity of the proposed models. However, as discussed elsewhere, methods to probe protein structure at interfaces do not currently exist. One possible approach may be to record NMR spectra from foams or emulsions created using the surfactant proteins. In these materials, the majority, if not all the protein would be in the interface and thus any information with regards to the structure that could be extracted, would provide insight into the active conformation. There are however many obstacles to this approach. To achieve a sufficient concentration of protein in the interface, the foam/emulsion would be required to consist of high number extremely small volumes of dispersant, requiring close control over dispersion creation and stability. Producing a sufficiently homogenous magnetic field throughout the sample in the presence of varying magnetic susceptibilities between the bulk solution and the dispersant may also prove a challenge. A potential solution may be to use substances with matched magnetic susceptibilities. For example, foam produced from nitrogen (magnetic susceptibility, χ_m = -12.4 x 10^{-6} cm^3mol^{-1} [254]) in water (χ_m = -12.96 x 10^{-6} cm^3mol^{-1} [255]) would be a better match than air (-0.37 x 10^{-6} cm^3mol^{-1} [256]) in water. Another issue would be the stability of the foam. Unlike, the natural foam nests, foams created from Rsn-2 collapse after several minutes. This may be problematic as NMR experiments often require samples stable over several hours or days, dependent upon protein concentration, to achieve sufficient sensitivity.

Another potential method may be to utilise the proteins' affinity for hydrophobic surfaces [73,130]. The form of the protein adsorbed to a hydrophobic surface would presumably be identical, or least similar, to that adsorbed into the air-water interface. An advantage of this approach is that the active form could be isolated from the solution form, resulting in a stable, ordered, homogenous sample. The sample could obviously not be analysed by X-ray crystallography or solution state NMR. However, recent advances in the application of solid state NMR to protein structure determination, presents this technique as a potential method to yield high resolution structural data from the bound protein [257].

In the absence of any applicable high resolution structural technique, biophysical analysis and indirect methods could be employed to investigate various aspects of the model. Mutagenesis could be utilised to probe the roles of the amphiphilic/hydrophobic features proposed to be responsible for surface adsorption in both proteins. In Rsn-2 more precise

investigation into the role of the aliphatic residues in the N-terminus could be carried out by site directed mutagenesis. A similar approach could be applied to investigate the activity of the hydrophobic loops in latherin. A potentially interesting experiment would be to mutate hydrophobic residues into the corresponding region of BPI and investigate any effects upon its adsorption into the interface. Additionally any role for the lone tryptophan could also be investigated using site-directed mutagenesis.

Further analysis of the effects of the insertion of disulphides into specific locations of Rsn-2 could also be carried out. By using different combinations of disulphides to inhibit particular potential unfolding processes, a greater understanding of any process may be achieved. A similar approach could be applied to the proposed latherin model whereby the molecule unfolds along a seam between the two α-helices. The validity of this model could be investigated by covalently linking the two helices and investigating how inhibition of their ability to reorient independently of one another affects the protein's surfactant activity. The close proximity of the two helices presents a number of locations where disulphide bonds could be introduced with minimal effects upon the secondary structure. Similarly, disulphide bonds could be introduced between other secondary structure features such as helix A and strand 1 to probe alternative unfolding processes.

IRRAS of Rsn-2 at the air-water interface has shown that the proportions of secondary structure are retained. For latherin two extreme cases where the molecule may either retain or relinquish its secondary structure features have been proposed. IRRAS data could reveal the proportions of regular secondary structure at the interface, identifying which of these models is the more applicable.

Another potential approach could be to utilise chemical cross-linking to extract information with regards to form of the protein at the interface. For example, amide to amide cross-linkers of various lengths of spacer arm could be used to probe the intra and inter-molecular distances of the four lysine residues. A potentially attractive approach may be to use photo-reactive chemical cross-linkers with the aim of limiting cross-linking to the interface.

10 Methods and Materials

10.1 Molecular Biology

10.1.1 Plasmid Purification

Plasmid DNA was purified from *E.coli* cultures using a QIAprep Spin Miniprep Kit (Qiagen, #27104). Purifications were carried out following the supplied protocol. DNA was eluted in sterile, deionised water.

10.1.2 DNA Quantification

DNA solutions were quantified by measuring the absorbance at 260 nm (A_{260}) using the nanodrop spectrophotometer (Thermo Scientific). Concentration was calculated based upon the equation DNA Concentration (µg/mL) = A260/0.02 [258].

10.1.3 DNA Electrophoresis

DNA electrophoresis was carried out using TAE, 1% agarose gels. Electrophoresis grade agarose (Invtirogen, #15510-027) was added to TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH= 8.0). The solution was then microwaved until all agarose had melted and was then left until it was cool enough to handle. At this point ethidium bromide (Promega, #H5041) was added to a final concentration of 100 µg/mL and the gel was cast and left to set. The gel was then submerged in TAE buffer.

Samples were prepared by adding 1 part Orange G loading dye (10 mM Orange G, 200 mM Tris, 100 mM acetic acid, 5 mM EDTA, 50% glycerol (v/v)) to 3 parts DNA sample and were loaded onto the agarose gel alongside DNA molecular weight markers (NEB, #N3232S). Electrophoresis was carried out at 100V for 40 min. Gels were then inspected under UV illumination.

10.1.4 DNA Gel Extraction

DNA was extracted from agarose gels using a QIAquick Gel Extraction kit (Qiagen, #28704). Extractions were carried out following the supplied protocol. DNA was eluted in sterile, deionised water.

10.1.5 Subcloning

10.1.5.1 PCR

PCR reactions were carried out using Pfu Turbo DNA polymerase (Stratagene, # 600250). Reactions were set up following the supplied protocol. A typical reaction was set up as follows:

Sterile, deionised water	40.6 µL
10x Pfu reaction buffer	5 µL
dNTPs (25 mM each dNTP)	0.4 µL
DNA template (100 ng/µL)	1 µL
Primer #1 (100 ng/µL)	1 µL
Primer #2 (100 ng/µL)	1 µL
Pfu Turbo DNA polymerase (2.5 U/µL)	1 µL
Final volume	50 µL

A typical PCR cycling programme was set up as follows:

Segment	Number of cycles	Temperature	Duration
1	1	95°C	2 min
2	30	95°C	30 sec
		Primer T_m - 5°C	30 sec
		72°C	1 min
3	1	72°C	10 min

Results of PCR were analysed by DNA electrophoresis.

10.1.5.2 Attaching A-tail Overhangs

A-tail overhangs were attached to blunt-ended PCR products using GoTaq Flexi DNA Polymerase (Promega, #M8301). The A-tail reaction was based upon the supplied protocol. A typical reaction was set up as follows:

PCR product	6 µL
5x GoTaq Flexi Buffer	1 µL
$MgCl_2$ solution (25 mM)	1 µL
dATP (25 mM)	1 µL
GoTaq DNA Polymerase (5 U/µL)	1 µL
Final volume	10 µL

10.1.5.3 TOPO Cloning

TOPO cloning reactions were set up following the supplied protocol. A typical reaction was set up as follows:

A-tailed PCR product	4 µL
Salt solution	1 µL
TOPO vector	1 µL
Final volume	6 µL

Reactions were left for 30 mins on the bench at room temperature. 2 µL of TOPO cloning reaction was then transformed into TOP10 competent cells (Invitrogen, #C4040-50). The transformations were plated on LB agar containing the relevant selection antibiotic and X-Gal (Promega, #V394A) to allow for blue/white screening.

10.1.5.4 Restriction Digests

Restriction digests were set-up following supplied protocols. A typical restriction digest was set-up as follows:

Sterile, deionised water	15.8 µL
10x RE buffer	2 µL
Acetylated BSA, 10 µg/mL	0.2 µL
DNA, 1 µg/mL	1 µL
Restriction Enzyme 10 U/µL	1 µL
Final volume	20 µL

Digests were then left at 37°C for 4 h with gentle mixing after 2 h. Restriction enzymes were then analysed by DNA electrophoresis .

10.1.5.5 Dephosphorylation

Dephosphorylation of vector DNA was carried out using Calf intestine alkaline phosphotase (CIAP) (Promega, # M182A). The reaction was set-up following the supplied protocol. A typical reaction was set up as follows:

DNA	44 µL
10x CIAP reaction buffer	5 µL
Diluted CIAP (0.01 U/pmol of DNA ends)	1 µL
Final volume	50 µL

The concentration of DNA ends was calculated based upon an average molecular weight of a nucleotide pair of 660pg/pmol. The reaction was left at 37°C for 30 min, then a second identical amount of diluted CIAP was added and left for a further 30 min at 37°C. The CIAP was then removed by DNA electrophoresis.

10.1.5.6 Ligation

Ligation reactions were carried out using T4 DNA ligase (Promega, #M180A). Molar ratios of 1:0 (negative control to determine background), 1:2 and 1:10 vector:insert were used with 50ng of RE digested destination vector. A typical reaction was set up as follows:

Sterile, deionised water	2 µL
Vector DNA (50 ng/µL)	1 µL
Insert DNA (5 ng/µL)	5 µL
10x Ligase buffer	1 µL
T4 DNA Ligase	1 µL
Final volume	10 µL

Reactions were left overnight at 4°C then transformed into subcloning efficiency DH5α competent cells (Invitrogen, #18265-017).

10.1.6 Site-Directed Mutagenesis

Site-directed mutagenesis was carried out using a QuikChange Site-Directed Mutagenesis Kit (Stratagene, #200518).

10.1.6.1 Primer Design

Primers were designed in accordance with the supplied guidelines. In brief: primers were designed to contain the desired mutation in the middle of the primer, anneal to the same sequence on opposite strands of the plasmid, have a melting temperature, T_m, > 78°C and have a GC content > 40%.

10.1.6.2 Procedure

Reactions were set up following the supplied protocol. A typical reaction was set up as follows:

Sterile, deionised water	40 µL
10x reaction buffer	5 µL
dNTP mix	1 µL
DNA template (50 ng/µL)	1 µL
Primer #1 (125 ng/µL)	1 µL
Primer #2 (125 ng/µL)	1 µL
Pfu Turbo DNA polymerase (2.5 U/µL)	1 µL
Final volume	50 µL

A typical PCR cycling programme was set up as follows:

Segment	Number of cycles	Temperature	Duration
1	1	95°C	30 min
2	16	95°C	30 sec
		55°C	1 min
		68°C	5 min

On completion of temperature cycling, samples were cooled to below 37°C on ice. Template DNA was then digested by adding 1 µL DpnI (10 U/µL) to each reaction. The reaction was then left at 37°C for 1 h with gentle mixing after 30 min. 1 µL of the digested DNA was then transformed into 50 µL of the supplied XL1-Blue Supercompetent cells.

10.1.7 Transformations

Transformations were carried out in accordance with the protocol supplied with the competent cells. Briefly, a typical transformation was carried out as follows: 20 µL of competent cells were taken from -85°C storage and defrosted on ice. 10 ng of plasmid DNA was then added to the competent cells and mixed by gentle stirring. The transformation mixture was then stored on ice for 5 mins, before heating at 42°C for 30 sec in a water bath. The tubes were again stored on ice for 2 min, and then 80 µL room temperature SOC medium was added to the reaction which was then stored on its side at 37°C with 250rpm shaking for 1 h. Transformations were then plated on LB agar containing the appropriate selection antibiotic and stored at 37°C overnight.

10.1.8 Recipes

10.1.8.1 LB Agar Plates

For 500 mL weigh out:

Tryptone	5 g
Yeast extract	2.5 g
NaCl	5 g
Micro agar	6 g

Dissolve in 450 mL ddH_2O.
Adjust pH to 7.5.
Make up to 500 mL with ddH_2O.
Autoclave.

10.1.8.2 SOC Medium

For 200 mL weigh out:

Tryptone	4 g
Yeast extract	1 g
NaCl	0.1 g
KCl	40 mg

Dissolve in 180 mL ddH_2O.
Adjust pH to 7.0.
Make up to 200 mL with ddH_2O.
Autoclave.
Once medium is cool add:

Glucose	720 mg
MgCl2	190 mg

10.2 Protein Expression and Purification

10.2.1 Protein Over-Expression

Expression media was treated with selective antibiotics to recommended concentrations (Ref, Novagen pET manual), warmed to 37°C and inoculated with a 100x dilution of high density, overnight culture. The cultures were then left to grow at 37°C, with 200rpm shaking until reaching an OD600 ≈ 0.8 at which point expression was induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma, #MB1003) at the desired concentration. The cells were then left to express at the selected conditions.

10.2.2 Cell Lysis

10.2.2.1 Detergent lysis

A lysis stock solution of 90% binding buffer, 10% 10X Bugbuster solution (Novagen, #70921) and 0.01% Benzonase nuclease (Novagen, #70746-3) was produced. Cell pellets were resuspended in this solution to a desired concentration and left to lyse for 20 min. Soluble and insoluble fractions were separated by centrifugation at 10000 xg for 1min.

10.2.2.2 Sonication

Cells were resuspended in binding buffer containing 0.01% Benzonase (Novagen, #70746-3). Sonication was carried out on a Sanyo Soniprep 150 at 10 amplitude microns for 30 cycles of 30 sec on, 30 sec off. An ice bath was used to prevent sample over-heating. Soluble and insoluble fractions were separated by centrifugation at 40000 xg for 20 min.

10.2.3 Ni-affinity Chromatography

Soluble cell lysates were passed through a 0.8µm syringe filter (Nalgene, #190-2580) to remove any remaining insoluble cell debris. The lysate was then passed through 5mL Ni-NTA resin (Novagen, #N75212), and washed with 40mL Binding buffer, 40mL Wash buffer and finally bound proteins were eluted with 40mL of elution buffer.

Binding buffer	20 mM Tris, 500 mM NaCl, 5 mM Imidazole, 1 mM NaN_3, pH= 8.0.
Wash buffer	20 mM Tris, 500 mM NaCl, 30 mM Imidazole, 1 mM NaN_3, pH= 8.0.
Elution buffer	20 mM Tris, 500 mM NaCl, 300 mM Imidazole, 1 mM NaN_3, pH= 8.0.

10.2.4 Fusion Tag Cleavage

10.2.4.1 Enterokinase

Enterokinase cleavage of fusion tags was carried out using Tag-off High Activity rEk (Novagen, #71537-3). A typical reaction was set-up as follows:

Protein solution	4500 µL
10x rEK cleavage buffer	500 µL
Tag-off high activity rEK (1 U/µL)	5 µL
Final volume	5005 µL

The reaction was left at 4°C for 48 h to improve specificity. The cleavage was then investigated by SDS-PAGE. If cleavage was not complete a further 5 µL (1U/µL) Tag-off high activity rEK was added and the reaction left at 4°C for a further 48 h. This step was repeated until cleavage was complete.

10.2.4.2 Thrombin

Thrombin cleavage of fusion tags was carried out using restriction grade thrombin (Novagen, #69671-3). A typical reaction was set-up as follows:

Protein solution	4500 µL
10x Thrombin cleavage buffer	500 µL
Restriction grade thrombin (1 U/µL)	5 µL
Final volume	5005 µL

The reaction was left at room temperature (≈ 22°C) for 48 h. The cleavage was then investigated by SDS-PAGE. If cleavage was not complete a further 5 µL (1U/µL) restriction grade thrombin was added and the reaction left at room temperature for a further 48 h. This step was repeated until cleavage was complete.

10.2.4.3 HRV 3C

HRV 3C cleavage of fusion tags was carried out using HRV 3V Protease (Novagen, #71493-3). A typical reaction was set-up as follows:

Protein solution	900 µL
10x HRV 3C protease cleavage buffer	100 µL
HRV 3C protease (1 U/µL)	5 µL
Final volume	1005 µL

The reaction was left at 4°C for 48 hours. The cleavage was then investigated by SDS-PAGE. If cleavage was not complete a further 5 µL (1U/µL) HRV 3C protease was added and the reaction left at 4°C for a further 48 hours. This step was repeated until cleavage was complete.

10.2.5 Size Exclusion Chromatography

Size exclusion chromatography was carried out on a pre-packed Superdex 75 10/300 GL column (GE healthcare, #17-5174-01), connected to the AKTA chromatography system housed in a cold room. After equilibration with 30 mL of running buffer, a 500 µL protein sample was injected into the sample loop. Size exclusion chromatography was then carried out at 0.5 mL/min running buffer for 1.5 column volumes (35.34 mL) and was monitored by the UV absorbance at 280 nm. Fractions of 0.5 mL elutant were collected and fractions that showed absorbance at 280 nm were analysed by SDS-PAGE.

Running buffer	20 mM NaPi, 50 mM NaCl, 1 mM NaN_3, pH= 7.5. De-gassed before use.

10.2.6 Protein Sample Concentration

Protein samples were concentrated using a Vivaspin20 5000 MWCO (molecular weight cut-off) column (Vivascience, #VS2012). Columns were first pre-rinsed with ddH_2O to remove any trace contaminants. Protein samples were centrifuged at 5000 x g until the desired concentration was achieved.

10.2.7 Recipes

10.2.7.1 LB Media

For 1L weigh out:

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

Dissolve in 950 mL ddH_2O.
Adjust pH to 7.5.
Make up to 1L with ddH_2O.
Autoclave.

10.2.7.2 M9 Minimal Media

For 1L M9 minimal media [259]:

dd H_2O	800 mL
5x M9 salts solution	200 mL
Salt mix	1 mL
$MgSO_4$ (1 M)	2 mL
$CaCl_2$ (50 mM)	2 mL
Thiamine (50 mg/mL)	800 µL
D-glucose	3 g
NH_4Cl	1 g

For ^{15}N-labelled media, replace NH_4Cl with $^{15}NH_4Cl$ (98% ^{15}N) (Isotec, #299251).

For 15N, 13C-labelled media, replace NH_4Cl with $^{15}NH_4Cl$ (98% ^{15}N) (Isotec, #299251) and D-glucose with D-Glucose (99% ^{13}C) (Isotec, # 389374).

For 500 mL 5x M9 salts solution:

Na_2HPO_4	34.08 g
KH_2PO_4	15 g
NaCl	2.5 g

Dissolve in 450 mL ddH_2O.
Adjust pH to 7.5.
Make up to 500 mL with ddH_2O.
Autoclave.

For 100 mL Salt mix:

$ZnSO_4$	115 mg
$MnSO_4$	22.3 mg
H_3BO_3	4.3 mg
$CuSO_4$	11.2 mg

Dissolve in 100 mL ddH_2O.

10.3 Protein Characterisation

10.3.1 UV/Vis Absorbance

Extinction co-efficients were calculated from amino acid sequence's [260]. Routinely this was carried out using the Protparam programme (http://web.expasy.org/protparam/). A baseline was calculated using the relevant buffer. The A_{280} absorbance of protein solutions were then measured using either a NanoDrop 1000 (Thermo Scientific) or a JASCO V-550 spectrophotometer. It was observed that the V-550 gave more accurate results and therefore this spectrometer was used when a high level of accuracy was required. The Beer-lambert law was used to calculate the protein concentration from A_{280} absorbance and the calculated extinction coefficient.

10.3.2 SDS-PAGE

Protein electrophoresis was carried out using Invitrogen's NuPAGE system. Loading buffer was prepared by adding Ditrioethol (DTT) (Melford, #MB1015) to 4x LDS loading buffer (Invitrogen, #NP0007) to a final concentration of 0.1M in all gels unless described as under non-reducing conditions, where the loading buffer was not modified. Samples were added to loading buffer at a ratio of 3:1 and heat at 85°C for 5 min to aid unfolding. Samples were then loaded onto NuPAGE 4-12% Bis Tris gels (Invitrogen, #NP0323BOX) in 1x MES running buffer (Invitrogen, #NP0002) and the gel was ran at 200v for 35min. Gels were then stained with coomassie stain and destained with destain solution.

Coomassie Stain	45% methanol, 45% ddH_2O, 10% acetic acid, 0.25% (w/v) coomassie R250.
Destain Solution	5% methanol, 10% acetic acid, 85% ddH_2O.

10.3.3 Western Blot

Preceding western blotting, protein samples were separated by mass by SDS-PAGE. Pres-stained protein markers (NEB, #P7708S) were utilised to allow for estimation of molecular weight following western blotting. The protein bands from the unstained gels were then transferred to nitrocellulose membrane (Invitrogen, #LC2000) using a XCell II Blot Module (Invitrogen, EI9051) following the supplied protocol. Western transfer was carried out in transfer buffer at a constant voltage of 30 V for 1 h.

Immunoblotting of the nitrocellulose membrane was carried out using monoclonal anti-polyhistidine-alkaline phosphatase, clone HIS-1 (Sigma, #A5588) following the supplied protocol. The membrane was then treated with 5-bromo-4-chloro-3-indolyl phosphate (BCIP)/nitro blue tetrazolium (NBT) liquid substrate (Sigma, #B1911) for detection of bound antibody.

Transfer Buffer	5% NuPAGE Transfer Buffer (20X) (Invitrogen, #NP0006), 10% Methanol, 85% ddH_2O.

10.3.3.1 Experimental Estimation of Stokes Radius

The Stokes radii of macromolecules were calculated based upon their rate of progression through a pre-packed Superdex 75 10/300 GL column (GE healthcare, #17-5174-01) via a standard curve. The void volume, V_0, of the column was calculated using Blue Dextran 2000. A geometric column volume, V_c, of 24.0 mL from the supplied documentation was used. The K_{av}'s calculated from the V_e of various protein standards from a LMW calibration kit (GE healthcare, #28-4038-41) and the Stokes radii from the supplied documentation were used to create a calibration curve (Figure 10.1). The formula of the best fit line was then used to calculate the Stokes radii from K_{av}.

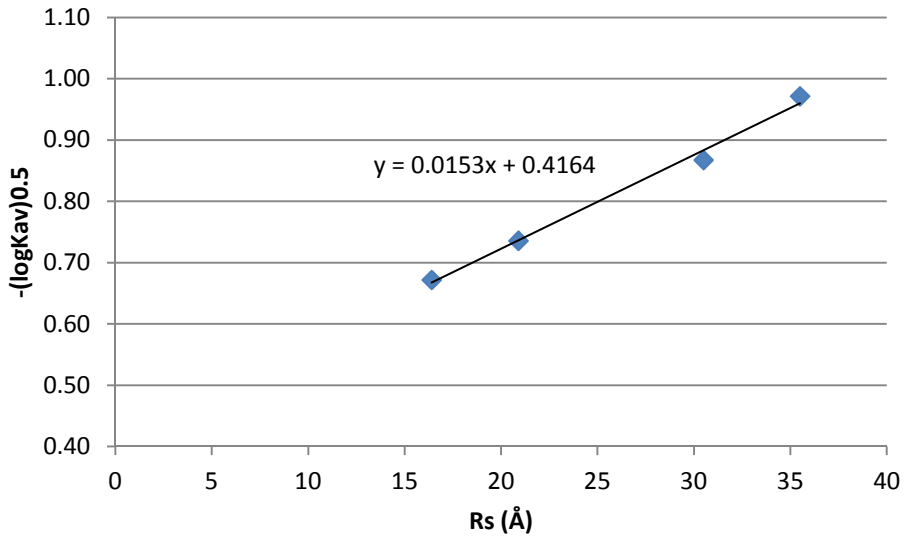


Figure 10.1- Size exclusion chromatography calibration curve. Standard proteins used were: Ribonuclease A (16.4 Å), Chymotrypsinogen (20.9 Å), Ovalalbumin (30.5 Å), and Albumin (35.5 Å).

10.3.4 Ellman Assay

A standard curve was created by measuring the absorbance from Ellman's reagent with cysteine at various concentrations. A stock solution of 1.5 mM cysteine was created by dissolving 26.34 mg of cysteine hydrochloride monohydrate (Sigma, #C7880) in 100 mL of reaction buffer (0.1 M sodium phosphate, 1 mM EDTA, pH 8.0). Solutions at concentrations of 0.10, 0.25, 0.50, 0.75, 1.00, 1.25 and 1.50 mM as well as a blank containing no cysteine, were created from dilutions of the stock solution and the reaction buffer. The Ellman's reagent solution was prepared by dissolving 4 mg of Ellman's reagent (Pierce, #69-78-3) in 1 mL of reaction buffer. 250 µL of each standard was added to 2.5 mL reaction buffer and 50 µL Ellman's reagent solution, mixed and incubated at room temperature for 15 min. The absorbance at 412 nm was then measured and plotted against cysteine concentration to produce the standard curve (Figure 10.2).

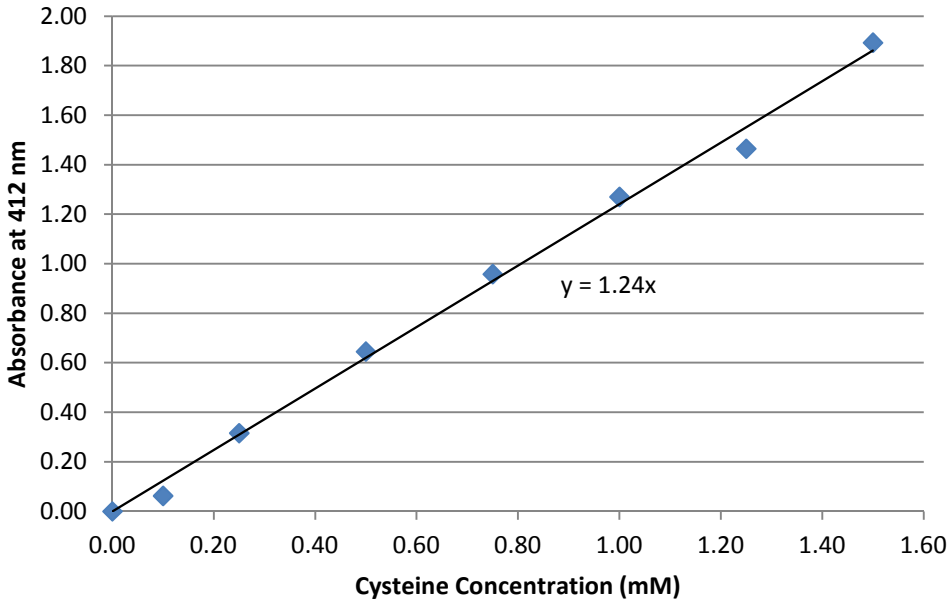


Figure 10.2- Standard curve for the determination of free cysteine concentration by the Ellman assay.

To measure the cysteine concentration of protein solutions, 250 µL protein solution was added to 2.5 mL reaction buffer and 50 µL Ellman's reagent solution, mixed and incubated at room temperature for 15 min. The absorbance at 412 nm was then measured and used to calculate cysteine content from the standard curve.

10.4 Surfactant Assays

10.4.1 Bikerman Assay

The Bikerman assay was carried out by dispensing 2.0 mL of 100 µg/mL protein solution in a 100 mL graduated flask. Air was then bubbled through the solution via a glass pipette (for improved reproducibility and ease of cleaning) at a constant rate of 0.5 mL/min using a peristaltic pump. The assay was carried out until steady-state equilibrium, the point at which the height of the foam column remains constant with time, was achieved. The height of the foam column was measured in mL and the 2 mL of solution of the bottom of the flask subtracted. The measuring cylinder was cleaned thoroughly with ddH_2O between samples. This measurement was carried out in triplicate for each sample to allow for the calculation of the standard deviation.

10.4.2 Du Noüy Ring

The Du Noüy Ring surface tension measurements were carried out on a classical torsion balance (White Electrical Instrument Company) with a platinum Du Noüy Ring. The balance was zeroed against air. 1 mL samples were placed on a glass evaporating dish and allowed to equilibrate for five minutes before measurement. The ring was then lowered into the surface of the solution and then raised up again until the point at which it broke free of the surface. The force measured was recorded and the measurement repeated in triplicate to allow for the calculation of the standard deviation. Between samples the ring was flamed in a Bunsen and the glass evaporating dish washed with ethanol and dried thoroughly to prevent sample contamination.

10.4.3 MicroTrough Tensiometer

The MicroTrough tensiometer surface tension measurements were carried out on a Kibron Microtrough S (Kibron, Espoo, Finland). The instrument was calibrated against air and pure water as per the manufacturer's instructions and then zeroed against buffer. 500 µL samples were placed on a multi-well plate immediately before measurement. The probe was lowered into solution to ensure complete wetting, raised up and lowered back down until in contact with the surface. The measurement was then initiated. The time taken from sample deposition to the start of the measurement was normally less than 10 s. Measurements where excessive delays (>10 seconds) between deposition and measurement occurred were abandoned. Measurements were left to run for 1000 seconds. Between

samples the probe was flamed in a Bunsen to prevent sample contamination. When required the multi-well plate was washed with ethanol then ddH_2O and dried thoroughly.

10.5 Lipid Vesicle Production

Lipid vesicles were produced with the assistance of Dr. G. Brand. Throughout the production process care was taken to minimise exposure to light to prevent oxidation of the phospholipids. Fresh vesicles were produced before each set of experiments.

Dimyristoylphosphatidylcholine (DMPC) (#850345) and dimyristoylphosphatidylglycerol (DMPG) (#840445) lipids were sourced from Avanti Polar Lipids. 20 µmoles of the required lipid was dissolved in a few mL of 50 % chloroform/ 50% methanol (v/v). The solvent was then removed using a rotary evaporator at room temperature leaving a thin film of lipid on the sides of the flask. The lipid was then resuspended in 1 mL 20 mM NaPi, 150 mM NaCl, pH 7.5, to give a final concentration of 20 mM. Once all lipid had been dissolved, the resultant vesicles were repeatedly passed through a 100 nm polycarbonate filter to ensure reasonable homogeneity of vesicle size.

Appendices

Appendix A- Latherin Chemical Shift Assignments

All experiments carried out at 310 K in 20 mM NaPi, 50 mM NaCl, 1 mM NaN_3, pH 7.5.

Residue	H	N	ND	NE	HA	HB	HG	HD	HE	HZ	HH	C	CA	CB	CG	CD	CE	CZ	CH
0 Ala	-	-	-	-	4.31	1.39	-	-	-	-	-	172.29	52.77	19.28	-	-	-	-	-
1 Gln	8.33	119.16	-	-	4.37	2.14 1.97	2.33 2.33	-	-	-	-	175.62	56.00	29.82	34.02	-	-	-	-
2 Gln	8.18	120.15	-	-	4.50	2.11 1.97	2.36 2.36	-	-	-	-	175.82	55.60	30.22	33.94	-	-	-	-
3 Ile	8.42	124.59	-	-	4.20	1.61	1.20 0.53 0.72	0.17	-	-	-	174.02	59.16	38.56	28.40 16.86	12.82	-	-	-
4 Pro	-	-	-	-	4.69	2.57 1.89	2.14 2.14	4.08 3.43	-	-	-	-	62.13	31.79	28.01	51.50	-	-	-
5 Pro	-	-	-	-	4.38	2.35 1.98	2.04 2.12	3.88 3.88	-	-	-	178.59	65.04	31.94	-	51.13	-	-	-
6 Glu	9.56	118.14	-	-	4.14	2.00 2.00	2.30 2.30	-	-	-	-	178.46	59.54	28.41	36.39	-	-	-	-
7 Val	7.33	118.70	-	-	3.94	2.01	0.97 0.81	-	-	-	-	177.54	64.58	32.28	21.99 21.18	-	-	-	-
8 Ser	8.19	113.88	-	-	3.89	3.85 3.77	-	-	-	-	-	177.39	61.57	62.89	-	-	-	-	-
9 Ser	8.49	117.07	-	-	4.25	3.91 3.91	-	-	-	-	-	175.68	61.74	62.60	-	-	-	-	-
10 Gln	7.37	119.78	-	110.59	4.15	2.14 2.14	2.58 2.33	-	6.67 7.38	-	-	179.44	59.24	29.00	34.30	-	-	-	-
11 Ile	7.75	118.05	-	-	3.79	1.95	1.16 0.79 1.69	0.73	-	-	-	177.73	64.46	37.53	29.11 17.83	13.39	-	-	-
12 Thr	8.91	113.00	-	-	3.77	4.22	1.11	-	-	-	-	178.68	67.40	67.36	22.19	-	-	-	-
13 Asp	8.75	126.01	-	-	4.47	2.86 2.62	-	-	-	-	-	178.27	58.13	40.17	-	-	-	-	-
14 Ala	7.38	122.29	-	-	4.18	1.5	-	-	-	-	-	181.69	55.29	18.56	-	-	-	-	-

15 Leu	8.46	118.76	-	-	3.98	1.12 1.75	1.79	0.69 0.70	-	-	-	178.13	57.83	42.59	26.75	26.48 23.51	-	-	-
16 Thr	8.35	116.36	-	-	4.38	3.65	1.22	-	-	-	-	175.88	57.73	68.44	20.95	-	-	-	-
17 Gln	8.00	119.91	-	114.28	3.89	2.15 2.08	2.39 2.45	-	6.74 7.82	-	-	178.20	58.80	28.02	33.40	-	-	-	-
18 Gly	7.93	105.66	-	-	3.92 3.70	-	-	-	-	-	-	177.13	47.26	-	-	-	-	-	-
19 Leu	8.22	123.10	-	-	4.01	1.09 2.01	1.65	0.54 0.06	-	-	-	179.20	57.95	41.74	26.25	22.86 26.27	-	-	-
20 Leu	8.07	120.60	-	-	4.25	1.89 1.61	1.84	0.89 0.90	-	-	-	180.77	57.90	41.37	26.74	22.69 25.19	-	-	-
21 Asp	9.01	121.74	-	-	4.45	2.74 2.67	-	-	-	-	-	177.83	56.87	40.29	-	-	-	-	-
22 Gly	7.62	104.85	-	-	3.65 4.27	-	-	-	-	-	-	174.06	44.93	-	-	-	-	-	-
23 Asn	8.06	116.78	112.48	-	4.63	2.84 3.06	-	7.52 6.84	-	-	-	176.10	53.90	37.31	-	-	-	-	-
24 Phe	8.33	119.82	-	-	4.14	2.75 3.32	-	7.00	7.08	6.99	-	176.52	62.35	41.13	-	131.98	130.34	-	-
25 Leu	8.82	114.51	-	-	3.82	1.80 1.44	1.81	0.97 0.81	-	-	-	179.97	57.64	40.71	27.30	25.30 22.88	-	-	-
26 Ser	7.69	115.47	-	-	4.20	3.93 3.93	-	-	-	-	-	177.03	61.42	62.60	-	-	-	-	-
27 Leu	7.76	123.50	-	-	3.97	1.65 1.35	1.66	0.86 0.92	-	-	-	180.04	57.76	41.43	26.97	23.43 25.44	-	-	-
28 Leu	7.28	118.01	-	-	3.82	1.63 1.47	1.49	0.81 0.75	-	-	-	177.39	57.89	41.13	28.06	25.75 25.26	-	-	-
29 Asn	7.50	113.46	-	-	4.48	2.88 2.73	-	-	-	-	-	175.28	55.14	38.89	-	-	-	-	-
30 Ala	7.31	120.33	-	-	4.48	1.43	-	-	-	-	-	177.32	51.76	19.46	-	-	-	-	-
31 Ile	7.01	120.65	-	-	3.93	1.92	1.61 0.95 1.45	0.77	-	-	-	175.29	61.31	38.01	17.24 27.67	12.91	-	-	-
32 Asn	8.57	125.52	-	-	4.60	2.89 2.98	-	-	-	-	-	176.71	54.25	38.93	-	-	-	-	-
33 Leu	8.91	127.17	-	-	4.37	1.59 1.53	1.76	0.53 0.79	-	-	-	177.14	54.65	42.95	26.46	26.39 23.56	-	-	-
34 Glu	8.29	120.83	-	-	3.79	2.12 2.00	2.33 2.24	-	-	-	-	178.56	60.25	29.67	36.74	-	-	-	-
35 Gly	8.35	107.97	-	-	3.83 3.83	-	-	-	-	-	-	176.60	46.96	-	-	-	-	-	-

36 Leu	7.55	125.27	-	-	4.11	1.51 0.97	1.07	0.68 0.07	-	-	-	177.90	57.45	40.48	26.60	25.07 22.96	-	-	-
37 Leu	8.01	118.84	-	-	3.71	1.29 1.91	1.44	0.75 0.64	-	-	-	177.81	58.24	41.09	26.45	25.50 22.89	-	-	-
38 Asn	8.20	116.21	111.89	-	4.33	2.83 2.83	-	7.44 6.88	-	-	-	177.74	56.66	38.21	-	-	-	-	-
39 Thr	8.15	117.87	-	-	4.04	4.34	1.24	-	-	-	-	176.35	66.78	68.79	21.27	-	-	-	-
40 Ile	8.17	121.04	-	-	3.37	1.84	0.73 0.61 1.80	0.62	-	-	-	177.80	66.53	37.88	30.20 18.48	13.65	-	-	-
41 Leu	8.92	119.85	-	-	3.85	1.53 1.83	1.64	0.82 0.81	-	-	-	179.96	58.82	41.84	26.99	24.05 25.16	-	-	-
42 Asp	8.29	120.58	-	-	4.33	2.71 2.87	-	-	-	-	-	178.76	57.72	41.11	-	-	-	-	-
43 Gln	7.83	116.54	-	110.85	4.17	1.97 2.09	2.64 2.41	-	6.69 7.24	-	-	179.36	58.54	29.78	34.60	-	-	-	-
44 Val	8.89	116.31	-	-	3.92	2.17	0.95 0.88	-	-	-	-	177.47	65.63	31.95	21.81 22.42	-	-	-	-
45 Thr	8.22	115.57	-	-	3.72	4.26	1.22	-	-	-	-	176.77	67.58	68.50	21.62	-	-	-	-
46 Gly	7.78	107.61	-	-	3.89 3.89	-	-	-	-	-	-	176.08	47.03	-	-	-	-	-	-
47 Leu	7.46	121.06	-	-	4.21	1.81 1.49	1.73	0.87 0.92	-	-	-	179.06	57.26	42.87	27.06	23.84 25.22	-	-	-
48 Leu	8.07	117.22	-	-	4.15	1.48 1.79	1.61	0.80 0.79	-	-	-	177.62	56.42	41.91	-	25.98 22.67	-	-	-
49 Asn	8.15	117.74	-	-	4.45	2.83 2.82	-	-	-	-	-	175.86	55.16	38.20	-	-	-	-	-
50 Ile	7.42	116.68	-	-	4.03	1.83	1.50 1.16 0.86	0.84	-	-	-	176.43	62.05	39.05	27.68 17.44	13.18	-	-	-
51 Leu	7.97	122.04	-	-	4.30	1.63 1.52	1.67	0.82 0.82	-	-	-	177.46	56.28	43.32	26.89	23.28 23.28	-	-	-
52 Val	8.24	117.01	-	-	4.20	2.17	0.88 0.88	-	-	-	-	176.05	62.41	33.21	20.63 21.39	-	-	-	-
53 Gly	7.87	111.18	-	-	4.21 3.95	-	-	-	-	-	-	171.31	44.93	-	-	-	-	-	-
54 Pro	-	-	-	-	4.35	2.19 1.82	1.97 1.97	3.57 3.57	-	-	-	176.99	63.34	32.00	27.21	49.71	-	-	-
55 Leu	8.13	121.99	-	-	4.37	1.53 1.62	1.62	0.82 0.88	-	-	-	176.83	54.85	42.31	26.86	25.42 23.84	-	-	-
56 Leu	8.09	123.25	-	-	4.38	1.57 1.57	1.56	0.80 0.86	-	-	-	177.03	54.88	42.70	26.94	25.25 23.53	-	-	-

57 Gly	8.15	109.29	-	-	3.95 4.08	-	-	-	-	-	-	171.87	44.52	-	-	-	-	-	-
58 Pro	-	-	-	-	-	-	-	3.58 3.58	-	-	-	-	-	-	-	49.81	-	-	-
59 Ser	-	-	-	-	4.41	3.80 3.80	-	-	-	-	-	174.08	58.32	63.90	-	-	-	-	-
60 Asp	8.00	121.59	-	-	4.66	2.52 2.62	-	-	-	-	-	175.04	54.35	41.77	-	-	-	-	-
61 Ala	7.90	122.66	-	-	4.70	1.39	-	-	-	-	-	176.17	52.04	21.16	-	-	-	-	-
62 Glu	8.75	119.12	-	-	4.68	1.92 1.99	2.15 2.15	-	-	-	-	173.34	55.60	34.87	36.51	-	-	-	-
63 Ile	8.38	121.26	-	-	5.44	1.5	0.84 1.66 0.78	0.56	-	-	-	173.19	58.98	41.74	28.21 16.63	14.40	-	-	-
64 Lys	8.87	123.54	-	-	4.61	1.84 1.69	1.26 1.34	1.55 1.55	2.75 2.75	-	-	174.61	55.26	37.02	24.74	29.15	42.14	-	-
65 Leu	8.78	123.60	-	-	4.96	1.71 1.42	1.72	0.77 0.89	-	-	-	176.35	54.23	42.02	27.12	25.51 24.46	-	-	-
66 Gln	8.99	122.29	-	109.78	4.61	1.85 1.97	2.07 2.07	-	7.15 6.47	-	-	174.19	55.21	31.50	33.63	-	-	-	-
67 Asp	8.26	116.91	-	-	4.48	2.65 2.67	-	-	-	-	-	174.96	55.11	41.49	-	-	-	-	-
68 Thr	7.85	110.82	-	-	4.92	3.97	1.15	-	-	-	-	174.97	60.66	70.53	22.95	-	-	-	-
69 Arg	9.08	122.25	-	-	4.62	1.84 1.77	1.54 1.54	3.17 3.17	-	-	-	173.65	55.81	34.28	27.14	43.63	-	-	-
70 Leu	8.39	123.52	-	-	4.90	1.75 1.49	1.80	0.89 0.89	-	-	-	176.95	54.23	44.39	27.05	25.64 25.64	-	-	-
71 Leu	8.13	125.43	-	-	4.62	1.74 1.52	1.45	0.76 0.74	-	-	-	175.21	53.43	40.84	28.11	22.83 24.89	-	-	-
72 Gln	7.46	116.92	-	112.86	4.09	1.93 1.97	2.44 2.36	-	7.64 6.83	-	-	175.48	56.91	29.19	33.65	-	-	-	-
73 Leu	8.19	127.02	-	-	4.72	1.61 1.54	2.07	0.77 0.99	-	-	-	177.06	56.59	44.27	29.91	26.73 24.18	-	-	-
74 Ser	8.62	112.46	-	-	4.74	3.82 3.80	-	-	-	-	-	171.63	57.67	66.16	-	-	-	-	-
75 Leu	8.37	119.85	-	-	5.46	1.37 1.36	1.47	0.69 0.68	-	-	-	175.81	53.78	46.07	27.17	26.31 26.04	-	-	-
76 Glu	8.97	119.73	-	-	4.66	2.23 2.01	2.28 2.28	-	-	-	-	175.57	54.71	34.18	35.65	-	-	-	-
77 Phe	8.93	120.97	-	-	4.76	2.86 3.24	-	7.30	7.06	-	-	176.52	59.80	39.35	-	131.55	131.57	-	-

78 Ser	8.24	118.42	-	-	4.62	3.70 3.70	-	-	-	-	-	174.82	57.14	62.93	-	-	-	-	-
79 Pro	-	-	-	-	4.36	1.94 2.35	-	3.95 3.75	-	-	-	176.98	65.15	31.84	-	50.87	-	-	-
80 Asp	7.70	114.59	-	-	4.53	2.85 2.54	-	-	-	-	-	176.07	53.44	40.68	-	-	-	-	-
81 Ser	8.01	111.47	-	-	4.12	4.62 4.62	-	-	-	-	-	-	59.72	62.50	-	-	-	-	-
82 Lys	-	-	-	-	4.36	1.64 1.93	1.38 1.25	1.58 1.57	2.92 2.93	-	-	175.27	55.51	32.03	24.90	29.15	42.25	-	-
83 Gly	7.55	106.36	-	-	4.30 4.16	-	-	-	-	-	-	172.34	45.93	-	-	-	-	-	-
84 Ile	8.36	119.30	-	-	4.66	1.51	1.35 0.97 0.67	0.60	-	-	-	172.55	59.20	42.68	27.78 17.56	13.76	-	-	-
85 Asp	9.20	126.50	-	-	5.22	1.62 2.40	-	-	-	-	-	174.36	53.04	42.59	-	-	-	-	-
86 Ile	8.96	121.92	-	-	4.42	1.57	0.70 1.71 1.46	0.60	-	-	-	174.24	60.13	39.88	26.98 17.25	13.83	-	-	-
87 Trp	8.91	126.88	-	129.06	5.31	3.11 2.87	-	6.71	10.20 7.30	7.40 6.85	7.03	175.97	55.44	30.32	-	126.41	119.87	114.69 121.01	123.79
88 Ile	9.37	122.76	-	-	4.58	2.02	0.76 1.47 0.66	0.61	-	-	-	173.24	58.32	39.98	18.25 26.95	15.09	-	-	-
89 Pro	-	-	-	-	5.26	1.74 2.40	1.96 1.80	3.39 3.96	-	-	-	175.65	62.20	31.78	27.79	51.68	-	-	-
90 Leu	9.00	121.76	-	-	5.30	1.66 1.44	1.70	0.85 0.79	-	-	-	176.82	53.57	46.37	26.96	25.17 25.29	-	-	-
91 Glu	8.85	121.77	-	-	5.27	1.96 1.96	2.19 2.19	-	-	-	-	174.18	54.50	33.45	37.45	-	-	-	-
92 Leu	8.96	119.19	-	-	4.84	1.75 1.55	1.58	0.71 0.71	-	-	-	174.37	54.58	44.56	26.54	26.40 26.40	-	-	-
93 Ser	8.43	114.61	-	-	5.63	3.67 3.67	-	-	-	-	-	173.48	56.83	65.06	-	-	-	-	-
94 Val	9.33	124.77	-	-	4.66	1.85	0.69 0.72	-	-	-	-	173.83	61.00	34.22	20.83 20.71	-	-	-	-
95 Tyr	9.19	129.71	-	-	5.11	2.76 2.76	-	6.86	6.59	-	-	173.80	56.54	40.25	-	133.18	117.97	-	-
96 Leu	8.75	126.86	-	-	4.79	1.30 1.23	1.34	0.70 0.70	-	-	-	174.02	53.74	45.87	27.15	26.57 25.72	-	-	-
97 Lys	8.81	129.36	-	-	4.48	1.38 1.85	1.40 1.39	1.62 1.62	2.93 2.93	-	-	174.46	54.76	34.16	25.09	29.03	42.38	-	-

98 Leu	8.59	123.92	-	-	4.41	1.35 1.59	1.44	0.68 0.65	-	-	-	176.26	53.67	43.37	-	25.96 23.70	-	-	-
99 Leu	8.59	117.21	-	-	3.91	1.43 2.09	1.47	0.82 0.85	-	-	-	-	58.14	39.50	27.71	25.85 23.18	-	-	-
100 Ile	7.55	113.23	-	-	4.12	2.1	0.92 1.03 1.19	0.82	-	-	-	174.97	61.57	37.88	26.59 18.80	14.55	-	-	-
101 Leu	7.59	120.09	-	-	4.26	1.55 1.64	1.64	0.81 0.79	-	-	-	175.93	54.24	44.13	27.25	24.46 25.68	-	-	-
102 Glu	8.34	121.57	-	-	4.37	-	2.24 2.24	-	-	-	-	-	55.35	28.26	36.23	-	-	-	-
103 Pro	-	-	-	-	4.38	1.77 1.99	2.15 1.93	3.63 3.74	-	-	-	175.95	63.59	32.21	28.01	50.19	-	-	-
104 Leu	8.36	121.95	-	-	4.81	1.59 1.30	1.86	0.81 0.68	-	-	-	175.31	53.50	45.81	26.15	25.79 24.40	-	-	-
105 Thr	8.23	114.99	-	-	5.33	3.92	1.07	-	-	-	-	173.70	61.50	70.45	21.70	-	-	-	-
106 Leu	9.47	127.39	-	-	5.06	1.47 1.25	1.48	0.66 0.67	-	-	-	174.81	53.67	43.42	28.03	25.94 25.00	-	-	-
107 Tyr	9.39	125.31	-	-	4.86	2.96 2.75	-	6.88	6.55	-	-	173.94	57.37	40.05	-	133.13	117.71	-	-
108 Val	8.61	126.46	-	-	4.72	1.82	0.78 0.70	-	-	-	-	174.09	60.15	35.04	21.54 20.98	-	-	-	-
109 Arg	8.59	125.43	-	-	5.07	1.83 1.65	1.53 1.53	3.18 3.34	-	-	-	175.90	53.76	31.77	28.71	42.43	-	-	-
110 Thr	8.73	121.60	-	-	4.95	3.71	0.89	-	-	-	-	170.06	59.85	71.05	19.87	-	-	-	-
111 Asp	8.37	127.12	-	-	5.52	2.63 2.74	-	-	-	-	-	175.17	53.34	43.68	-	-	-	-	-
112 Ile	8.86	123.49	-	-	4.58	2.08	0.72 0.85 1.86	0.93	-	-	-	174.28	61.15	38.48	26.90 17.76	14.95	-	-	-
113 Arg	9.06	128.32	-	-	4.39	0.71 1.35	- 0.27 0.37	2.57 2.57	-	-	-	175.57	54.16	32.13	26.49	43.45	-	-	-
114 Val	8.97	128.38	-	-	4.37	1.78	0.57 0.75	-	-	-	-	174.19	60.65	33.90	22.43 22.07	-	-	-	-
115 Gln	8.22	125.06	-	112.67	4.99	1.58 1.76	1.94 1.97	-	6.69 8.01	-	-	175.22	54.83	31.99	34.58	-	-	-	-
116 Leu	9.11	124.64	-	-	5.06	1.18 1.80	1.54	0.71 0.70	-	-	-	176.44	53.03	42.94	26.84	25.36 23.74	-	-	-
117 Arg	8.97	121.79	-	-	4.82	1.81 1.60	1.34 1.48	3.08 3.08	-	-	-	174.71	53.78	33.41	26.76	42.89	-	-	-
118 Leu	8.40	123.32	-	-	4.83	1.27	1.51	0.62	-	-	-	176.17	55.13	43.60	26.98	25.55	-	-	-

						1.69		0.70								25.81			
119 Glu	9.00	126.19	-	-	4.76	2.00 1.69	2.10 2.10	-	-	-	-	174.74	54.80	33.67	36.23	-	-	-	-
120 Ser	8.43	115.43	-	-	4.91	3.46 3.46	-	-	-	-	-	174.33	57.15	65.05	-	-	-	-	-
121 Asp	8.41	123.99	-	-	4.75	2.58 2.93	-	-	-	-	-	177.80	52.58	42.32	-	-	-	-	-
122 Glu	8.80	118.69	-	-	4.03	2.01 1.94	2.24 2.24	-	-	-	-	176.69	58.58	29.46	36.41	-	-	-	-
123 Asp	8.16	118.42	-	-	4.60	2.71 2.62	-	-	-	-	-	176.61	54.22	41.33	-	-	-	-	-
124 Gly	8.03	108.25	-	-	3.49 4.02	-	-	-	-	-	-	174.01	45.45	-	-	-	-	-	-
125 Lys	7.91	121.24	-	-	4.24	1.74 1.84	1.28 1.33	1.62 1.62	2.94 2.94	-	-	176.15	55.34	32.71	24.90	28.82	42.20	-	-
126 Tyr	8.22	120.46	-	-	5.11	2.72 2.87	-	7.02	6.72	-	-	175.66	58.34	41.25	-	133.07	118.41	-	-
127 Arg	9.05	118.96	-	-	4.57	1.74 1.74	1.45 1.31	3.13 3.13	-	-	-	173.65	54.23	33.54	26.85	43.46	-	-	-
128 Leu	8.53	121.55	-	-	4.85	1.47 1.36	1.43	0.71 0.74	-	-	-	175.31	53.48	43.15	26.59	25.31 25.05	-	-	-
129 Ala	9.15	127.73	-	-	4.57	1.21	-	-	-	-	-	176.45	50.46	21.62	-	-	-	-	-
130 Phe	8.76	123.53	-	-	4.20	3.09 2.75	-	6.96	7.00	7.11	-	174.87	59.01	39.07	-	132.17	131.00	-	-
131 Gly	8.50	117.90	-	-	4.06 3.31	-	-	-	-	-	-	172.71	46.01	-	-	-	-	-	-
132 His	7.27	118.05	-	-	4.34	2.91 2.67	-	6.68	7.59	-	-	172.54	56.81	32.91	-	121.07	138.77	-	-
133 Cys	7.70	124.83	-	-	5.29	2.93 3.20	-	-	-	-	-	-	55.12	43.57	-	-	-	-	-
134 Ser	8.61	121.98	-	-	4.58	3.79 3.82	-	-	-	-	-	171.03	57.70	65.52	-	-	-	-	-
135 Leu	8.61	121.88	-	-	5.17	1.35 1.52	1.66	0.71 0.76	-	-	-	176.01	55.54	43.91	31.09	26.67 24.84	-	-	-
136 Leu	8.88	121.28	-	-	4.97	1.66 1.60	1.66	0.97 0.97	-	-	-	175.47	51.73	43.67	27.99	25.33 25.33	-	-	-
137 Pro	-	-	-	-	4.43	2.07 1.98	1.90 2.32	3.78 3.83	-	-	-	175.75	63.05	31.03	27.45	51.11	-	-	-
138 Arg	8.81	123.06	-	-	4.44	1.52 1.81	1.55 1.55	2.90 3.07	-	-	-	176.16	55.17	31.49	25.66	42.67	-	-	-
139 Ala	7.80	119.84	-	-	4.52	1.39	-	-	-	-	-	175.42	53.06	21.96	-	-	-	-	-

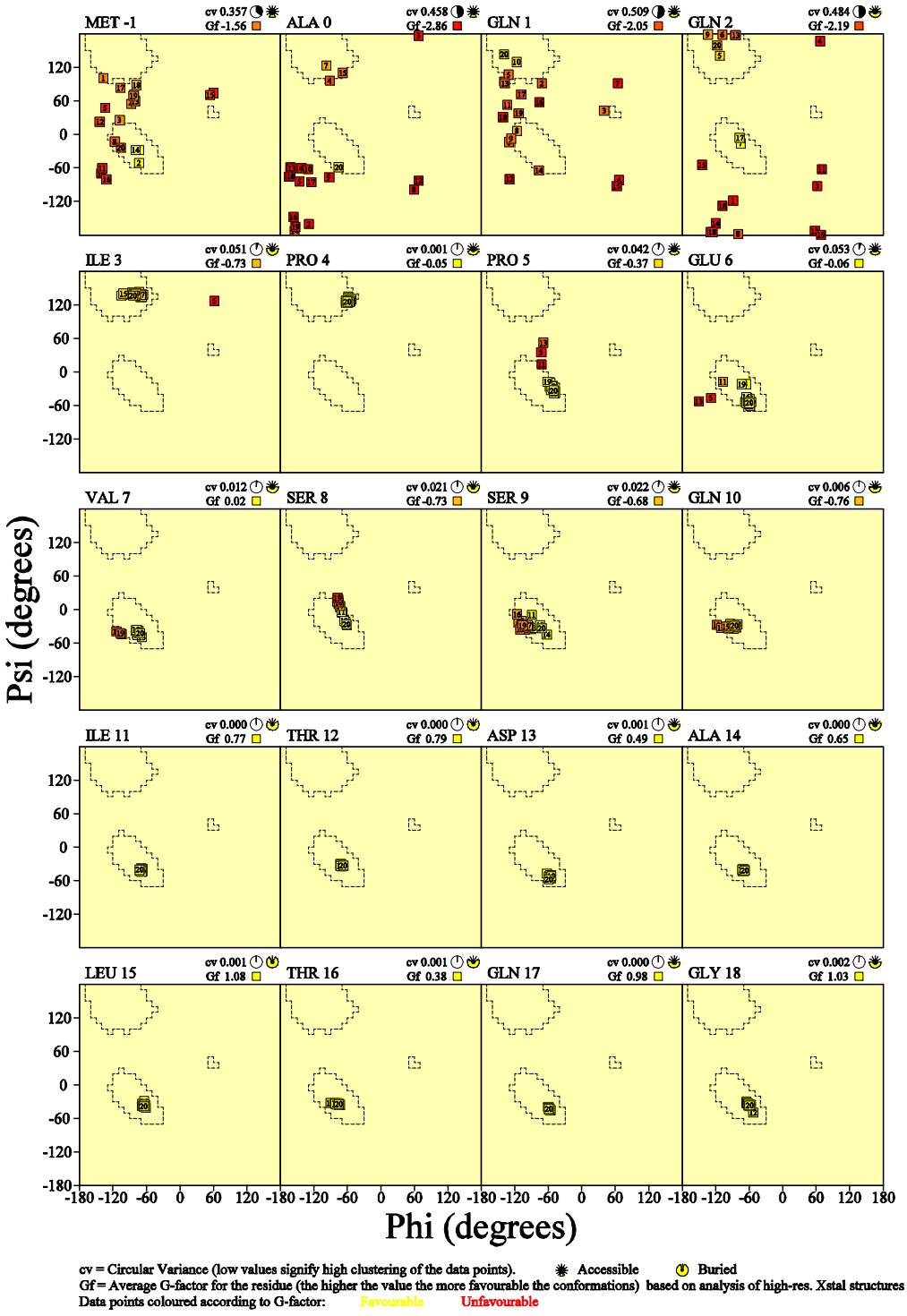
140 Ile	8.27	121.68	-	-	4.87	1.75	0.99 0.75 1.64	0.84	-	-	-	175.12	61.60	40.59	18.15 27.74	14.25	-	-	-
141 Glu	9.34	124.88	-	-	4.83	2.07 1.91	2.19 2.19	-	-	-	-	175.34	54.24	34.22	35.92	-	-	-	-
142 Leu	8.98	125.70	-	-	4.47	1.41 1.82	1.52	0.92 0.81	-	-	-	176.27	55.91	42.38	27.35	25.92 25.10	-	-	-
143 Gln	8.93	127.45	-	111.05	4.43	1.93 1.93	2.22 2.30	-	7.19 6.68	-	-	175.77	56.45	30.47	34.21	-	-	-	-
144 Ser	8.08	114.55	-	-	4.52	3.80 3.87	-	-	-	-	-	-	58.27	64.35	-	-	-	-	-
145 Gly	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
146 Asn	-	-	-	-	5.02	2.71 2.96	-	-	-	-	-	-	50.75	39.16	-	-	-	-	-
147 Pro	-	-	-	-	4.29	2.20 1.93	1.98 1.98	3.80 3.90	-	-	-	176.69	63.69	31.95	27.22	50.64	-	-	-
148 Leu	8.01	119.57	-	-	4.32	1.57 1.71	1.64	0.91 0.82	-	-	-	177.82	54.90	41.95	27.26	22.93 25.08	-	-	-
150 Leu	-	-	-	-	4.38	1.56 1.74	-	0.89	-	-	-	-	-	40.51	-	23.92	-	-	-
151 Pro	-	-	-	-	4.26	1.83 2.15	-	3.77 3.77	-	-	-	177.38	65.01	31.39	-	50.18	-	-	-
152 Val	7.15	115.88	-	-	3.61	2.14	1.00 0.87	-	-	-	-	-	65.60	32.05	22.91 21.76	-	-	-	-
153 Asn	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
154 Ala	-	-	-	-	4.21	1.47	-	-	-	-	-	179.80	54.53	18.91	-	-	-	-	-
155 Val	7.46	110.55	-	-	4.18	2.27	0.93 0.95	-	-	-	-	177.14	62.80	32.69	20.40 21.38	-	-	-	-
156 Leu	8.05	123.47	-	-	3.87	1.74 1.66	1.75	0.85 0.86	-	-	-	178.44	58.88	41.34	26.77	24.85 23.98	-	-	-
157 Gly	8.57	105.95	-	-	3.85 3.85	-	-	-	-	-	-	176.60	47.25	-	-	-	-	-	-
158 Thr	7.64	119.04	-	-	4.03	4.2	1.16	-	-	-	-	176.92	66.71	68.72	22.00	-	-	-	-
159 Ile	7.82	122.65	-	-	3.41	1.96	1.71 1.71 0.69	0.61	-	-	-	177.11	66.21	37.90	17.38 29.20	13.84	-	-	-
160 Glu	8.86	119.77	-	-	3.67	2.00 2.15	2.00 2.34	-	-	-	-	179.37	60.61	29.65	36.26	-	-	-	-
161 Asn	8.06	117.97	-	-	4.43	2.87 2.81	-	-	-	-	-	177.30	56.53	38.73	-	-	-	-	-

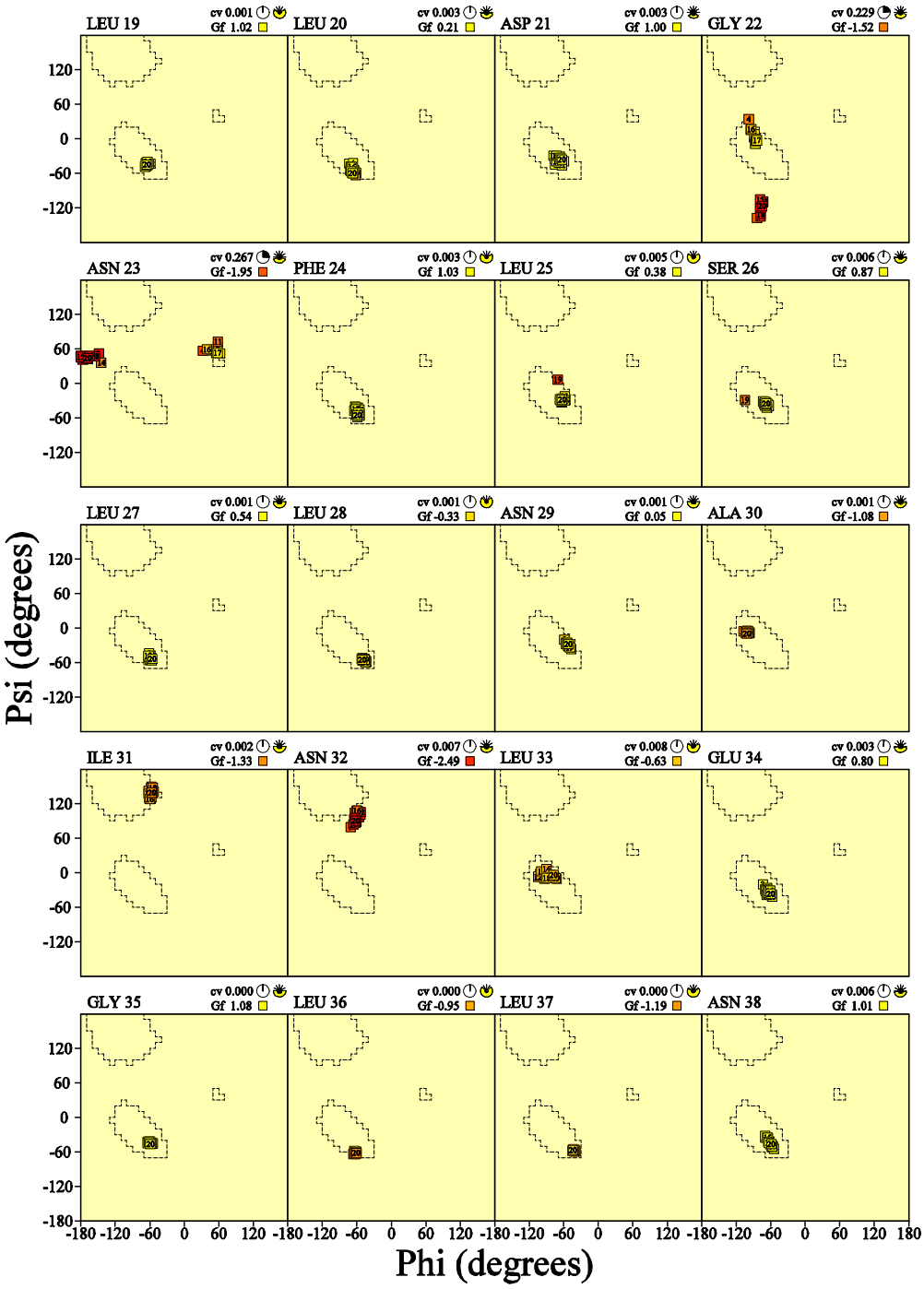
162 Ala	7.99	122.20	-	-	4.16	1.37	-	-	-	-	-	181.10	55.10	18.07	-	-	-	-	-
163 Leu	8.68	119.35	-	-	4.10	1.80 1.51	1.68	0.76 0.75	-	-	-	178.63	57.87	42.02	27.65	25.25 24.92	-	-	-
164 Gly	8.48	107.05	-	-	3.57 4.12	-	-	-	-	-	-	175.78	47.96	-	-	-	-	-	-
165 Asn	7.84	119.57	112.58	-	4.53	2.85 2.85	-	7.61 6.82	-	-	-	176.22	56.24	38.56	-	-	-	-	-
166 Phe	7.97	121.17	-	-	4.65	3.31 3.31	-	7.18	7.07	-	-	177.21	60.68	40.18	-	132.51	130.66	-	-
167 Ile	8.45	118.65	-	-	3.37	2.04	1.94 0.73 1.06	0.76	-	-	-	176.07	64.78	38.12	29.18 18.77	14.01	-	-	-
168 Thr	7.82	108.00	-	-	4.05	4.32	1.40	-	-	-	-	176.72	65.56	69.43	21.96	-	-	-	-
169 Glu	8.40	119.56	-	-	4.42	2.03 1.98	2.32 2.24	-	-	-	-	177.18	57.67	30.14	36.12	-	-	-	-
170 Asp	8.01	119.82	-	-	4.44	2.27 2.13	-	-	-	-	-	179.28	57.24	41.55	-	-	-	-	-
171 Leu	8.04	119.20	-	-	3.82	1.67 1.52	1.27	0.71 0.60	-	-	-	179.01	58.01	40.66	26.64	24.34 24.60	-	-	-
172 Gly	7.54	105.79	-	-	3.85 3.59	-	-	-	-	-	-	175.73	48.15	-	-	-	-	-	-
173 Ala	7.58	122.62	-	-	4.11	1.42	-	-	-	-	-	179.61	54.49	18.12	-	-	-	-	-
174 Gly	7.66	101.44	-	-	3.49 4.15	-	-	-	-	-	-	175.34	45.51	-	-	-	-	-	-
175 Leu	7.91	118.75	-	-	4.57	1.73 1.51	1.76	0.69 0.41	-	-	-	177.56	56.39	44.04	26.50	25.50 21.75	-	-	-
176 Cys	8.53	116.92	-	-	4.51	3.47 2.96	-	-	-	-	-	174.03	61.30	39.17	-	-	-	-	-
177 Pro	-	-	-	-	4.20	2.29 1.82	2.17 1.93	3.38 3.63	-	-	-	179.81	66.76	30.65	28.32	51.22	-	-	-
178 Thr	7.45	115.17	-	-	3.98	4.25	1.32	-	-	-	-	176.35	66.99	68.42	21.47	-	-	-	-
179 Leu	7.61	122.58	-	-	3.95	1.30 2.32	1.78	0.85 0.46	-	-	-	177.64	57.79	40.91	27.09	26.58 23.71	-	-	-
180 Asn	7.97	118.41	110.36	-	3.25	2.59 1.57	-	6.68 6.53	-	-	-	177.90	56.17	37.93	-	-	-	-	-
181 Ser	7.56	115.38	-	-	-	3.92 3.92	-	-	-	-	-	176.26	61.85	62.63	-	-	-	-	-
182 Leu	7.95	121.24	-	-	4.03	1.59 1.69	1.64	0.76 0.75	-	-	-	180.00	58.25	42.23	26.77	24.49 25.26	-	-	-
183 Val	8.18	116.28	-	-	3.58	2.2	1.05 0.92	-	-	-	-	178.02	65.70	31.28	22.30 21.49	-	-	-	-

184 Ser	8.03	115.16	-	-	4.28	3.90 3.90	-	-	-	-	-	174.36	61.15	63.04	-	-	-	-	-
185 Asn	7.30	117.88	113.34	-	4.85	2.95 2.60	-	6.77 7.55	-	-	-	174.62	52.91	39.91	-	-	-	-	-
186 Leu	7.07	121.34	-	-	4.07	1.66 1.33	2.02	0.75 0.76	-	-	-	176.37	55.52	43.32	26.23	24.89 23.17	-	-	-
187 Asp	8.39	121.81	-	-	4.32	2.75 2.54	-	-	-	-	-	177.10	55.10	43.08	-	-	-	-	-
188 Leu	8.41	125.88	-	-	3.96	1.72 1.59	1.73	0.94 0.91	-	-	-	178.84	58.15	42.10	26.95	24.08 24.31	-	-	-
189 Gln	8.61	117.83	-	-	4.06	2.10 2.06	2.35 2.39	-	-	-	-	178.81	58.97	28.19	34.06	-	-	-	-
190 Leu	7.41	119.70	-	-	4.11	1.93 1.56	1.68	0.84 0.93	-	-	-	179.05	57.77	41.30	27.34	24.03 25.34	-	-	-
191 Val	7.88	119.24	-	-	3.35	2.13	0.96 0.96	-	-	-	-	177.91	67.46	31.78	23.46 21.41	-	-	-	-
192 Asn	8.45	117.62	111.01	-	4.39	2.72 2.82	-	7.40 6.78	-	-	-	177.72	56.36	37.91	-	-	-	-	-
193 Asn	7.87	118.68	110.04	-	4.49	2.68 3.04	-	6.91 7.28	-	-	-	178.15	56.06	38.02	-	-	-	-	-
194 Leu	7.99	120.58	-	-	3.95	2.08 1.25	1.73	0.67 0.68	-	-	-	178.16	58.33	42.32	26.56	24.81 23.45	-	-	-
195 Ile	8.41	119.12	-	-	3.31	1.86	0.67	0.77	-	-	-	177.47	66.76	38.20	29.97 16.56	13.87	-	-	-
196 Asn	7.95	116.22	111.68	-	4.31	2.81 2.81	-	6.72 7.47	-	-	-	177.55	56.48	38.02	-	-	-	-	-
197 Leu	7.78	119.94	-	-	4.18	1.69 2.06	1.92	0.93 0.89	-	-	-	179.28	58.11	42.47	26.80	24.42 25.91	-	-	-
198 Ile	7.85	118.71	-	-	3.57	1.84	1.96 1.96 0.80	0.66	-	-	-	178.51	65.56	38.41	29.10 18.51	14.48	-	-	-
199 Leu	8.32	119.71	-	-	3.83	1.33 1.72	1.47	0.35 0.51	-	-	-	179.36	57.70	41.83	26.26	22.10 25.55	-	-	-
200 Asp	8.02	118.41	-	-	4.45	2.67 2.72	-	-	-	-	-	177.89	56.42	41.04	-	-	-	-	-
201 Arg	7.53	119.01	-	-	4.05	1.73 1.57	1.31 1.31	2.52 2.52	-	-	-	176.87	56.94	30.18	26.75	42.91	-	-	-
202 Ala	7.86	121.42	-	-	4.17	1.35	-	-	-	-	-	177.78	53.22	18.89	-	-	-	-	-
203 Asn	7.91	116.26	-	-	4.65	2.85 2.72	-	-	-	-	-	175.38	53.47	38.77	-	-	-	-	-
204 Val	7.68	119.16	-	-	4.03	2.05	0.87 0.88	-	-	-	-	175.57	62.63	32.74	21.16 20.62	-	-	-	-

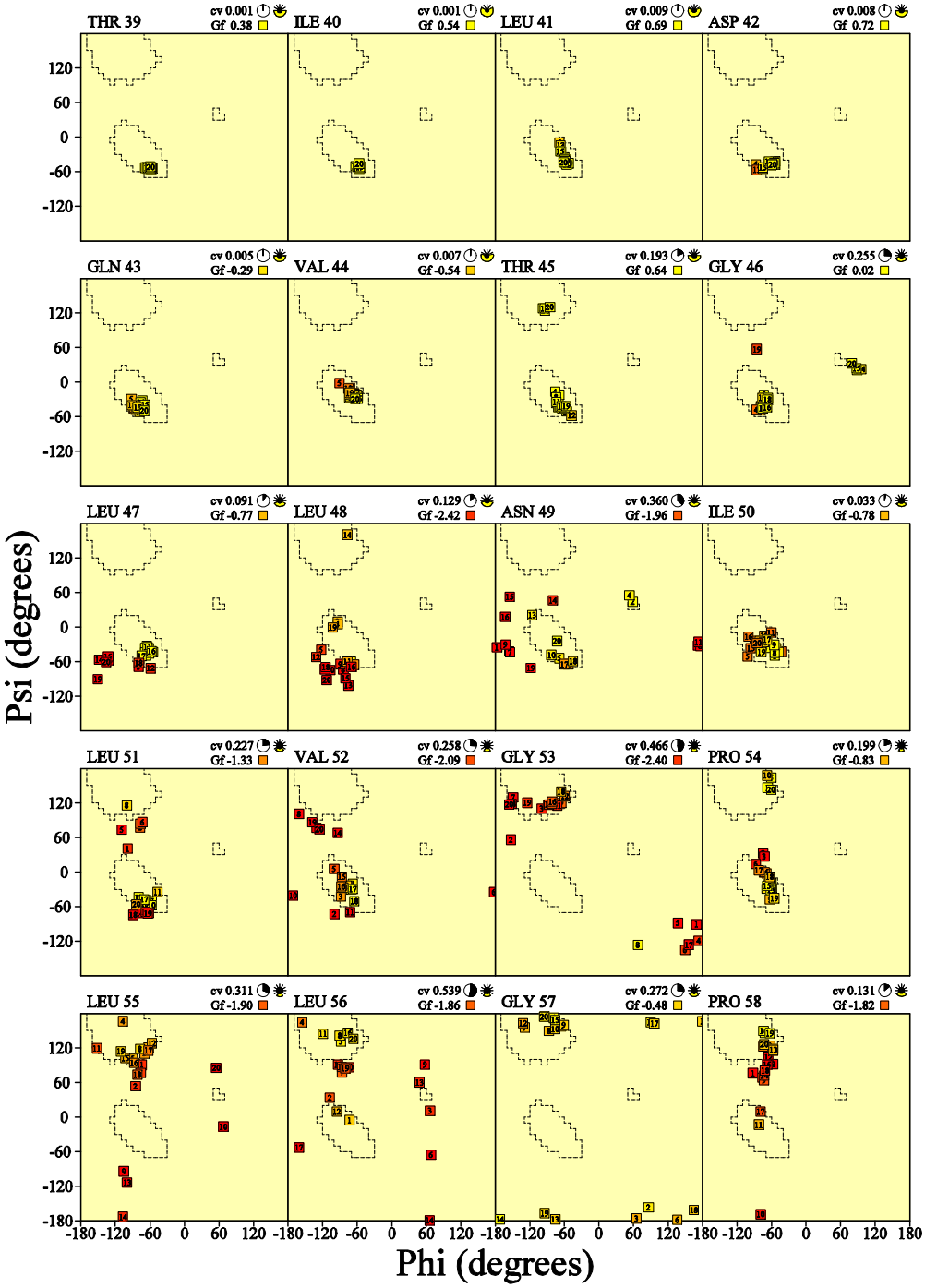
205 Asp	8.20	123.28	-	-	4.58	2.53 2.70	-	-	-	-	-	176.09	54.32	41.14	-	-	-	-	-
206 Leu	8.07	123.12	-	-	4.36	1.61 1.61	1.61	0.88 0.82	-	-	-	177.36	55.08	42.33	26.88	23.30 25.11	-	-	-
207 Ser	8.28	117.26	-	-	4.41	3.83 3.83	-	-	-	-	-	173.59	58.57	63.96	-	-	-	-	-
208 Val	7.51	125.09	-	-	4.02	2.06	0.84 0.86	-	-	-	-	180.95	63.57	33.20	21.60 19.99	-	-	-	-

Appendix B- PROCHECK Analysis of Latherin Ensemble

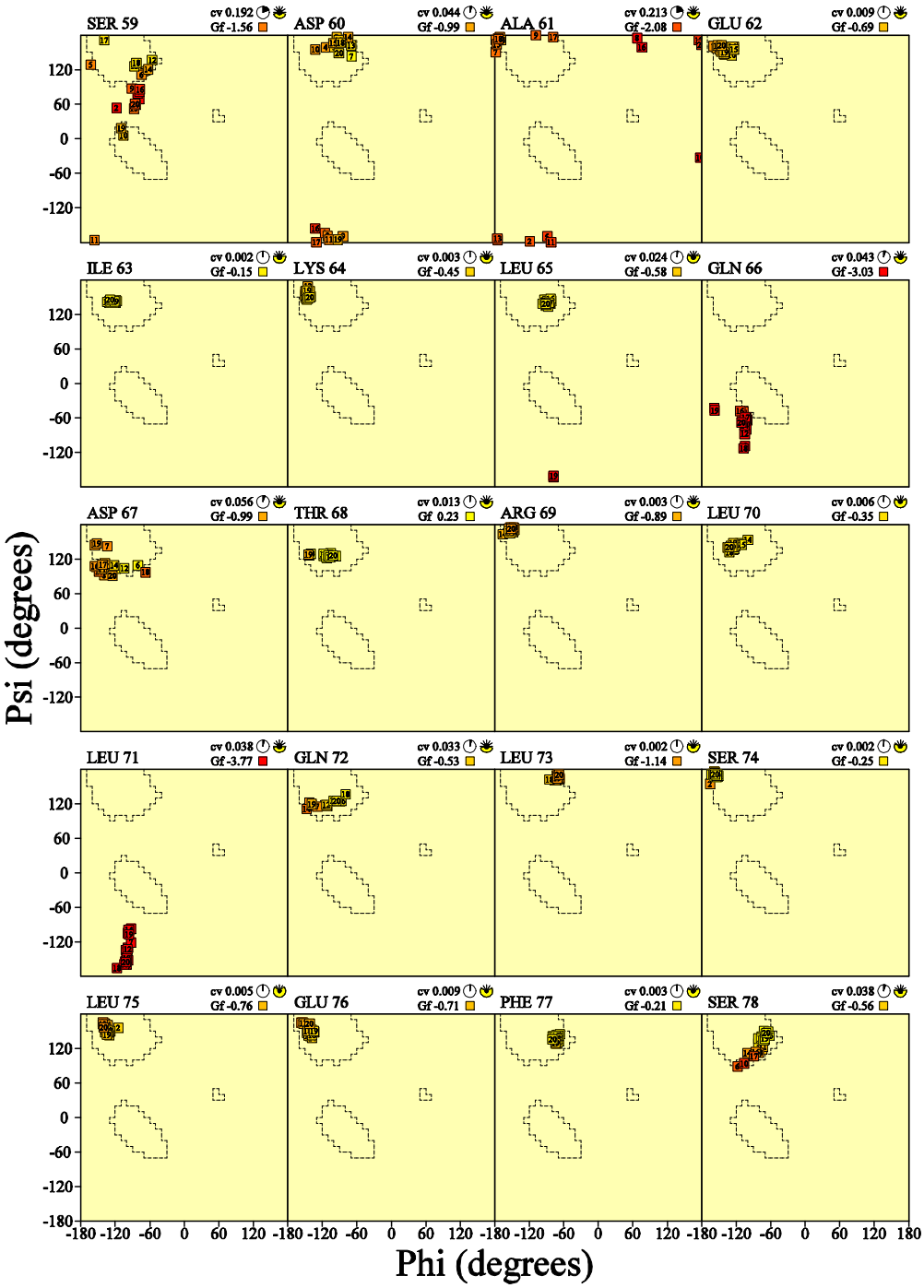




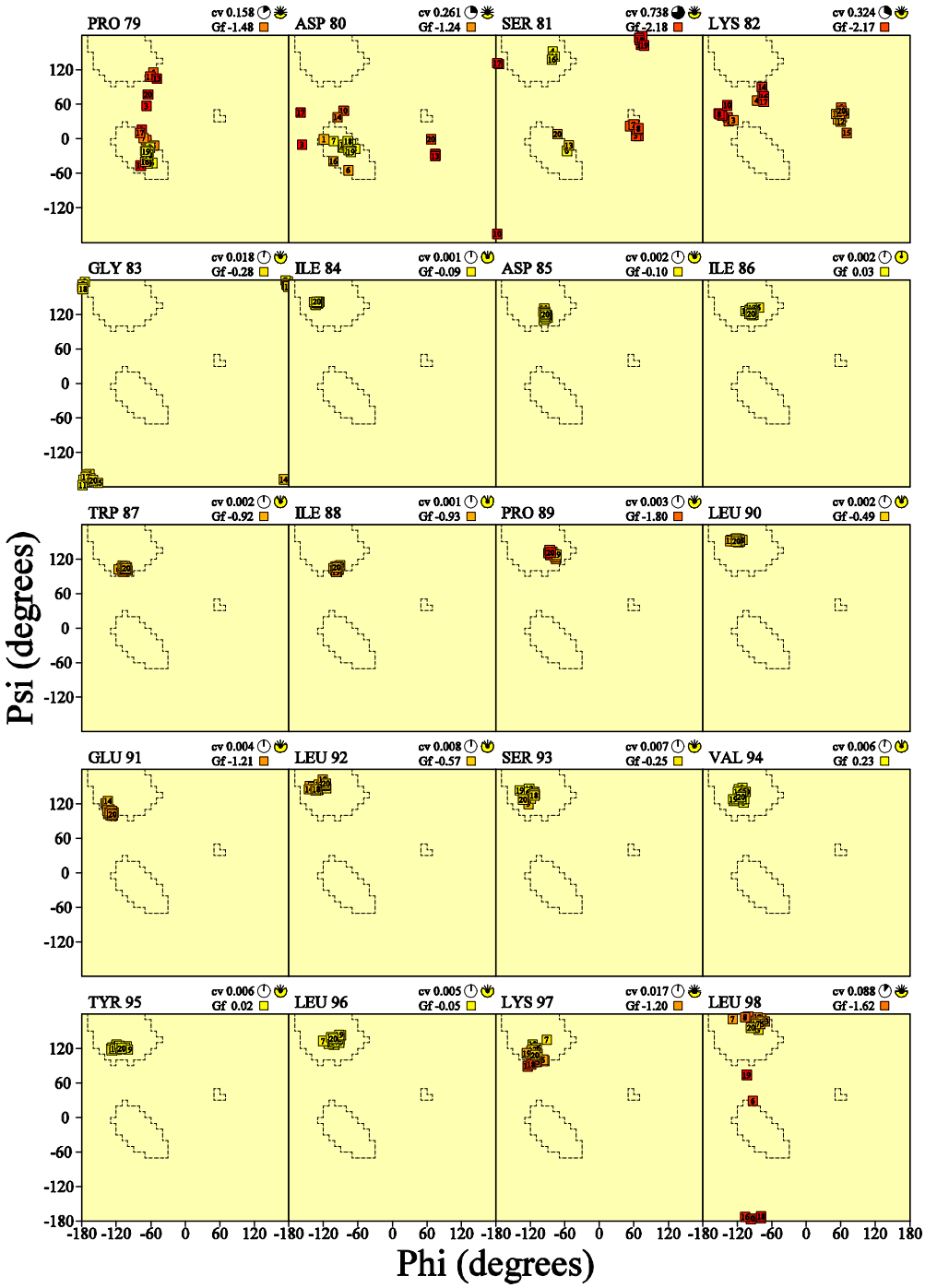
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable



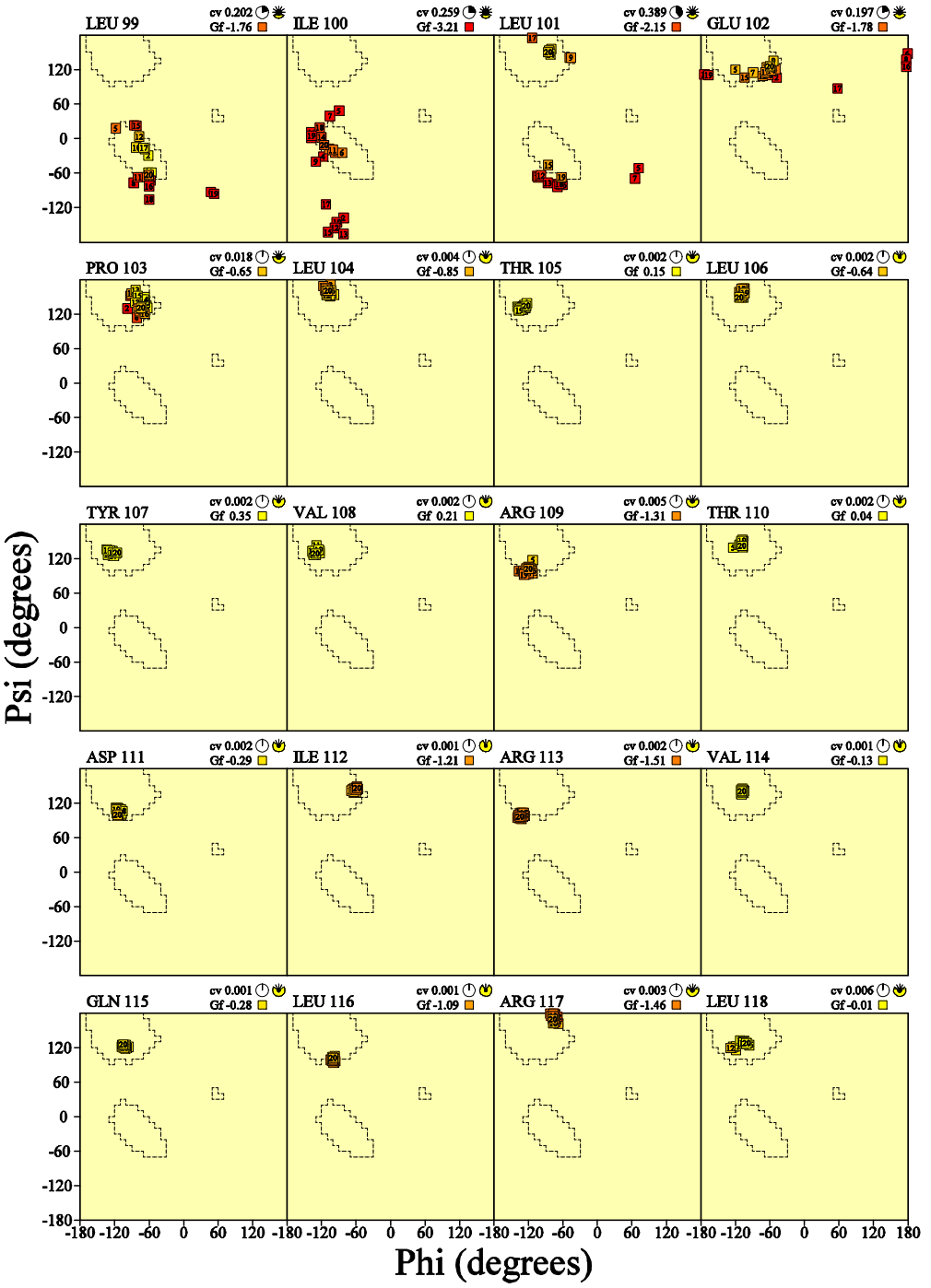
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable



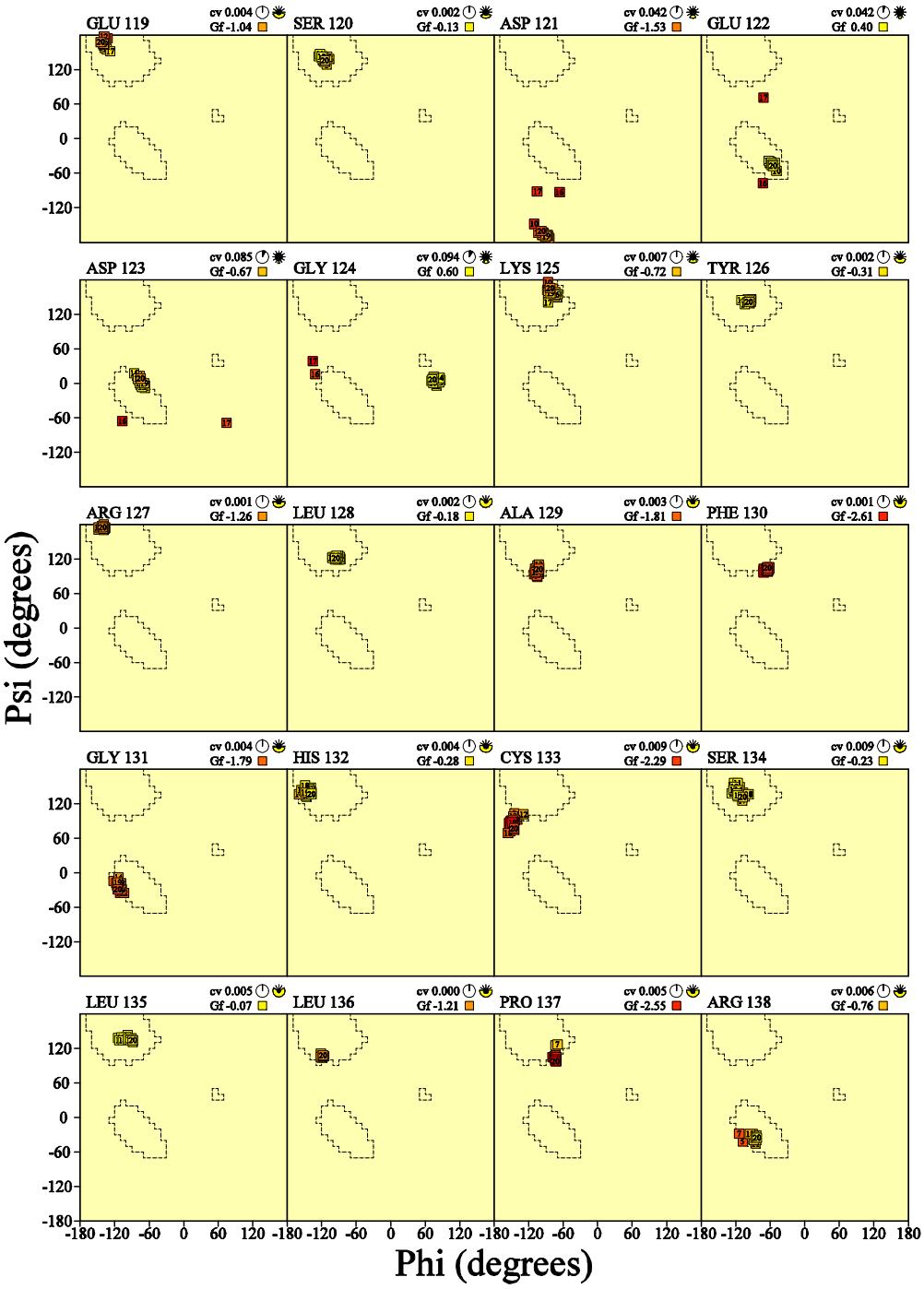
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable



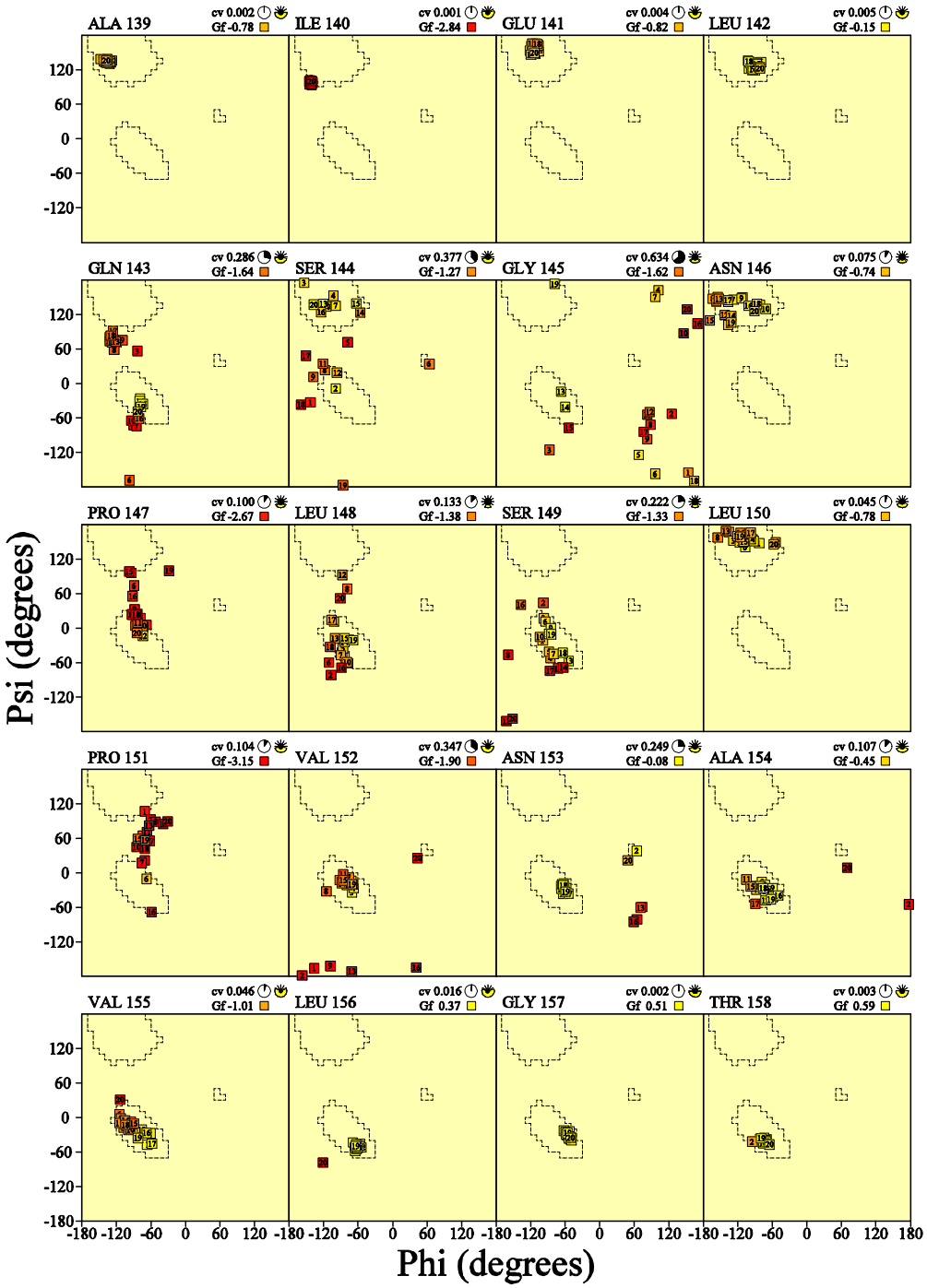
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable



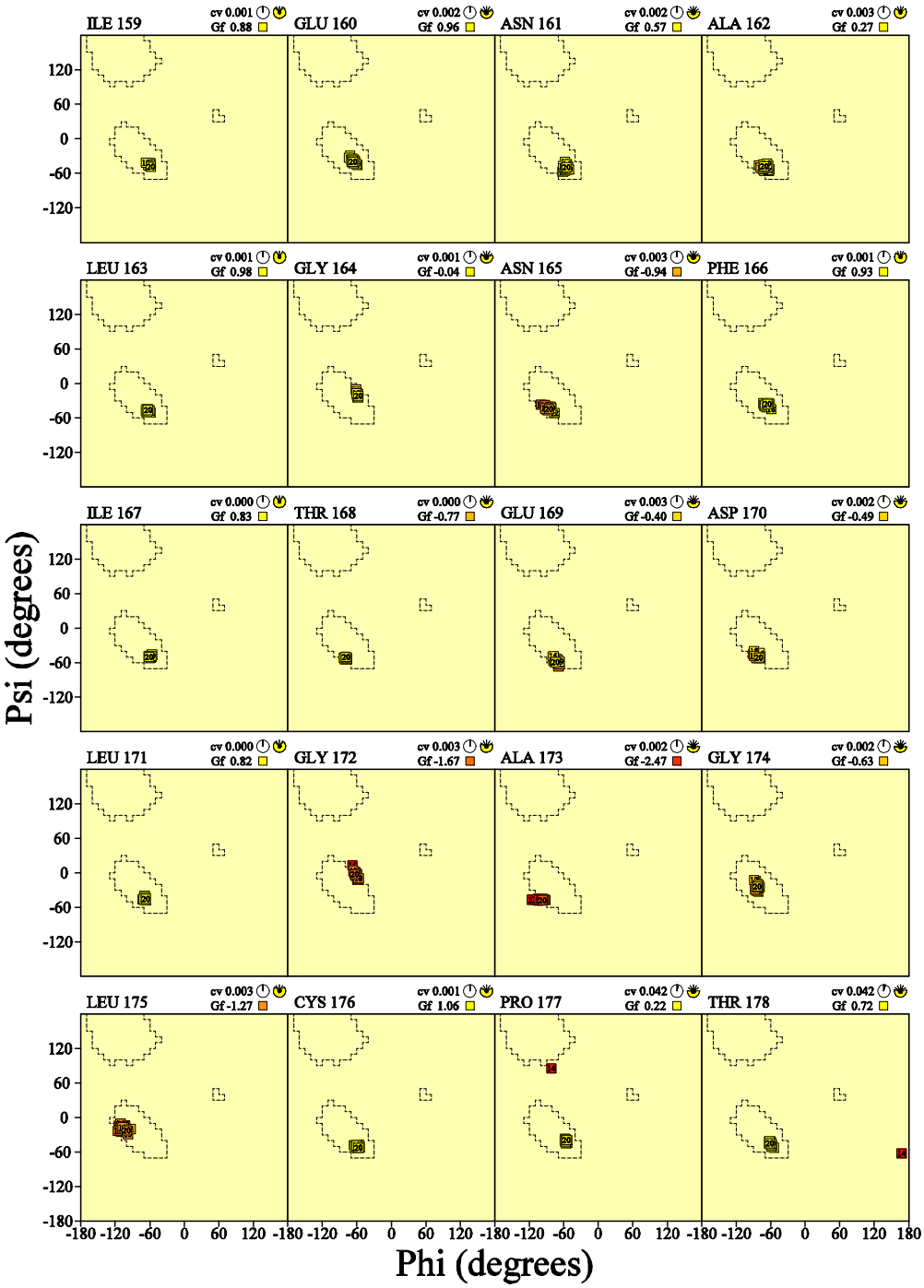
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable



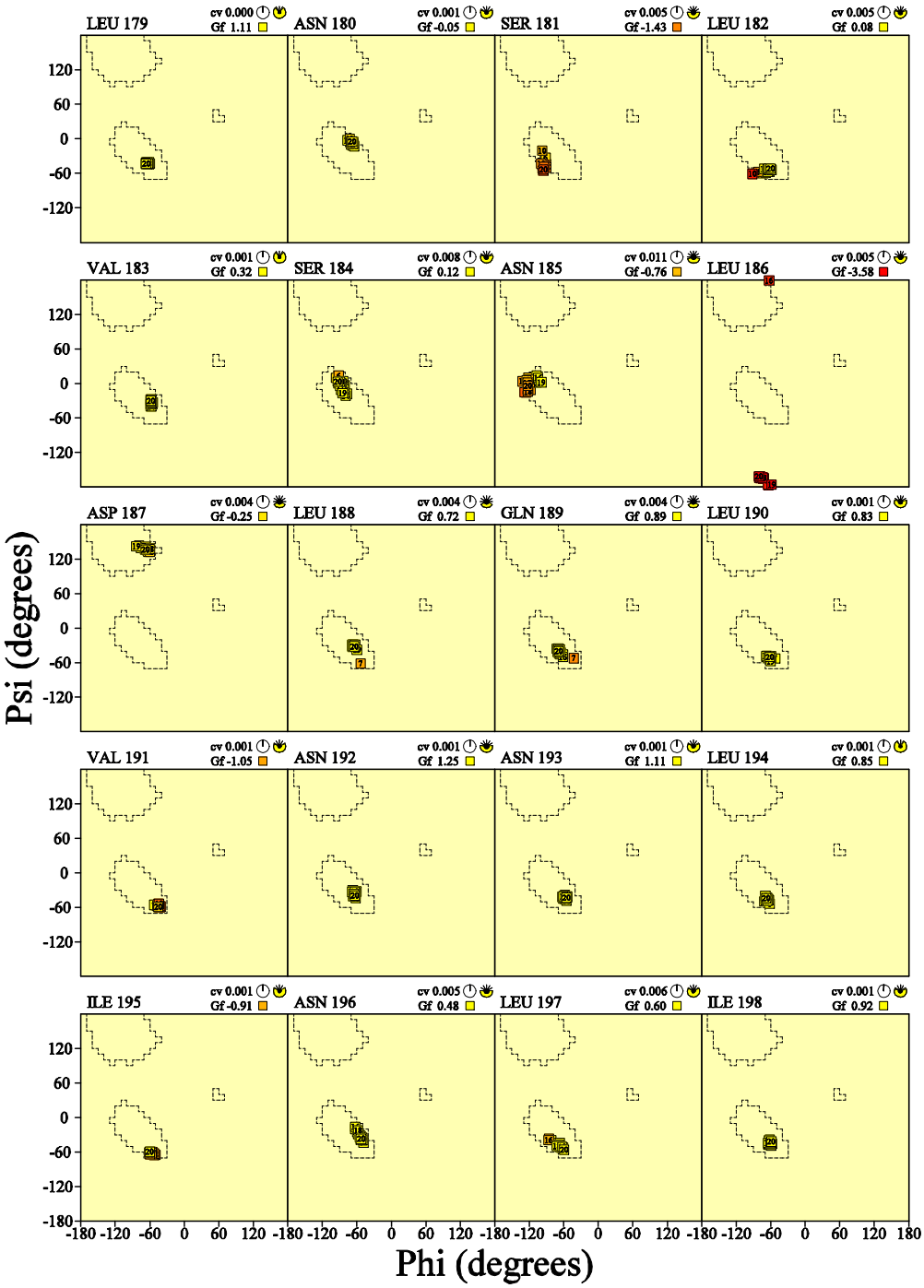
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable



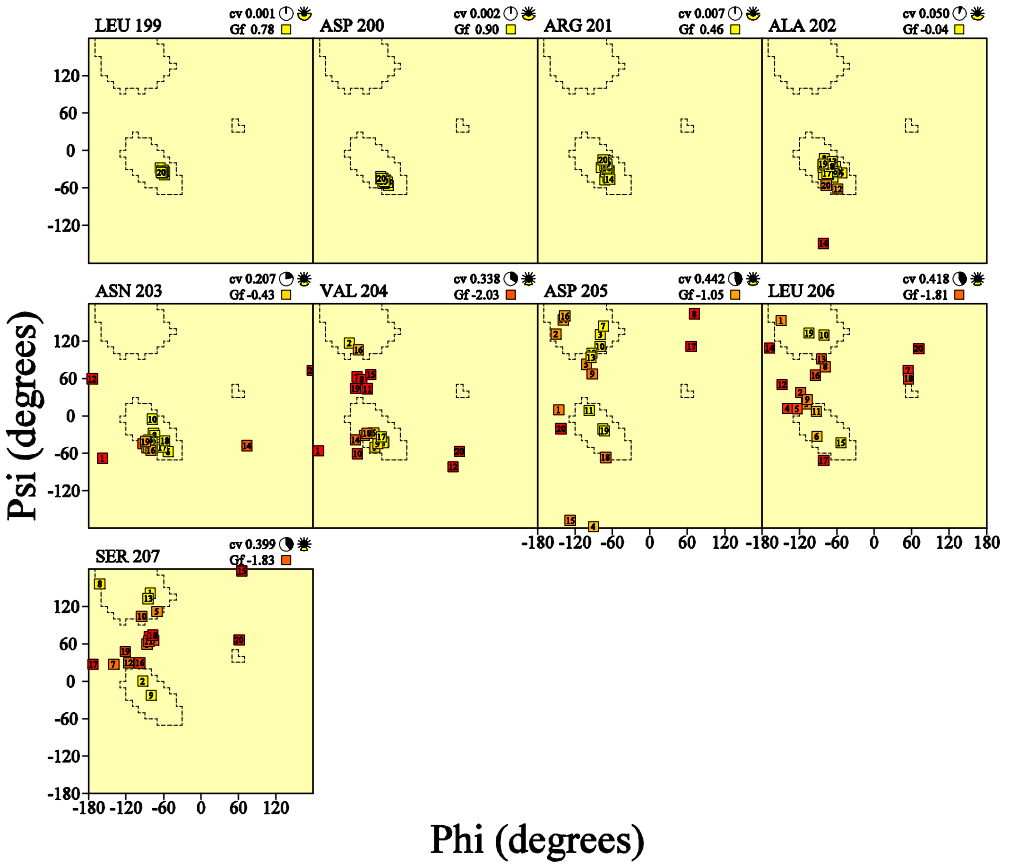
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable



cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable

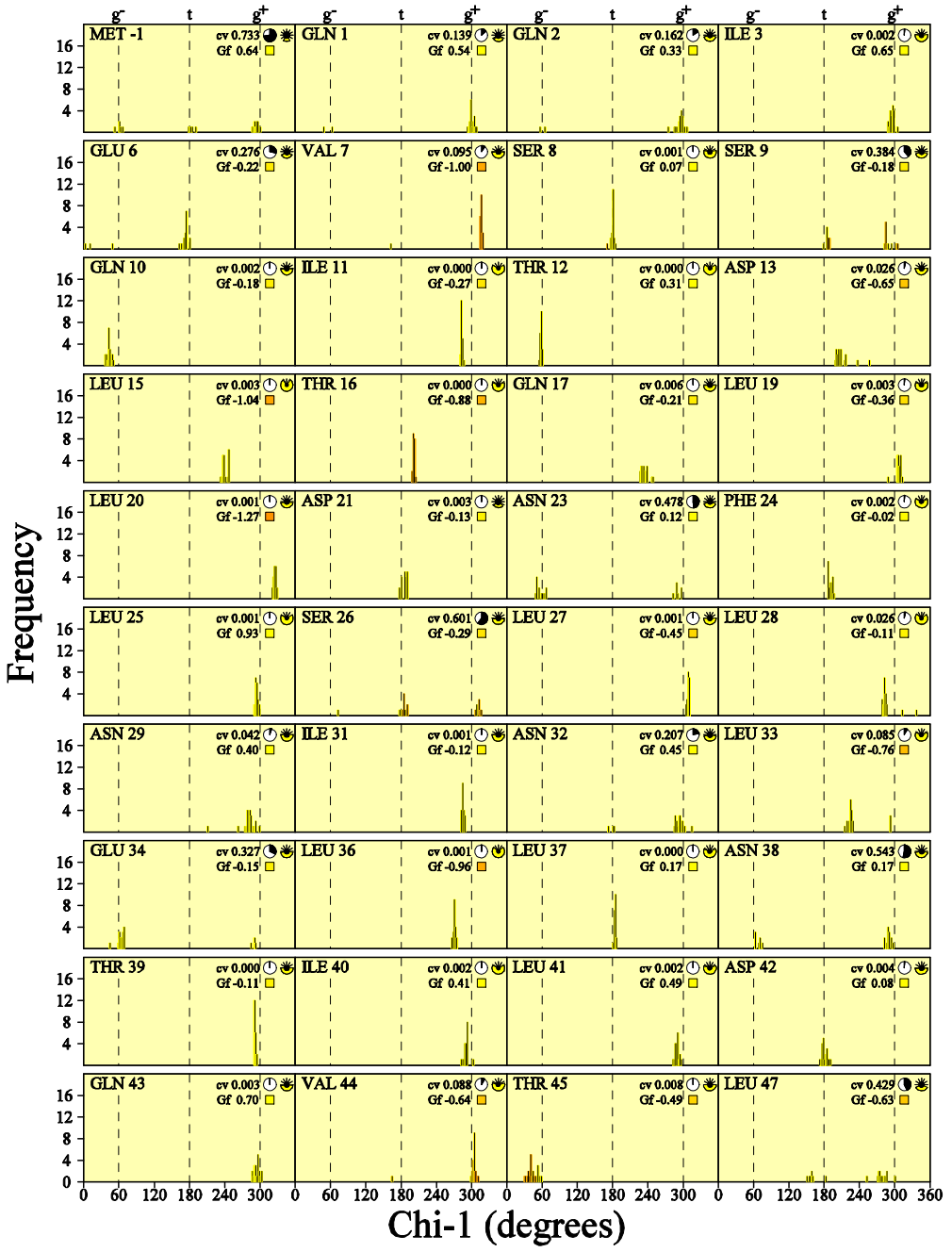


cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable

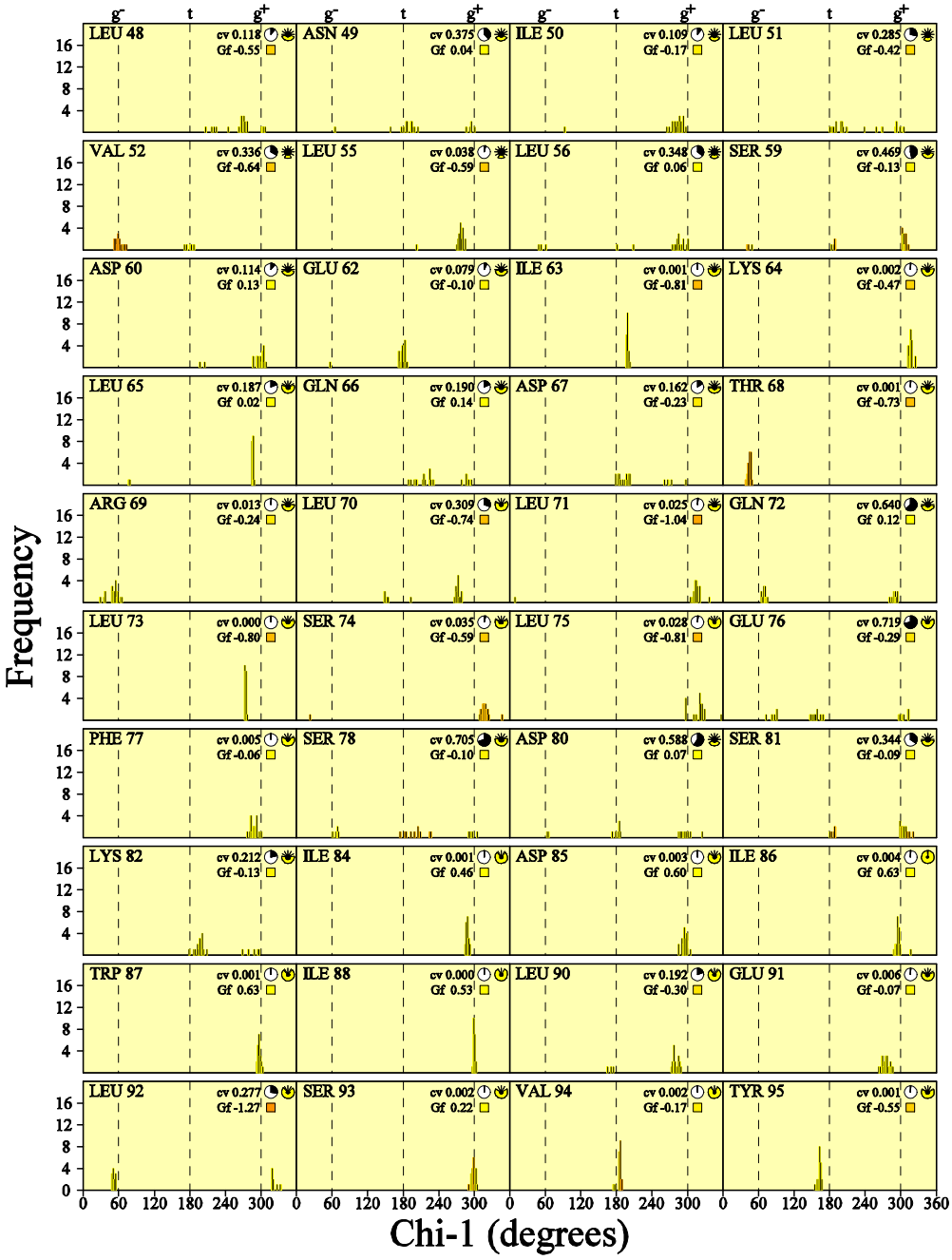


cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data points coloured according to G-factor: Favourable Unfavourable

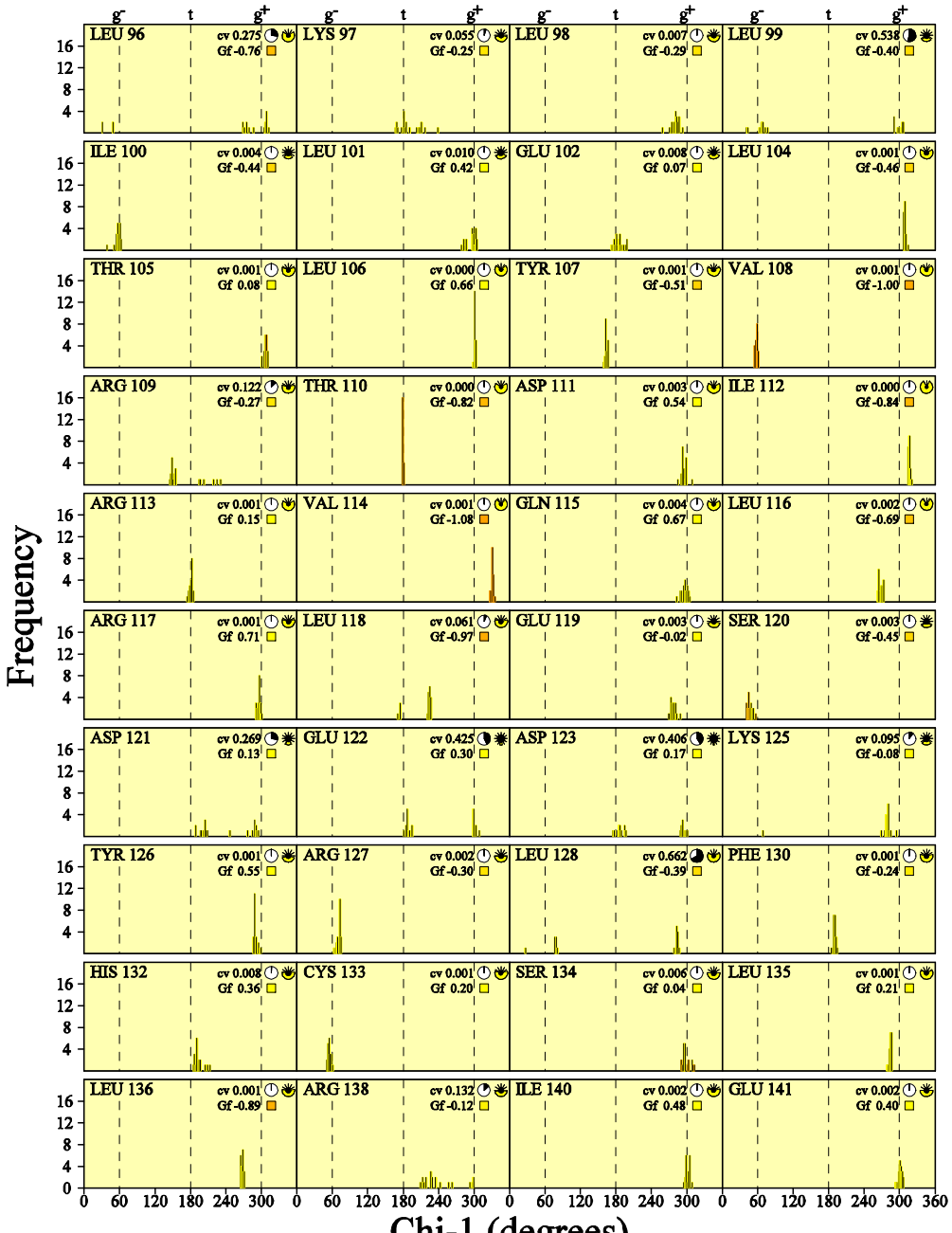
Chi-1 frequency distributions
ensemble (20 models)



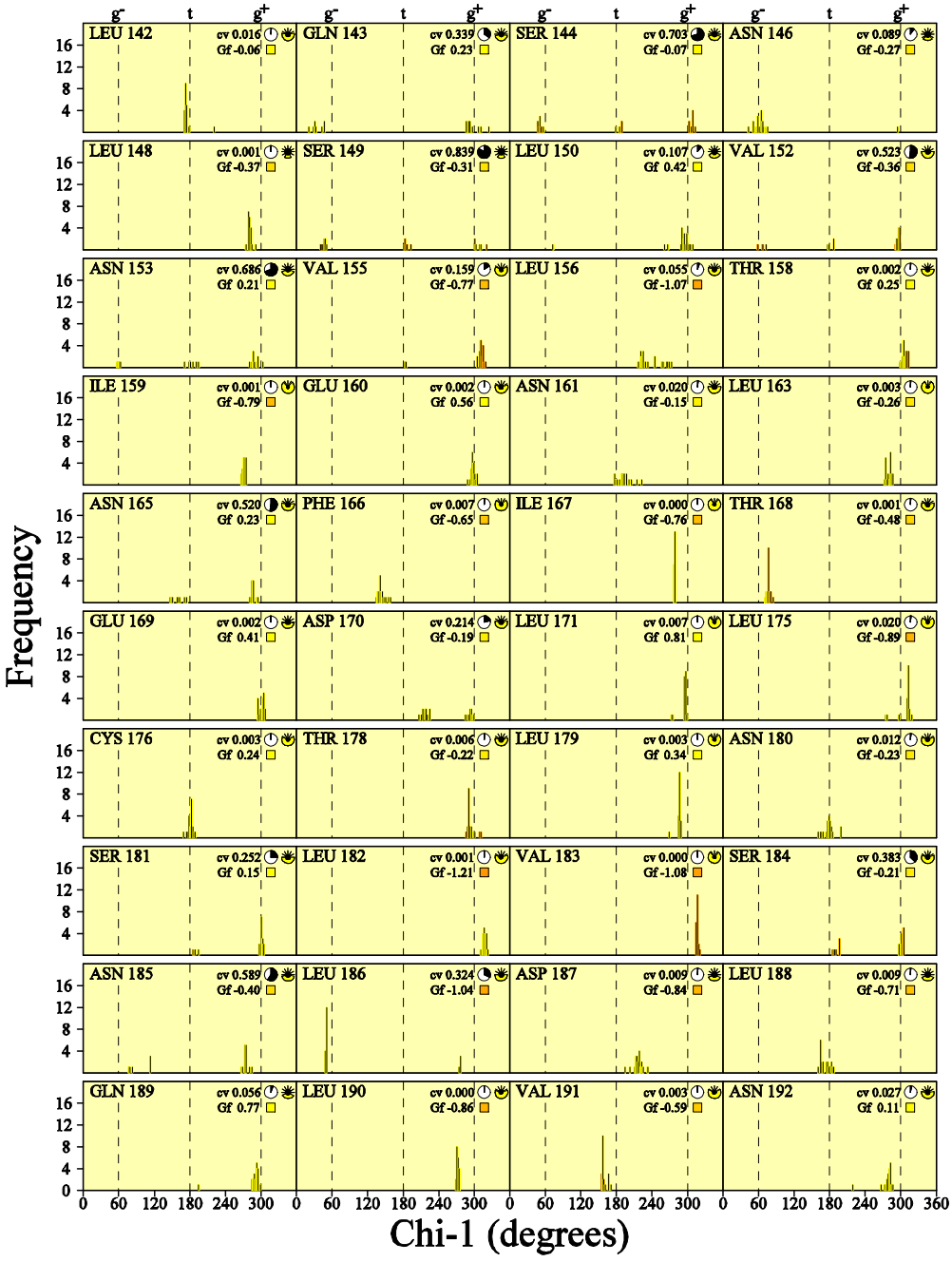
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data bars coloured according to G-factor: Favourable Unfavourable



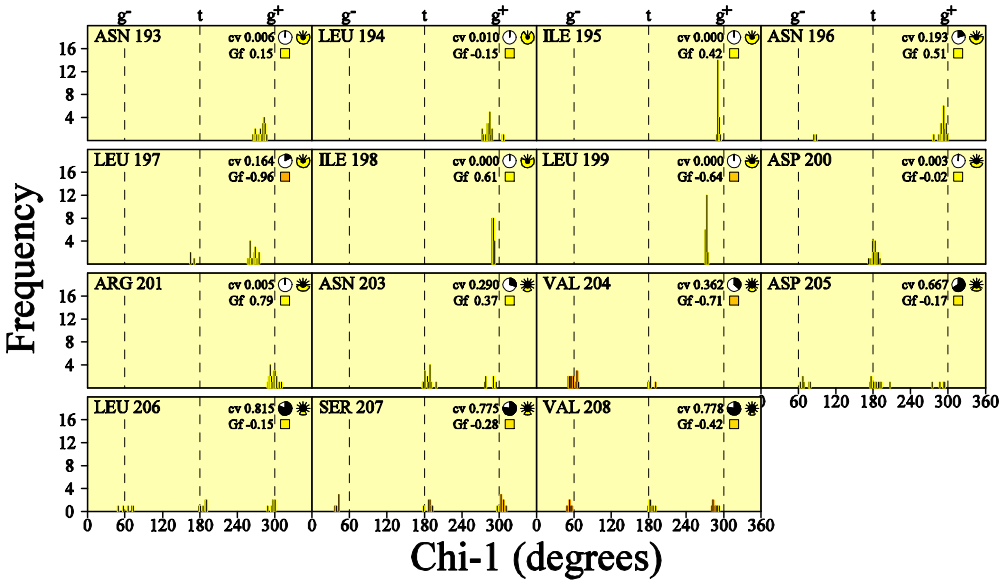
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data bars coloured according to G-factor: Favourable Unfavourable



cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data bars coloured according to G-factor: Favourable Unfavourable

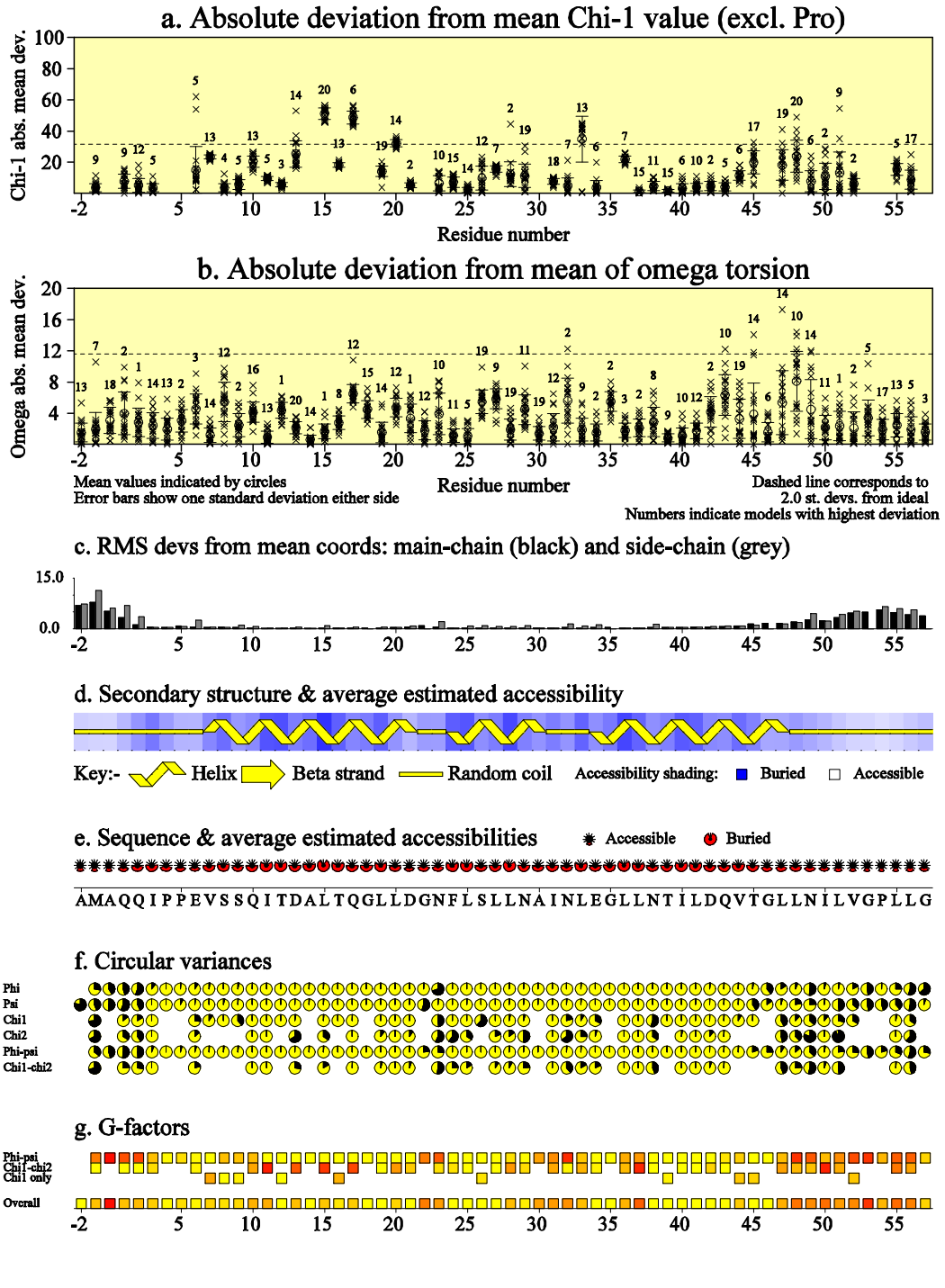


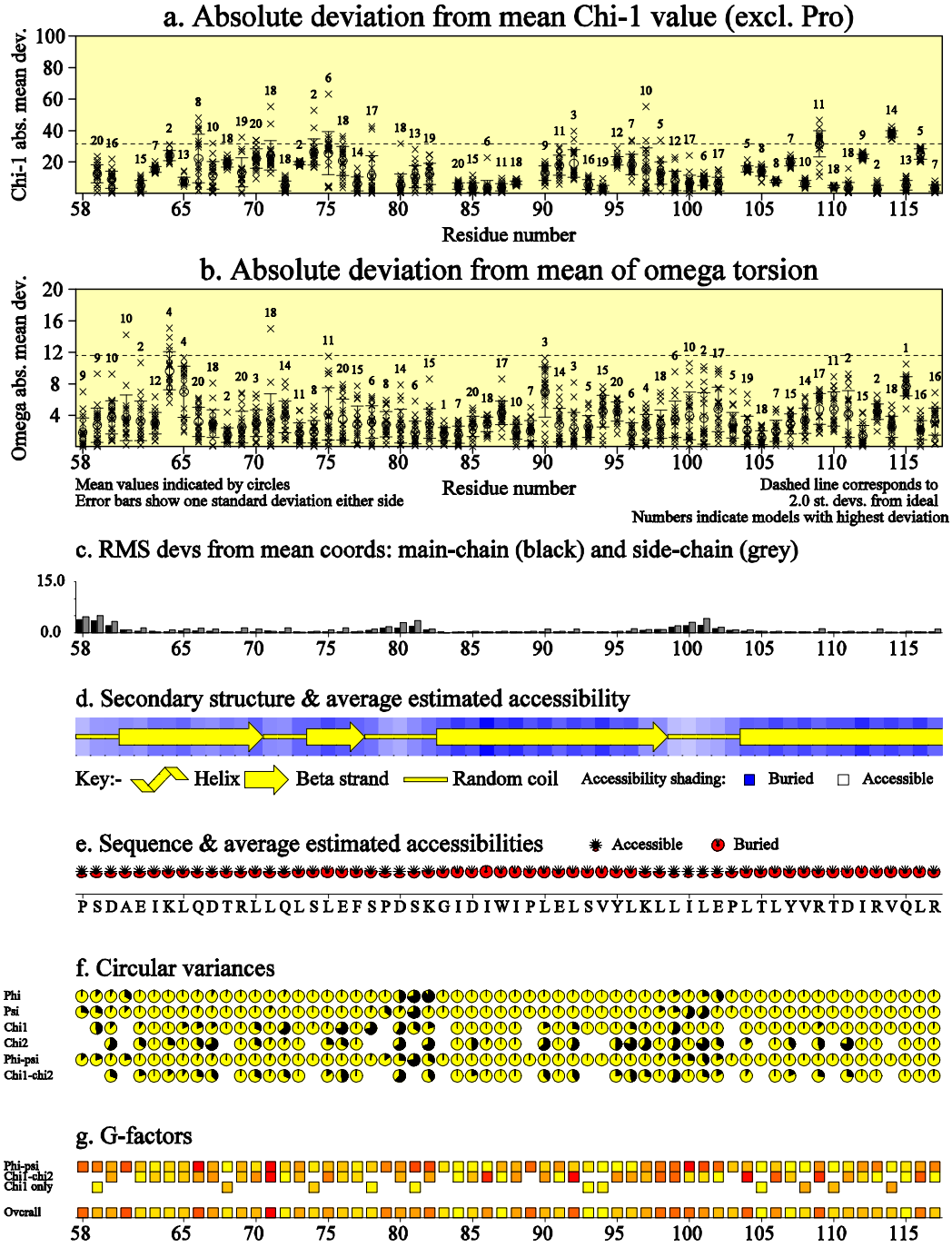
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data bars coloured according to G-factor: Favourable Unfavourable

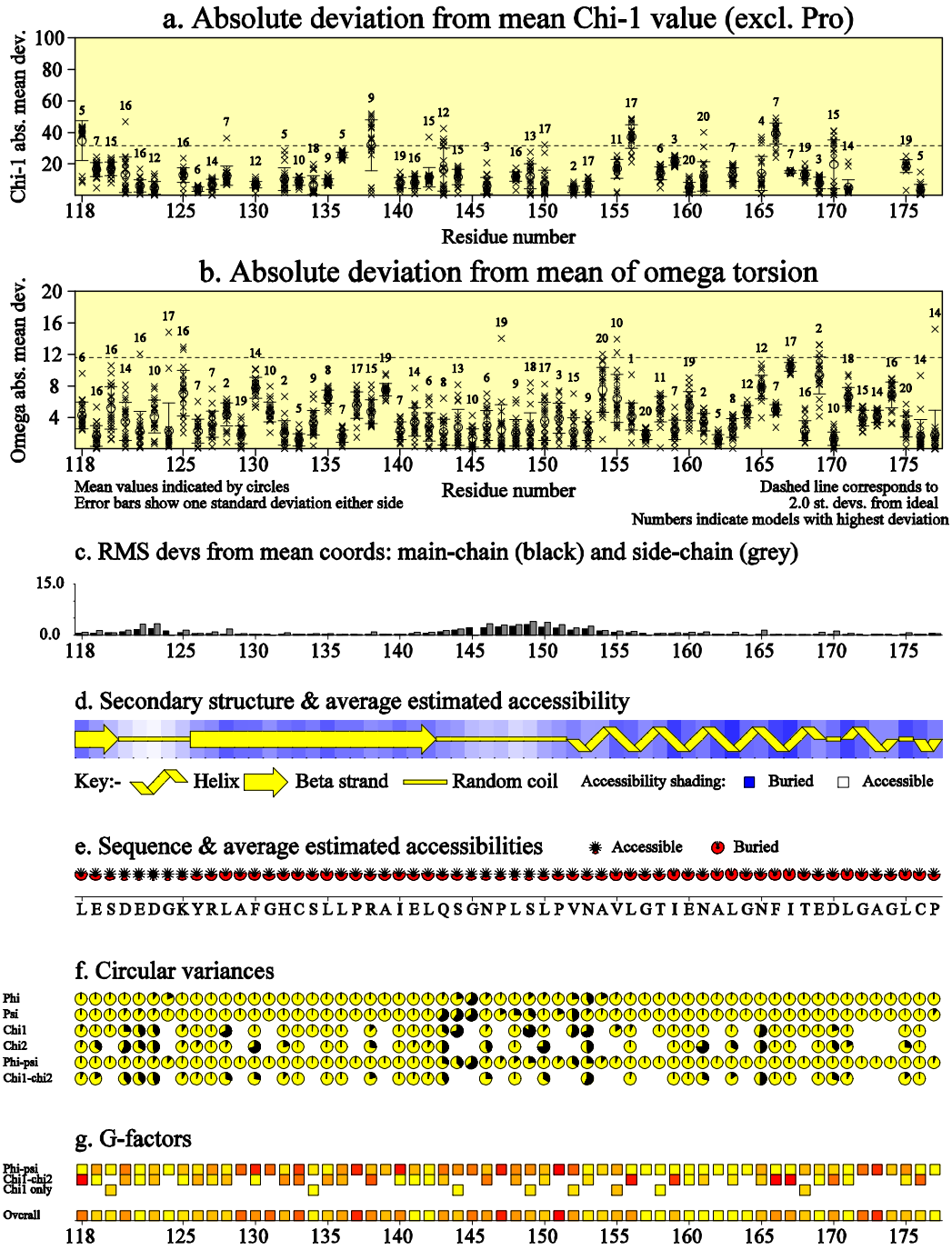


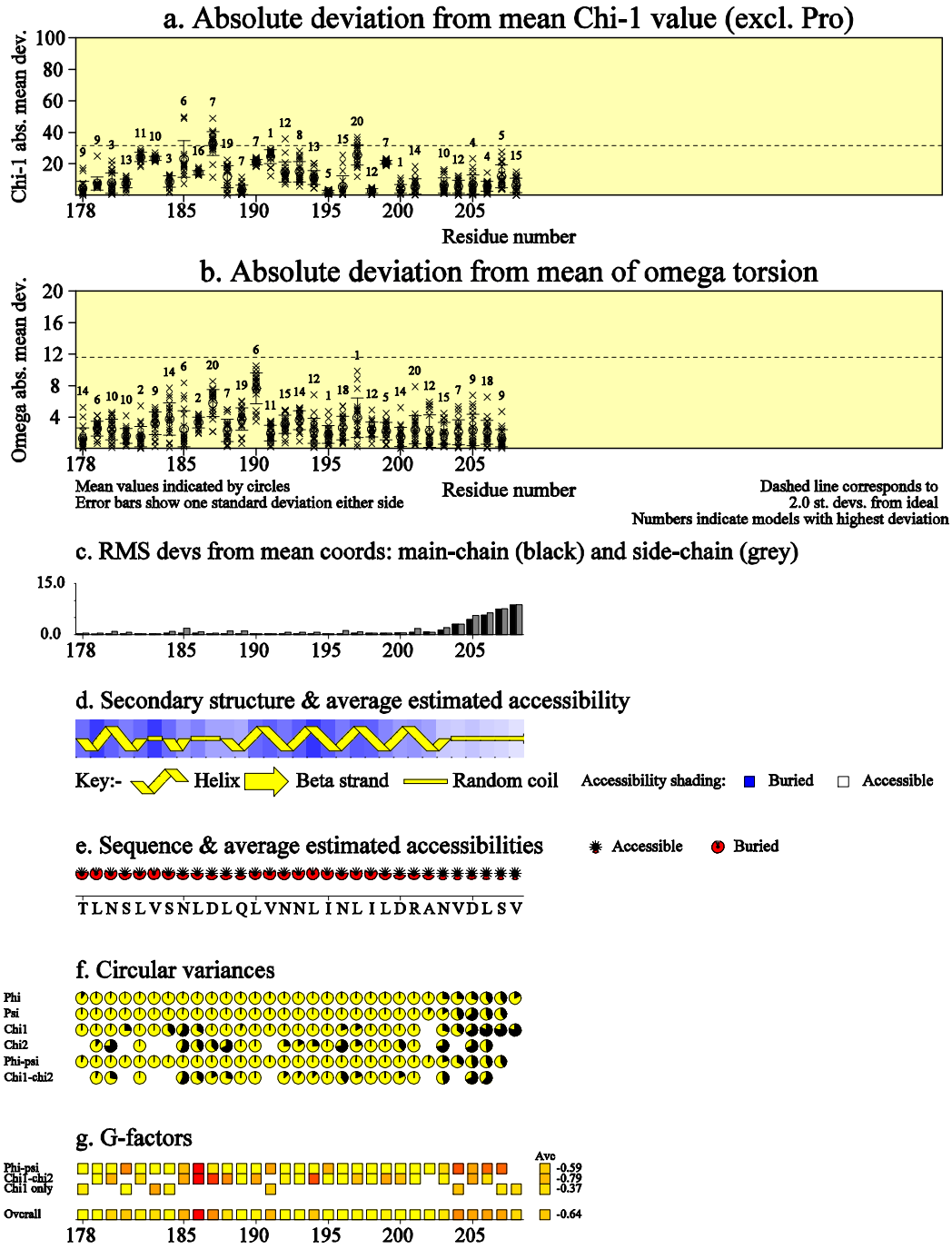
cv = Circular Variance (low values signify high clustering of the data points). Accessible Buried
Gf = Average G-factor for the residue (the higher the value the more favourable the conformations) based on analysis of high-res. Xstal structures
Data bars coloured according to G-factor: Favourable Unfavourable

Residue properties
ensemble (20 models)




a. Absolute deviation from mean Chi-1 value (excl. Pro)
Chi-1 abs. mean dev.
Residue number
b. Absolute deviation from mean of omega torsion
Omega abs. mean dev.
Residue number
Mean values indicated by circles
Error bars show one standard deviation either side
Dashed line corresponds to
2.0 st. devs. from ideal
Numbers indicate models with highest deviation
c. RMS devs from mean coords: main-chain (black) and side-chain (grey)
d. Secondary structure & average estimated accessibility
Key:- Helix Beta strand Random coil Accessibility shading: Buried Accessible
e. Sequence & average estimated accessibilities
Accessible
Buried
P S D A E I K L Q D T R L L Q L S L E F S P D S K G I D I W I P L E L S V Y L K L L I L E P L T L Y V R T D I R V Q L R
f. Circular variances
Phi
Psi
Chi1
Chi2
Phi-psi
Chi1-chi2
g. G-factors
Phi-psi
Chi1-chi2
Chi1 only
Overall


a. Absolute deviation from mean Chi-1 value (excl. Pro)
Chi-1 abs. mean dev.
Residue number
b. Absolute deviation from mean of omega torsion
Omega abs. mean dev.
Mean values indicated by circles
Error bars show one standard deviation either side
Residue number
Dashed line corresponds to
2.0 st. devs. from ideal
Numbers indicate models with highest deviation
c. RMS devs from mean coords: main-chain (black) and side-chain (grey)
d. Secondary structure & average estimated accessibility
Key:- Helix Beta strand Random coil Accessibility shading: Buried Accessible
e. Sequence & average estimated accessibilities
Accessible
Buried
L E S D E D G K Y R L A F G H C S L L P R A I E L Q S G N P L S L P V N A V L G T I E N A L G N F I T E D L G A G L C P
f. Circular variances
Phi
Psi
Chi1
Chi2
Phi-psi
Chi1-chi2
g. G-factors
Phi-psi
Chi1-chi2
Chi1 only
Overall


a. Absolute deviation from mean Chi-1 value (excl. Pro)
Chi-1 abs. mean dev.
Residue number
b. Absolute deviation from mean of omega torsion
Omega abs. mean dev.
Residue number
Mean values indicated by circles
Error bars show one standard deviation either side
Dashed line corresponds to
2.0 st. devs. from ideal
Numbers indicate models with highest deviation
c. RMS devs from mean coords: main-chain (black) and side-chain (grey)
d. Secondary structure & average estimated accessibility
Key:- Helix Beta strand Random coil Accessibility shading: Buried Accessible
e. Sequence & average estimated accessibilities
Accessible
Buried
TLNSLVSNLDLQLVNNLINLILDRANVDLSV
f. Circular variances
Phi
Psi
Chi1
Chi2
Phi-psi
Chi1-chi2
g. G-factors
Phi-psi
Chi1-chi2
Chi1 only
Overall
Ave
-0.59
-0.79
-0.37
-0.64

Appendix C- The Optimal ModelFree Parameters for Backbone Dynamics

Res	Model	S^2	ΔS^2	S^2_f	ΔS^2_f	τ_e (ps)	$\Delta \tau_e$	R_{ex} (ms)	ΔR_{ex}	SSE
3	5	0.622	0.021	0.831	0.016	1078.6	56.2			0.000
7	5	0.664	0.008	0.844	0.008	1299.7	76.2			0.000
8	5	0.789	0.008	0.912	0.007	1058.6	88.1			0.000
9	5	0.844	0.011	0.925	0.016	1381.7	236.7			0.000
11	2	0.915	0.007			32.0	6.8			0.042
12	4	0.860	0.008			20.3	4.4	1.718	0.214	0.000
13	1	0.930	0.005							1.930
14	1	0.887	0.005							1.093
16										0.000
18	2	0.904	0.003			9.8	2.0			2.321
20	4	0.849	0.007			15.4	2.1	1.138	0.247	0.000
22	2	0.858	0.004			17.2	2.6			0.544
25	2	0.927	0.006			45.0	14.0			3.045
26	3	0.867	0.009					1.907	0.423	3.056
27	5	0.874	0.006	0.906	0.005	1590.4	153.6			0.000
29	4	0.819	0.011			26.3	2.2	1.687	0.457	0.000
30	2	0.859	0.004			38.0	1.4			0.801
31	5	0.725	0.007	0.818	0.006	1330.7	113.3			0.000
34	4	0.859	0.006			16.8	1.0	1.584	0.196	0.000
35	1	0.870	0.038							4.892
37	1	0.939	0.014							1.008
39	2	0.899	0.006			18.7	4.5			1.985
40										0.000
41	3	0.883	0.003					1.902	0.346	0.018
43	4	0.853	0.013			16.5	4.0	0.793	0.281	0.000
44	4	0.944	0.027			46.8	222.7	4.312	0.958	0.000
45	4	0.879	0.016			26.8	6.4	4.108	0.585	0.000
47	5	0.824	0.021	0.878	0.014	1001.0	209.0			0.000
48	4	0.806	0.013			23.3	4.4	2.530	0.388	0.000
50	1	0.855	0.012							8.697
53	5	0.703	0.024	0.860	0.014	600.5	80.4			0.000
55	5	0.450	0.016	0.776	0.017	897.7	40.7			0.000
57	4	0.402	0.024			195.1	47.3	0.091	0.195	0.000
61	5	0.630	0.015	0.820	0.011	972.6	52.9			0.000
62	2	0.862	0.016			38.2	9.2			2.889
63	2	0.877	0.008			13.8	6.1			2.484
65	2	0.828	0.008			50.0	3.6			0.523
66	5	0.690	0.045	0.804	0.027	2176.8	283.8			0.000
68	5	0.706	0.008	0.862	0.007	1409.5	99.8			0.000
69	1	0.871	0.006							0.460
70	5	0.005	0.030	0.007	0.072	3074.3	836.3			0.000
71	5	0.769	0.012	0.830	0.009	1514.6	956.8			0.000
72	4	0.687	0.008			20.9	1.1	1.263	0.192	0.000
73	5	0.819	0.015	0.884	0.010	613.5	120.4			0.000

74	2	0.846	0.005			17.7	4.2			2.920
76	2	0.885	0.008			12.5	2.2			0.405
77	2	0.884	0.003			42.9	4.4			0.079
78	2	0.913	0.007			63.2	6.8			3.575
81	1	0.979	0.027							5.629
83	2	0.917	0.009			164.9	46.9			2.477
84	1	0.888	0.004							5.918
85	2	0.882	0.007			27.8	4.6			1.647
88	1	0.838	0.008							4.853
91	2	0.895	0.010			47.1	8.1			0.001
92	3	0.886	0.002					1.313	0.337	0.014
93	1	0.863	0.008							7.957
95	2	0.840	0.013			6.0	1.8			0.061
96	4	0.895	0.015			15.8	3.1	0.851	0.318	0.000
97	4	0.832	0.008			31.5	3.0	2.117	0.319	0.000
98	1	0.855	0.025							17.310
99	4	0.872	0.016			30.4	7.5	1.323	0.500	0.000
101	1	1.000	0.014							19.590
105	4	0.789	0.010			21.3	2.4	1.611	0.257	0.000
106	2	0.901	0.009			19.3	5.6			1.726
107	2	0.842	0.006			17.4	1.8			0.446
108	5	0.832	0.014	0.875	0.009	1408.0	234.8			0.000
110	1	0.907	0.011							2.323
111	1	0.881	0.013							2.845
113	1	0.915	0.012							1.844
114	1	0.879	0.008							0.404
115	2	0.862	0.005			15.8	2.8			0.833
116	2	0.882	0.013			18.4	2.9			2.653
118	4	0.852	0.008			1769.6	282.0			0.001
119	5	0.782	0.023	0.850	0.014	1495.2	275.2			0.000
120	1	0.813	0.021							1.781
121	5	0.627	0.014	0.788	0.010	862.9	46.9			0.000
123	2	0.745	0.009			74.2	4.5			0.451
124	5	0.699	0.011	0.824	0.008	761.4	47.5			0.000
126	2	0.736	0.006			54.9	2.6			0.294
127	5	0.823	0.009	0.874	0.007	612.4	114.3			0.000
128	5	0.797	0.008	0.859	0.005	769.9	87.6			0.000
129	2	0.840	0.010			48.5	7.4			1.301
131	5	0.868	0.009	0.903	0.007	746.4	216.9			0.000
136	1	0.862	0.009							18.820
138	5	0.734	0.013	0.820	0.010	1629.6	168.4			0.000
140	2	0.805	0.006			30.5	2.1			1.327
142	2	0.801	0.013			24.8	2.8			1.989
143	5	0.749	0.043	0.835	0.027	1555.5	377.8			0.000
152	4	0.905	0.045			98.9	219.1	7.192	1.931	0.000
155	4	0.863	0.066			75.1	252.5	26.878	6.357	0.000
157	3	0.691	0.118					6.088	2.527	0.019
159	1	0.947	0.011							0.261
160	3	0.873	0.012					1.761	0.460	1.949

161	4	0.868	0.019			16.8	3.1	2.196	0.395	0.000
163	2	0.905	0.008			11.5	1.7			3.687
164										0.000
165	4	0.876	0.014			21.3	2.9	2.328	0.449	0.000
168	4	0.817	0.007			11.6	2.4	1.440	0.307	0.000
169										0.000
171	2	0.873	0.007			36.1	3.7			0.004
172	2	0.866	0.009			65.2	6.6			0.293
174	1	0.909	0.009							16.040
176	3	0.872	0.005					1.138	0.359	0.822
178										0.000
180	4	0.870	0.008			27.9	5.7	1.277	0.329	0.000
181	4	0.802	0.028			25.5	8.7	4.610	0.992	0.000
183	2	0.907	0.008			25.6	4.6			0.076
184	4	0.856	0.011			18.5	3.1	2.195	0.281	0.000
186	5	0.751	0.004	0.829	0.004	1481.7	116.8			0.000
188	2	0.867	0.011			50.8	8.4			1.272
189	4	0.851	0.008			43.8	4.2	0.880	0.178	0.000
191	2	0.901	0.008			41.1	9.2			2.671
192	4	0.910	0.012			22.9	5.9	0.884	0.262	0.000
193	4	0.874	0.006			43.7	3.6	0.505	0.149	0.000
194	4	0.870	0.009			31.6	3.2	0.875	0.359	0.000
195	2	0.890	0.008			29.5	3.3			0.536
196	4	0.879	0.004			36.2	3.4	0.957	0.206	0.000
197										14.350
200	2	0.893	0.005			38.1	3.5			0.169
201	5	0.770	0.011	0.902	0.007	887.1	49.7			0.000
202	5	0.618	0.011	0.861	0.008	1110.8	43.1			0.000
203										2.055
204	4	0.314	0.009			317.3	5.9	1.002	0.116	0.000
208										0.000

References

1. Cooper A, Kennedy MW (2009) Biofoams and natural protein surfactants. Biophysical Chemistry 151: 96-104.
2. Linder MB (2009) Hydrophobins: Proteins that self assemble at interfaces. Current Opinion in Colloid & Interface Science 14: 356-363.
3. Szent-Gyorgyi A (1979) Cell-Associated Water. Academic Press.
4. Stillinger FH (1980) Water Revisited. Science 209: 451-457.
5. Myers D (2005) Surfactant Science and Technology: John Wiley & Sons, Inc.
6. Rosen MJ (2004) Surfactants and Interfacial Phenomena: John Wiley & Sons, Inc.
7. Flick EW (1993) Industrial surfactants: Noyes Publications.
8. Acmite Market Intelligence (2010) Market Report: World Surfactant Market.
9. Schramm LL (2001) Dictionary of Colloid and Interface Science: John Wiley & Sons,.
10. Chistyakov BE (2001) 6. Theory and practical application aspects of surfactants. In: V.B. Fainerman DM, Miller R, editors. Studies in Interface Science: Elsevier. pp. 511-618.
11. Schramm LL, Stasiuk EN, Marangoni DG (2003) 2 Surfactants and their applications. Annual Reports Section "C" (Physical Chemistry) 99: 3-48.
12. Jones MN (1999) Surfactants in membrane solubilisation. International Journal of Pharmaceutics 177: 137-159.
13. Almgren M (2000) Mixed micelles and other structures in the solubilization of bilayer lipid membranes by surfactants. Biochimica et Biophysica Acta (BBA) - Biomembranes 1508: 146-163.
14. Scott MJ, Jones MN (2000) The biodegradation of surfactants in the environment. Biochimica Et Biophysica Acta-Biomembranes 1508: 235-251.
15. Ying GG (2006) Fate, behavior and effects of surfactants and their degradation products in the environment. Environment International 32: 417-431.
16. Banat IM, Makkar RS, Cameotra SS (2000) Potential commercial applications of microbial surfactants. Applied Microbiology and Biotechnology 53: 495-508.
17. Lu JR, Zhao XB, Yaseen M (2007) Biomimetic amphiphiles: Biosurfactants. Current Opinion in Colloid & Interface Science 12: 60-67.
18. Arima K, Kakinuma A, Tamura G (1968) Surfactin a Crystalline Peptidelipid Surfactant Produced by Bacillus Subtilis- Isolation Characterisation and its Inhibition of Fibrin Clot Formation. Biochemical and Biophysical Research Communications 31: 488-494.
19. Yeh MS, Wei YH, Chang JS (2005) Enhanced production of surfactin from Bacillus subtilis by addition of solid carriers. Biotechnology Progress 21: 1329-1334.
20. Peypoux F, Bonmatin JM, Wallach J (1999) Recent trends in the biochemistry of surfactin. Applied Microbiology and Biotechnology 51: 553-563.
21. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. Trends in Biochemical Sciences 25: 331-339.
22. Dill KA (1990) Dominant Forces in Protein Folding. Biochemistry 29: 7133-7155.
23. Clarkson JR, Cui ZF, Darton RC (1999) Protein Denaturation in Foam: I. Mechanism Study. Journal of Colloid and Interface Science 215: 323-332.
24. Campbell GM, Mougeot E (1999) Creation and characterisation of aerated food products. Trends in Food Science & Technology 10: 283-296.
25. Goerke J (1998) Pulmonary surfactant: functions and molecular composition. Biochimica Et Biophysica Acta-Molecular Basis of Disease 1408: 79-89.
26. Kishore U, Greenhough TJ, Waters P, Shrive AK, Ghai R, et al. (2006) Surfactant proteins SP-A and SP-D: Structure, function and receptors. Molecular Immunology 43: 1293-1315.
27. Kuroki Y, Takahashi M, Nishitani C (2007) Pulmonary collectins in innate immunity of the lung. Cellular Microbiology 9: 1871-1879.
28. Melton KR, Nesslein LL, Ikegami M, Tichelaar JW, Clark JC, et al. (2003) SP-B deficiency causes respiratory failure in adult mice. American Journal of Physiology-Lung Cellular and Molecular Physiology 285: L543-L549.
29. Glasser SW, Detmer EA, Ikegami M, Na CL, Stahlman MT, et al. (2003) Pneumonitis and emphysema in sp-C gene targeted mice. Journal of Biological Chemistry 278: 14291-14298.
30. Weaver TE, Conkright JJ (2001) Functions of surfactant proteins B and C. Annual Review of Physiology 63: 555-578.
31. Zaltash S, Palmblad M, Curstedt T, Johansson J, Persson B (2000) Pulmonary surfactant protein B: a structural model and a functional analogue. Biochimica Et Biophysica Acta-Biomembranes 1466: 179-186.

32. Shanmukh S, Howell P, Baatz JE, Dluhy RA (2002) Effect of hydrophobic surfactant proteins SP-B and SP-C on phospholipid monolayers. Protein structure studied using 2D IR and beta nu correlation analysis. Biophysical Journal 83: 2126-2141.
33. Perez-Gil J (2008) Structure of pulmonary surfactant membranes and films: The role of proteins and lipid-protein interactions. Biochimica Et Biophysica Acta-Biomembranes 1778: 1676-1695.
34. Johansson J, Szyperski T, Curstedt T, Wuthrich K (1994) The NMR Structure of the Pulmonary Surfactant-Associated Polypeptide SP-C in an Apolar Solvent Contains a Valyl-rich Alpha-helix. Biochemistry 33: 6015-6023.
35. Vandenbussche G, Clercx A, Curstedt T, Johansson J, Jornvall H, et al. (1992) Structure and Orientation of the Surfactant-associated Protein-C in a Lipid Bilayer. European Journal of Biochemistry 203: 201-209.
36. Plasencia I, Cruz A, Casals C, Perez-Gil J (2001) Superficial disposition of the N-terminal region of the surfactant protein SP-C and the absence of specific SP-B-SP-C interactions in phospholipid bilayers. Biochemical Journal 359: 651-659.
37. Oosterlakendijksterhuis MA, Haagsman HP, Vangolde LMG, Demel RA (1991) Characterization of Lipid Insertion into Monomolecular Layers Mediated by Lung Surfactant Protein SP-B and SP-C. Biochemistry 30: 10965-10971.
38. Stahlman MT, Gray MP, Falconieri MW, Whitsett JA, Weaver TE (2000) Lamellar body formation in normal and surfactant protein B-deficient fetal mice. Laboratory Investigation 80: 395-403.
39. Foster CD, Zhang PX, Gonzales LW, Guttentag SH (2003) In vitro surfactant protein B deficiency inhibits lamellar body formation. American Journal of Respiratory Cell and Molecular Biology 29: 259-266.
40. Vorbroker DK, Profitt SA, Nogee LM, Whitsett JA (1995) Aberrant Processing of Surfactant Protein-C in Hereditary SP-B Deficiency. American Journal of Physiology-Lung Cellular and Molecular Physiology 268: L647-L656.
41. Linder MB, Szilvay GR, Nakari-Setala T, Penttila ME (2005) Hydrophobins: the protein-amphiphiles of filamentous fungi. Fems Microbiology Reviews 29: 877-896.
42. Wosten HAB, van Wetter MA, Lugones LG, van der Mei HC, Busscher HJ, et al. (1999) How a fungus escapes the water to grow into the air. Current Biology 9: 85-88.
43. Kershaw MJ, Talbot NJ (1998) Hydrophobins and Repellents: Proteins with Fundamental Roles in Fungal Morphogenesis. Fungal Genetics and Biology 23: 18-33.
44. Schuren FHJ, Wessels JGH (1990) 2 Genes Specifically Expressed in Fruiting Dikaryons of Schizopyllum-Commune - Homoligies with a Gene Not Regulated by Mating-Type Genes. Gene 90: 199-205.
45. Wessels JGH (1994) Developmental Regulation of Fungal Cell-wall Formation. Annual Review of Phytopathology 32: 413-437.
46. Hakanpää J, Paananen A, Askolin S, Nakari-Setälä T, Parkkinen T, et al. (2004) Atomic Resolution Structure of the HFBII Hydrophobin, a Self-assembling Amphiphile. Journal of Biological Chemistry 279: 534-539.
47. The PyMOL Molecular Graphics System. 1.3 ed: Schrodinger, LLC.
48. Wang XQ, Graveland-Bikker JF, De Kruif CG, Robillard GT (2004) Oligomerization of hydrophobin SC3 in solution: From soluble state to self-assembly. Protein Science 13: 810-821.
49. Hektor HJ, Scholtmeijer K (2005) Hydrophobins: proteins with potential. Current Opinion in Biotechnology 16: 434-439.
50. Lu JR, Thomas RK (1998) Neutron reflection from wet interfaces. Journal of the Chemical Society-Faraday Transactions 94: 995-1018.
51. Penfold J, Richardson RM, Zarbakhsh A, Webster JRP, Bucknall DG, et al. (1997) Recent advances in the study of chemical surfaces and interfaces by specular neutron reflection. Journal of the Chemical Society, Faraday Transactions 93: 3899-3917.
52. Mendelsohn R, Flach CR (2006) Infrared Reflection – Absorption Spectrometry of Monolayer Films at the Air – Water Interface. Handbook of Vibrational Spectroscopy: John Wiley & Sons, Ltd.
53. Tarano Z (1998) Cover and Ambient Light Influence Nesting Preferences in the Tungara Frog Physalaemus Pustulosus. Copeia: 250-251.
54. Ryan MJ (1985) The Tungara Frog- A Study in Sexual Selection and Communication: University of Chicago Press.
55. Fleming RI, Mackenzie CD, Cooper A, Kennedy MW (2009) Foam nest components of the túngara frog: a cocktail of proteins conferring physical and biological resilience. Proceedings of the Royal Society B: Biological Sciences 276: 1787-1795.
56. Cooper A, Kennedy MW, Fleming RI, Wilson EH, Videler H, et al. (2005) Adsorption of Frog Foam Nest Proteins at the Air-Water Interface. Biophysical Journal 88: 2114-2125.

57. Bode W, Engh R, Musil D, Thiele U, Huber R, et al. (1988) The 2.0 A X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. European Molecular Biology Organization Journal 7: 2593-2599.
58. CLC Bio A/S, Aarhus, Denmark.
59. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404-405.
60. Kyte J, Doolittle RF (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. Journal of Molecular Biology 157: 105-132.
61. Wilkins MR, Gasteiger E, Bairoch A, Sanchez J-C, Williams KL, et al. (1999) Protein identification and analysis tools in the ExPASy server. Methods in Molecular Biology; 2-D proteome analysis protocols: Humana Press Inc. {a}. pp. 531-552.
62. Mackenzie CD, Smith BO, Meister A, Blume A, Zhao X, et al. (2009) Ranaspumin-2: Structure and Function of a Surfactant Protein from the Foam Nests of a Tropical Frog. Biophysical Journal 96: 4984-4992.
63. Wendell D, Todd J, Montemagno C (2010) Artificial Photosynthesis in Ranaspumin-2 Based Foam. Nano Letters 10: 3231-3236.
64. Miyake K, Tanaka T, McNeil PL (2007) Lectin-Based Food Poisoning: A New Mechanism of Protein Toxicity. PLoS ONE 2.
65. Brody S, Lardy HA (1946) Bioenergetics and Growth. The Journal of Physical Chemistry 50: 168-169.
66. Terrien J, Perret M, Aujard F (2011) Behavioral thermoregulation in mammals: a review. Frontiers in Bioscience-Landmark 16: 1428-1444.
67. McConaghy F. (1994) Thermoregulation. The Athletic Horse: Principles and Practice of Equine Sports Medicine: W.B, Saunders Company. pp. 181-202.
68. Jenkinson DM, Elder HY, Bovell DL (2007) Equine sweating and anhidrosis Part 2: anhidrosis. Veterinary Dermatology 18: 2-11.
69. Eckersall PD, Beeley JG, Snow DH, Thomas A (1984) Characterization of Glycoproteins in the Sweat of the Horse (Equus-caballus). Research in Veterinary Science 36: 231-234.
70. Beeley JG, Eason R, Snow DH (1986) Isolation and Characterization of latherin, a surface-active protein from horse sweat. Biochem Journal 235: 645-650.
71. Jenkinson DM, Elder HY, Bovell DL (2006) Equine sweating and anhidrosis Part 1 - equine sweating. Veterinary Dermatology 17: 361-392.
72. Eckersall PD, Kerr MG, Snow DH (1982) An investigation into the proteins of horse sweat (Equus caballus). Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 73: 375-378.
73. McDonald RE, Fleming RI, Beeley JG, Bovell DL, Lu JR, et al. (2009) Latherin: A Surfactant Protein of Horse Sweat and Saliva. PLoS ONE 4: e5726.
74. Bingle CD, Bingle L, Craven CJ (2011) Proteins with a BPI/LBP/PLUNC-Like Domain: Revisiting the Old and Characterizing the New. Biochemical Society Transactions 39: 961-965.
75. Egland KA, Vincent JJ, Strausberg R, Lee B, Pastan I (2003) Discovery of the breast cancer gene BASE using a molecular approach to enrich for genes encoding membrane and secreted proteins. Proceedings of the National Academy of Sciences of the United States of America 100: 1099-1104.
76. Bingle CD, Gorr SU (2004) Host defense in oral and airway epithelia: chromosome 20 contributes a new protein family. International Journal of Biochemistry & Cell Biology 36: 2144-2152.
77. Hahn Y, Lee B (2005) Identification of nine human-specific frameshift mutations by comparative analysis of the human and the chimpanzee genome sequences. Bioinformatics 21: I186-I194.
78. Bingle CD, LeClair EE, Havard S, Bingle L, Gillingham P, et al. (2004) Phylogenetic and evolutionary analysis of the PLUNC gene family. Protein Science 13: 422-430.
79. Schumann R, Leong, Flaggs G, Gray P, Wright S, et al. (1990) Structure and function of lipopolysaccharide binding protein. Science 249: 1429-1431.
80. Beamer LJ, Carroll SF, Eisenberg D (1997) Crystal Structure of Human BPI and Two Bound Phospholipids at 2.4 Angstrom Resolution. Science 276: 1861-1864.
81. Beamer LJ, Carroll SF, Eisenberg D (1998) The BPI/LBP family of proteins: A structural analysis of conserved regions. Protein Science 7: 906-914.
82. Alexander C, Rietschel ET (2001) Invited review: Bacterial lipopolysaccharides and innate immunity. pp. 167-202.
83. Dellinger RP (2003) Inflammation and coagulation: Implications for the septic patient. Clinical Infectious Diseases 36: 1259-1265.

84. Marra MN, Wilde CG, Griffith JE, Snable JL, Scott RW (1990) Bactericidal Permeability-Increasing Protein Has Endotoxin-Neutralizing Activity. Journal of Immunology 144: 662-666.
85. Weiss J, Elsbach P, Shu C, Castillo J, Grinna L, et al. (1992) Human Bactericidal Permeability-Increasing Protein and a Recombinant Nh2-Terminal Fragment Cause Killing of Serum-Resistant Gram-Negative Bacteria in Whole-Blood and Inhibit Tumor-Necrosis-Factor Release Induced by the Bacteria. Journal of Clinical Investigation 90: 1122-1130.
86. Gazzanosantoro H, Parent JB, Grinna L, Horwitz A, Parsons T, et al. (1992) High-Affinity Binding of the Bactericidal Permeability-Increasing Protein and a Recombinant Amino-Terminal Fragment to the Lipid-a Region of Lipopolysaccharide. Infection and Immunity 60: 4754-4761.
87. Kleiger G, Beamer LJ, Grothe R, Mallick P, Eisenberg D (2000) The 1.7 Å crystal structure of BPI: a study of how two dissimilar amino acid sequences can adopt the same fold. Journal of Molecular Biology 299: 1019-1034.
88. Beamer LJ, Carroll SF, Eisenberg D (1999) The three-dimensional structure of human bactericidal/permeability-increasing protein: Implications for understanding protein-lipopolysaccharide interactions. Biochemical Pharmacology 57: 225-229.
89. Schultz H, Weiss JP (2007) The bactericidal/permeability-increasing protein (BPI) in infection and inflammatory disease. Clinica Chimica Acta 384: 12-23.
90. Gazzanosantoro H, Parent JB, Conlon PJ, Kasler HG, Tsai CM, et al. (1995) Characterization of the Structural Elements in Lipid-a Required for Binding of a Recombinant Fragment of Bactericidal/Permeability-Increasing Protein Rbpi(23). Infection and Immunity 63: 2201-2205.
91. Lamping N, Hoess A, Yu B, Park TC, Kirschning CJ, et al. (1996) Effects of site-directed mutagenesis of basic residues (Arg 94, Lys 95, Lys 99) of lipopolysaccharide (LPS)-binding protein on binding and transfer of LPS and subsequent immune cell activation. Journal of Immunology 157: 4648-4656.
92. Domingues MM, Lopes S, Santos NC, Quintas A, Castanho M (2009) Fold-Unfold Transitions in the Selectivity and Mechanism of Action of the N-Terminal Fragment of the Bactericidal/Permeability-Increasing Protein (rBPI(21)). Biophysical Journal 96: 987-996.
93. Lechner AJ, Lamprech KE, Johanns CA, Matuschak GM (1995) The Recombinant 23-Kda N-Terminal Fragment of Bactericidal Permeability-Increasing Protein (Rbpi(23)) Decreases Escherichia Coli-Induced Mortality and Organ Injury During Immunosuppression-Related Neutropenia. Shock 4: 298-306.
94. Mahley RW, Innerarity TL, Rall SC, Weisgraber KH (1984) Plasma-Lipoproteins - Apolilpoprotein Structure and Function. Journal of Lipid Research 25: 1277-1294.
95. Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, et al. (2005) Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet 366: 1267-1278.
96. Desforges JF, Gordon DJ, Rifkind BM (1989) High-Density Lipoprotein — The Clinical Implications of Recent Studies. New England Journal of Medicine 321: 1311-1316.
97. Barter P, Gotto AM, LaRosa JC, Maroni J, Szarek M, et al. (2007) HDL Cholesterol, Very Low Levels of LDL Cholesterol, and Cardiovascular Events. New England Journal of Medicine 357: 1301-1310.
98. Drayna D, Jarnagin AS, McLean J, Henzel W, Kohr W, et al. (1987) Cloning and sequencing of human cholesteryl ester transfer protein cDNA. Nature 327: 632-634.
99. Pape ME, Ulrich RG, Rea TJ, Marotti KR, Melchior GW (1991) Evidence that the nonparenchymal cells of the liver are the principal source of cholesteryl ester transfer protein in primates. Journal of Biological Chemistry 266: 12829-12831.
100. Gakhar L, Bartlett JA, Penterman J, Mizrachi D, Singh PK, et al. (2010) PLUNC Is a Novel Airway Surfactant Protein with Anti-Biofilm Activity. PLoS ONE 5: e9098.
101. Brown ML, Inazu A, Hesler CB, Agellon LB, Mann C, et al. (1989) Molecular basis of lipid transfer protein deficiency in a family with increased high-density lipoproteins. Nature 342: 448-451.
102. Sikorski JA (2006) Cholesteryl ester transfer protein inhibitors as potential new therapies for coronary artery disease. Expert Opinion on Therapeutic Patents 16: 753-772.
103. Day JR, Albers JJ, Lofton-Day CE, Gilbert TL, Ching AF, et al. (1994) Complete cDNA encoding human phospholipid transfer protein from human endothelial cells. pp. 9388-9391.
104. Tall AR, Krumholz S, Olivercrona T, Deckelbaum RJ (1985) Plasma Phospholipid Transfer Protein Enhances Transfer and Exchange of Phospholipids between Very Low-Density Lipoproteins and High-Density Lipoproteins During Lipolysis. Journal of Lipid Research 26: 842-851.

105. Jauhiainen M, Metso J, Pahlman R, Blomqvist S, van Tol A, et al. (1993) Human plasma phospholipid transfer protein causes high density lipoprotein conversion. pp. 4032-4036.
106. Settasatian N, Duong M, Curtiss LK, Ehnholm C, Jauhiainen M, et al. (2001) The Mechanism of the Remodeling of High Density Lipoproteins by Phospholipid Transfer Protein. Journal of Lipid Research 276: 26898-26905.
107. Hailman E, Albers JJ, Wolfbauer G, Tu AY, Wright SD (1996) Neutralization and transfer of lipopolysaccharide by phospholipid transfer protein. Journal of Biological Chemistry 271: 12172-12178.
108. Jiang XC, Tall AR, Qin S, Lin M, Schneider M, et al. (2002) Phospholipid transfer protein deficiency protects circulating lipoproteins from oxidation due to the enhanced accumulation of vitamin E. Journal of Biological Chemistry 277: 31850-31856.
109. Lagrost L, Desrumaux C, Masson D, Deckert V, Gambert P (1998) Structure and function of the plasma phospholipid transfer protein. Current Opinion in Lipidology 9: 203-209.
110. Qiu X, Mistry A, Ammirati MJ, Chrunyk BA, Clark RW, et al. (2007) Crystal structure of cholesteryl ester transfer protein reveals a long tunnel and four bound lipid molecules. Nat Struct Mol Biol 14: 106-113.
111. Huuskonen J, Wohlfahrt G, Jauhiainen M, Ehnholm C, Teleman O, et al. (1999) Structure and phospholipid transfer activity of human PLTP: analysis by molecular modeling and site-directed mutagenesis. Journal of Lipid Research 40: 1123-1130.
112. Huuskonen J, Jauhiainen M, Ehnholm C, Olkkonen VM (1998) Biosynthesis and secretion of human plasma phospholipid transfer protein. Journal of Lipid Research 39: 2021-2030.
113. Wang S, Kussie P, Deng L, Tall A (1995) Defective Binding of Neutral Lipids by a Carboxyl-terminal Deletion Mutant of Cholesteryl Ester Transfer Protein. Journal Of Biological Chemistry 270: 612-618.
114. Desrumaux C, Labeur C, Verhee A, Tavernier J, Vandekerckhove J, et al. (2001) A hydrophobic cluster at the surface of the human plasma phospholipid transfer protein is critical for activity on high density lipoproteins. Journal of Biological Chemistry 276: 5908-5915.
115. Bingle CD, Craven CJ (2002) PLUNC: A novel family of candidate host defence proteins expressed in the upper airways and nasopharynx. Human Molecular Genetics 11: 937-943.
116. Bingle CD, Craven CJ (2004) Meet the relatives: a family of BPI- and LBP-related proteins. Trends in Immunology 25: 53-55.
117. Vitorino R, Lobo MJC, Ferrer-Correira AJ, Dubin JR, Tomer KB, et al. (2004) Identification of human whole saliva protein components using proteomics. PROTEOMICS 4: 1109-1115.
118. Ghafouri B, Kihlstrom E, Stahlbom B, Tagesson C, Lindahl M (2003) PLUNC (palate, lung and nasal epithelial clone) proteins in human nasal lavage fluid. Biochemical Society Transactions 31: 810-814.
119. Campos MA, Abreu AR, Nlend MC, Cobas MA, Conner GE, et al. (2004) Purification and characterization of PLUNC from human - Tracheobronchial secretions. American Journal of Respiratory Cell and Molecular Biology 30: 184-192.
120. Chu HW, Thaikoottathil J, Rino JG, Zhang G, Wu Q, et al. (2007) Function and regulation of SPLUNC1 protein in mycoplasma infection and allergic inflammation. Journal of Immunology 179: 3995-4002.
121. Bartlett JA, Hicks BJ, Schlomann JM, Ramachandran S, Nauseef WM, et al. (2008) PLUNC is a secreted product of neutrophil granules. Journal of Leukocyte Biology 83: 1201-1206.
122. Ghafouri B, Kihlstrom E, Tagesson C, Lindahl M (2004) PLUNC in human nasal lavage fluid: multiple isoforms that bind to lipopolysaccharide. Biochimica Et Biophysica Acta-Proteins and Proteomics 1699: 57-63.
123. Gil J, Weibel ER (1971) Extracellular Lining of Bronchioles after Perfusion-Fixation of Rat Lungs for Electron Microscopy. Anatomical Record 169: 185-199.
124. Khoor A, Stahlman MT, Gray ME, Whitsett JA (1994) Temporal-Spatial Distribution of Sp-B and Sp-C Proteins and Messenger-Rnas in Developing Respiratory Epithelium of Human Lung. Journal of Histochemistry & Cytochemistry 42: 1187-1199.
125. Goubran Botros H, Poncet P, Rabillon J, Fontaine T, Laval JM, et al. (2001) Biochemical characterization and surfactant properties of horse allergens. European journal of biochemistry / FEBS 268: 3126-3136.
126. Kennedy MW, Beauchamp J (2000) Sticky-finger interaction sites on cytosolic lipid-binding proteins? Cellular and Molecular Life Sciences 57: 1379-1387.
127. Shuman S (1994) Novel-Approach to Molecular-Cloning and Polynucleotide Synthesis Using Vaccinia DNA Topoisomerase. Journal of Biological Chemistry 269: 32678-32684.
128. Sorensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. Journal of Biotechnology 115: 113-128.

129. Hochuli E, Bannwarth W, Dobeli H, Gentz R, Stuber D (1988) Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. Nat Biotech 6: 1321-1325.
130. Mackenzie CD (2007) The Biophysical Chemistry of Frog Foam Nest Proteins. Glasgow: University of Glasgow.
131. Wada KN, Wada Y, Ishibashi F, Gojobori T, Ikemura T (1992) Codon usage tabulated from the Genbank genetic sequence data. Nucleic Acids Research 20: 2111-2118.
132. Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Current Opinion in Biotechnology 6: 494-500.
133. Novagen Novagen Competent Cells Brochure.
134. Hans Peter Sørensen KKM (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. Journal of Biotechnology 115: 113-128.
135. Raab D, Graf M, Notka F, Schödl T, Wagner R (2010) The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. Systems and Synthetic Biology 4: 215-225.
136. Alibolandi M, Mirzahoseini H, Moshtaghi Nehi F, Tabatabaian G, Amini H, et al. (2010) Improving recombinant protein solubility in Escherichia coli: Identification of best chaperone combination which assists folding of human basic fibroblast growth factor. AJB 9: 8100-8109.
137. Deborah A. Berthold VJJ, Terry L. Sasser, John J. Shea, Heather L. Frericks, Gautam Shah, and Chad M. Rienstra Top Ten Tips for Producing 13C 15N Protein in Abundance.
138. Hwang TL, Shaka AJ (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. Journal of Magnetic Resonance, Series A 112: 275-279.
139. Rehm T, Huber R, Holak TA (2002) Application of NMR in Structural Proteomics: Screening for Proteins Amenable to Structural Analysis. Structure 10: 1613-1618.
140. Liew OW, Chong JPC, Yandle TG, Brennan SO (2005) Preparation of recombinant thioredoxin fused N-terminal proCNP: Analysis of enterokinase cleavage products reveals new enterokinase cleavage sites. Protein Expression and Purification 41: 332-340.
141. Eaton D, Rodriguez H, Vehar GA (1986) Proteolytic Processing of Human Factor-VIII - Correlation of Specific Cleavages by Thrombin, Factor Xa, and Activated Protein-C with Activation and Inactivation of Factor-VIII Coagulant Activity. Biochemistry 25: 505-512.
142. Gregory A. Cox RBJ, James A. Cook, Mark Wakulchik, Melvin G. Johnson, Elcira C. Villarreal, and Q. May Wang (1999) Identification and Characterization of Human Rhinovirus-14 3C Protease Deamidation Isoform. The Journal Of Biological Chemistry 274: 13211–13216.
143. Lavallie ER, Diblasio EA, Kovacic S, Grant KL, Schendel PF, et al. (1993) A Thioredoxin Gene Fusion Expression System that Circumvents Inclusion Body Formation in the Escheria-coli Cytoplasm. Bio-Technology 11: 187-193.
144. Bikerman JJ (1973) Foams [by] J. J. Bikerman. New York: Springer-Verlag.
145. Iglesias E, Anderez J, Forgiarini A, Salager J-L (1995) A new method to estimate the stability of short-life foams. Colloids and Surfaces A: Physicochemical and Engineering Aspects 98: 167-174.
146. du Noüy PL (1925) An Interfacial Tensiometer for Universal. The Journal of General Physiology 7: 625-631.
147. Sundaram S, Ferri JK, Vollhardt D, Stebe KJ (1998) Surface Phase Behavior and Surface Tension Evolution for Lysozyme Adsorption onto Clean Interfaces and into DPPC Monolayers: Theory and Experiment. Langmuir 14: 1208-1218.
148. Tripp BC, Magda JJ, Andrade JD (1995) Adsorption of Globular Proteins at the Air/Water Interface as Measured via Dynamic Surface Tension: Concentration Dependence, Mass-Transfer Considerations, and Adsorption Kinetics. Journal of Colloid and Interface Science 173: 16-27.
149. van der Vegt W, Norde W, van der Mei HC, Busscher HJ (1996) Kinetics of Interfacial Tension Changes during Protein Adsorption from Sessile Droplets on FEP-Teflon. Journal of Colloid and Interface Science 179: 57-65.
150. Siegel LM, Monty KJ (1966) Determination of Molecular Weights and Frictional Ratios of Proteins in Impure Systems by Use of Gel Filtration and Density Gradient Centrifugation. Application to Crude Preparations of Sulfite and Hydroxylamine Reductases. Biochimica Et Biophysica Acta 112: 346.
151. John Cavanagh WJF, Aurthur G. Palmer III, Nicholas J. Skelton (1996) Protein NMR Spectroscopy Principles and Practice: Academic Press Limited.
152. Perrin F (1936) Mouvement brownien d'un ellipsoïde (II): Rotation libre et depolarisation des flourescences. Translation et diffusion des molécules ellipsoïdales. J Phys Radium 7: 1-11.
153. Westley F, Cohen I (1966) Tables of values relating the axial ratio to the frictional ratio of an ellipsoid of revolution. Biopolymers 4: 201-204.

154. Zamyatnin AA (1972) Protein volume in solution. Progress in Biophysics and Molecular Biology 24: 107-123.
155. Cooper A (2004) Biophysical chemistry: Royal Society of Chemistry.
156. Op den Kamp JAF (1979) Lipid Asymmetry in Membranes. Annual Review of Biochemistry 48: 47-71.
157. van Meer G, Voelker DR, Feigenson GW (2008) Membrane lipids: where they are and how they behave. Nat Rev Mol Cell Biol 9: 112-124.
158. Ratledge C, Wilkinson SG (1989) Microbial lipids: Academic Press.
159. Fernandez D, Sani M-A, Gehman J, Hahm K-S, Separovic F (2011) Interactions of a synthetic Leu–Lys-rich antimicrobial peptide with phospholipid bilayers. European Biophysics Journal 40: 471-480.
160. Epand RF, Schmitt MA, Gellman SH, Epand RM (2006) Role of membrane lipids in the mechanism of bacterial species selective toxicity by two α/β-antimicrobial peptides. Biochimica et Biophysica Acta (BBA) - Biomembranes 1758: 1343-1350.
161. Brand G (2009) Unpublished Data.
162. Tweten RK (2005) Cholesterol-Dependent Cytolysins, a Family of Versatile Pore-Forming Toxins. Infection and Immunity 73: 6199-6209.
163. Muller-Eberhard HJ (1986) The Membrane Attack Complex of Complement. Annual Review of Immunology 4: 503-528.
164. Iacovache I, Bischofberger M, van der Goot FG (2010) Structure and assembly of pore-forming proteins. Current Opinion in Structural Biology 20: 241-246.
165. Anderluh G, Lakey JH (2008) Disparate proteins use similar architectures to damage membranes. Trends in Biochemical Sciences 33: 482-490.
166. Korzhnev DM, Kay LE (2008) Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. Accounts of Chemical Research 41: 442-451.
167. Carr HY, Purcell EM (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. Physical Review 94: 630.
168. Meiboom S, Gill D (1958) Modified Spin Echo Method for Measuring Nuclear Relaxation Times. Review of Scientific Instruments 29: 688-691.
169. Palmer AG, Kroenke CD, Loria JP (2001) Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules. Methods in Enzymology 339: 204-238.
170. Evans JNS (1995) Biomolecular NMR Spectroscopy: Oxford University Press.
171. Keeler J (2004) Understanding NMR spectroscopy.
172. Dimitry M. Korzhnev LEK (2007) Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding. Accounts of Chemical Research 41: 442-451.
173. Mittermaier A, Kay LE (2006) New Tools Provide New Insights in NMR Studies of Structural Protein Dynamics. Science 312: 224-228.
174. Mulder AA, Skrynnikov NR, Hon B, Dahlquist FW, Kay LE (2001) Measurement of Slow (us-ms) Time Scale Dynamics in Protein Side Chains by N Relaxation Dispersion NMR Spectroscopy: Application to Asn and Gln Residues in a Cavity Mutant of T4 Lysozyme. Journal of American Chemical Society 123: 967-975.
175. Davis DG, Perlman ME, London RE (1994) Direct Measurements of the Dissociation-Rate Constant for Inhibitor-Enzyme Complexes via the T1ρ and T2 (CPMG) Methods. Journal of Magnetic Resonance, Series B 104: 266.
176. Carver JP, Richards RE (1972) A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. Journal of Magnetic Resonance 6: 89.
177. Jen J (1978) Chemical Exchange and NMR T2 Relaxation-The Multisite Case. Journal of Magnetic Resonance 30: 111.
178. Mulder FAA, Tilborg PJAv, Kaptein R, Boelens R (1999) Microsecond time scale dynamics in the RXR DNA-binding domain from a combination of spin-echo and off-resonance rotating frame relaxation measurements. Journal of Biomolecular NMR 13: 275-288.
179. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, et al. (2005) The CCPN Data Model for NMR Spectroscopy: Developement of a Software Pipeline. Proteins 59: 687-696.
180. Balaji VN, Mobasser A, Rao SN (1989) Modification of protein stability by introduction of disulfide bridges and prolines: Geometric criteria for mutation sites. Biochemical and Biophysical Research Communications 160: 109-114.
181. Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 1751: 119-139.
182. Fasman GD (1996) Circular dichroism and the conformational analysis of biomolecules: Plenum Press.

183. Chapman S, Faulkner C, Kaiserli E, Garcia-Mata C, Savenkov EI, et al. (2008) The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. PNAS 105: 20038-20043.
184. Holzer W, Penzkofer A, Fuhrmann M, Hegemann P (2002) Spectroscopic Characterization of Flavin Mononucleotide Bound to the LOV1 Domain of Phot1 from Chlamydomonas reinhardtii¶. Photochemistry and Photobiology 75: 479-487.
185. Drepper T, Eggert T, Circolone F, Heck A, Krausz U, et al. (2007) Reporter proteins for in vivo fluorescence without oxygen. Nat Biotech 25: 443-445.
186. Stryer L (1965) The interaction of a naphthalene dye with apomyoglobin and apohemoglobin: A fluorescent probe of non-polar binding sites. Journal of Molecular Biology 13: 482-495.
187. McDonald RM, Smith BO Unpublished Data.
188. Sibisi S, Skilling J, Brereton RG, Laue ED, Staunton J (1984) Maximum entropy signal processing in practical NMR spectroscopy. Nature 311: 446-447.
189. Barkhuijsen H, De Beer R, Bovee WMMJ, Creyghton JHN, Van Ormondt D (1985) Application of linear prediction and singular value decomposition (LPSVD) to determine NMR frequencies and intensities from the FID. Magnetic Resonance in Medicine 2: 86-89.
190. Mandelshtam VA, Taylor HS, Shaka AJ (1998) Application of the Filter Diagonalization Method to One- and Two-Dimensional NMR Spectra. Journal of Magnetic Resonance 133: 304-312.
191. Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I (2006) Random sampling of evolution time space and Fourier transform processing. Journal of Biomolecular NMR 36: 157-168.
192. Orekhov VY, Ibraghimov I, Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. Journal of Biomolecular NMR 27: 165-173.
193. Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP (1987) Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. Journal of Magnetic Resonance (1969) 73: 69-77.
194. Rovnyak D, Frueh DP, Sastry M, Sun Z-YJ, Stern AS, et al. (2004) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. Journal of Magnetic Resonance 170: 15-21.
195. Wittekind M, Mueller L (1993) HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. Journal of Magnetic Resonance, Series B 101: 201-205.
196. Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. Journal of Biomolecular NMR 3: 185-204.
197. Clubb RT, Thanabal V, Wagner G (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C' chemical shifts in 15N-13C-labelled proteins. Journal of Magnetic Resonance (1969) 97: 213-217.
198. Montelione GT, Lyons BA, Emerson SD, Tashiro M (1992) An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. Journal of the American Chemical Society 114: 10974-10975.
199. Logan TM, Olejniczak ET, Xu RX, Fesik SW (1993) A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. Journal of Biomolecular NMR 3: 225-231.
200. Grzesiek S, Anglister J, Bax A (1993) Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in 13C/15N-Enriched Proteins by Isotropic Mixing of 13C Magnetization. Journal of Magnetic Resonance, Series B 101: 114-119.
201. Cavanagh J, Chazin WJ, Rance M (1990) The time dependence of coherence transfer in homonuclear isotropic mixing experiments. Journal of Magnetic Resonance (1969) 87: 110-131.
202. Uhrin D, Uhrinova S, Leadbeater C, Nairn J, Price NC, et al. (2000) 3D HCCH3-TOCSY for resonance assignment of methyl-containing side chains in C-13-labeled proteins. Journal of Magnetic Resonance 142: 288-293.
203. Yamazaki T, Forman-Kay JD, Kay LE (1993) Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./.epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. Journal of the American Chemical Society 115: 11054-11055.
204. Zhang H, Neal S, Wishart DS (2003) RefDB: A database of uniformly referenced protein chemical shifts. Journal of Biomolecular NMR 25: 173-195.
205. Neuhaus D, Williamson MP (1989) The nuclear Overhauser effect in structural and conformational analysis: VCH.
206. Sklenar V, Piotto M, Leppik R, Saudek V (1993) Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity. Journal of Magnetic Resonance, Series A 102: 241-245.

207. Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. Journal of Magnetic Resonance 202: 223-233.
208. Lipsitz RS, Tjandra N (2004) Residual Dipolar Couplings in NMR Structure Analysis. Annual Review of Biophysics and Biomolecular Structure 33: 387-413.
209. Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Mol Biol 5: 1065-1074.
210. Ottiger M, Delaglio F, Bax A (1998) Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. Journal of Magnetic Resonance 131: 373-378.
211. Clore GM, Gronenborn AM, Bax A (1998) A Robust Method for Determining the Magnitude of the Fully Asymmetric Alignment Tensor of Oriented Macromolecules in the Absence of Structural Information. Journal of Magnetic Resonance 133: 216-221.
212. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A (1997) Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Mol Biol 4: 732-738.
213. Sharma D, Rajarathnam K (2000) 13C NMR chemical shifts can predict disulfide bond formation. Journal of Biomolecular NMR 18: 165-171.
214. Nilges M (1995) Calculation of Protein Structures with Ambiguous Distance Restraints. Automated Assignment of Ambiguous NOE Crosspeaks and Disulphide Connectivities. Journal of Molecular Biology 245: 645-660.
215. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, et al. (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. Acta Crystallographica Section D 54: 905-921.
216. Linge JP, Nilges M (1999) Influence of non-bonded parameters on the quality of NMR structures: A new force field for NMR structure calculation. Journal of Biomolecular NMR 13: 51-59.
217. Habeck M, Rieping W, Linge JP, Nilges M (2004) NOE Assignment With ARIA 2.0 Protein NMR Techniques. Humana Press. pp. 379-402.
218. Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M (2003) Refinement of protein structures in explicit solvent. Proteins: Structure, Function, and Bioinformatics 50: 496-506.
219. Spronk C, Nabuurs S, Krieger E, Vriend G, Vuister G (2004) Validation of protein structures derived by NMR spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy 45: 315-337.
220. Nabuurs SB, Spronk CAEM, Vriend G, Vuister GW (2004) Concepts and tools for NMR restraint analysis and validation. Concepts in Magnetic Resonance Part A 22A: 90-105.
221. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. Journal of Biomolecular NMR 8: 477-486.
222. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577-2637.
223. Abragam A (1994) The principles of nuclear magnetism: Oxford University Press.
224. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. Journal of the American Chemical Society 104: 4546-4559.
225. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. Journal of the American Chemical Society 104: 4559-4570.
226. Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, et al. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. Journal of the American Chemical Society 112: 4989-4991.
227. Woessner DE (1962) Nuclear Spin Relaxation in Ellipsoids Undergoing Rotational Brownian Motion. The Journal of Chemical Physics 37: 647-654.
228. Halle B, Wennerström H (1981) Interpretation of magnetic resonance data from water nuclei in heterogeneous systems: AIP. 1928-1943 p.
229. Barbato G, Ikura M, Kay LE, Pastor RW, Bax A (1992) Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. Biochemistry 31: 5269-5278.
230. Farrow NA, Zhang O, Forman-Kay JD, Kay LE (1995) Comparison of the Backbone Dynamics of a Folded and an Unfolded SH3 Domain Existing in Equilibrium in Aqueous Buffer. Biochemistry 34: 868-878.
231. Press W, Teukolsky S, Vetterling W, Flannery B (1992) Numerical Recipes: The Art of Scientific Computing. 2 ed: Cambridge University Press.

232. Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28: 8972-8979.
233. Bruschweiler R, Liao X, Wright PE (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. Science (New York, NY) 268: 886-889.
234. Lee LK, Rance M, Chazin WJ, Palmer AG (1997) Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13Cα nuclear spin relaxation. Journal of Biomolecular NMR 9: 287-298.
235. Cole R, Loria JP (2003) FAST-Modelfree: A program for rapid automated analysis of solution NMR spin-relaxation data. Journal of Biomolecular NMR 26: 203-213.
236. Mandel AM, Akke M, Palmer IIIAG (1995) Backbone Dynamics of *Escherichia coli* Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. Journal of Molecular Biology 246: 144-163.
237. Chen J, Brooks CL, Wright PE (2004) Model-free Analysis of Protein Dynamics: Assessment of Accuracy and Model Selection Protocols Based on Molecular Dynamics Simulation. Journal of Biomolecular NMR 29: 243-257.
238. Ortega A, Amorós D, García de la Torre J (2011) Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. Biophysical Journal 101: 892-898.
239. Lee H, Choi J, Yoon S (2007) QHELIX: A Computational Tool for the Improved Measurement of Inter-Helical Angles in Proteins. The Protein Journal 26: 556-561.
240. Cooley RB, Arp DJ, Karplus PA (2010) Evolutionary Origin of a Secondary Structure: π-Helices as Cryptic but Widespread Insertional Variations of α-Helices That Enhance Protein Functionality. Journal of Molecular Biology 404: 232-246.
241. Hall J, Qiu X (2011) Structural and biophysical insight into cholesteryl ester-transfer protein. Biochemical Society Transactions 39: 1000-1005.
242. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology 247: 536-540.
243. Cuff AL, Sillitoe I, Lewis T, Redfern OC, Garratt R, et al. (2009) The CATH classification revisited—architectures reviewed and new ways to characterize structural divergence in superfamilies. Nucleic Acids Research 37: D310-D314.
244. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Research 38: W545-W549.
245. Mueller GA, Edwards LL, Aloor JJ, Fessler MB, Glesner J, et al. (2010) The structure of the dust mite allergen Der p 7 reveals similarities to innate immune proteins. The Journal of allergy and clinical immunology 125: 909-917.e904.
246. Kolodziejczyk R, Bujacz G, Jakób M, Ożyhar A, Jaskolski M, et al. (2008) Insect Juvenile Hormone Binding Protein Shows Ancestral Fold Present in Human Lipid-Binding Proteins. Journal of Molecular Biology 377: 870-881.
247. Hamiaux C, Stanley D, Greenwood DR, Baker EN, Newcomb RD (2009) Crystal Structure of Epiphyas postvittana Takeout 1 with Bound Ubiquinone Supports a Role as Ligand Carriers for Takeout Proteins in Insects. Journal of Biological Chemistry 284: 3496-3503.
248. Meyer P, Prodromou C, Liao C, Hu B, Mark Roe S, et al. (2004) Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. EMBO J 23: 511-519.
249. Chapman MD, Pomés A, Breiteneder H, Ferreira F (2007) Nomenclature and structural biology of allergens. The Journal of allergy and clinical immunology 119: 414-420.
250. N. Eswar MAM-R, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali (2006) Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics: John Wiley & Sons, Inc.
251. Shen M-y, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Science 15: 2507-2524.
252. Bakhtiar R, Chen H, Ogo SH, Fish R (1997) Gas phase molecular recognition of aromatic amino acid and aromatic carboxylic acid guests with a supramolecular $[(\eta^5$-pentamethylcyclopentadienyl)rhodium(2′-deoxyadenosine)$]_3^{3+}$ cyclic trimer host via non-covalent π–π interactions utilizing electrospray ionization mass spectroscopy. Chemical Communications: 2135-2136.
253. Daniel JM, Friess SD, Rajagopalan S, Wendt S, Zenobi R (2002) Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry. International Journal of Mass Spectrometry 216: 1-27.
254. Glick RE (1961) On the Diamagnetic Suseptibility of Gases. The Journal of Physical Chemistry 65: 1552-1555.

255. Arrighini GP, Maestro M, Moccia R (1968) Magnetic Properties of Polyatomic Molecules. I. Magnetic Susceptibility of H_2O, NH_3, CH_4, H_2O_2. The Journal of Chemical Physics 49: 882-889.
256. Vignaud A, Maître X, Guillot G, Durand E, de Rochefort L, et al. (2005) Magnetic susceptibility matching at the air–tissue interface in rat lung by using a superparamagnetic intravascular contrast agent: Influence on transverse relaxation time of hyperpolarized helium-3. Magnetic Resonance in Medicine 54: 28-33.
257. Bertini I, Bhaumik A, De Paëpe Gl, Griffin RG, Lelli M, et al. (2009) High-Resolution Solid-State NMR Structure of a 17.6 kDa Protein. Journal of the American Chemical Society 132: 1032-1040.
258. Gallagher SR, Desjardins PR (2001) Quantitation of DNA and RNA with Absorption and Fluorescence Spectroscopy. Current Protocols in Molecular Biology: John Wiley & Sons, Inc.
259. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual: Cold Spring Harbor Laboratory.
260. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Analytical Biochemistry 182: 319-326.