Biophysical and computational characterisation of the disorder-to-order structural transition of the small hydrophilic endoplasmic-reticulum associated protein, SHERP

Thesis presented by

Elliot Dudley Drew

To the University of London in the Faculty of Science for the degree of

Doctor of Philosophy

September 2017

Birkbeck College, University of London

Malet St

London WC1E 7HX

Declaration of conjoint work:

SAXS data collection was carried out by the candidate conjointly with Professor Rosangela Itri (*University* of *São Paulo*).

Modelling of SAXS data was carried out by the candidate conjointly with Gustavo Scanavachi (*University* of *São Paulo*, Brazil)

SRCD spectroscopy data was collected by Dr Andrew Miles (*Birkbeck College, University of London,* UK).

All other work in this thesis was produced solely by the candidate.

Signed: ______________________________
The Candidate

Signed: ______________________________
The Candidate’s Supervisor

Date: ______________________________

Abstract

This thesis explores the disorder-to-order structural transition of the small hydrophilic endoplasmic reticulum associated protein (SHERP). SHERP has been shown to be essential to the life cycle of *Leishmania major*, a parasite responsible for leishmaniasis which kills tens of thousands every year. The protein is almost entirely disordered in solution, but undergoes a dramatic increase in helicity upon binding to anionic lipids or detergents. Although the ordered structure of SHERP had previously been solved by solution nuclear magnetic resonance spectroscopy in the presence of sodium dodecyl sulphate (SDS), both the nature of the disordered ensemble of the protein and the organisation of the SHERP/detergent complex were unknown.

Using a combination of synchrotron radiation circular dichroism spectroscopy (SRCD), small angle X-ray scattering (SAXS) and molecular dynamics (MD), several projects were carried out exploring the disorder-to-order structural transition of SHERP in the presence of SDS. The effectiveness of sequence-based predictors to estimate the effect of single mutants was explored, with a number of mutants expressed and characterised by SRCD and MD. A mutant, the “permutant”, was designed with the aim of decreasing the disorder of the protein in solution while maintaining amino acid composition, by introduction of multiple potential $i \rightarrow i4$ salt bridges created by permutations of the wild-type sequence. Molecular dynamics simulations of the wild-type and “permutant” construct found a dramatic increase in salt bridge formation, and *in vitro* characterisation of the “permutant” construct showed it had significantly greater helical character than the wild-type in the absence of SDS.

The disordered ensemble of SHERP was characterised by replica exchange MD, SRCD and SAXS. Good agreement was found between simulation and experiment, with a predominantly unfolded ensemble deficient in secondary structure described by our results. The changes that occur upon SHERP binding to SDS were also characterised. MD simulation of the SHERP-SDS complex showed that the protein bound among the head-groups of the SDS micelle, and the helical content and helix-turn-helix structure was retained. It also allowed identification of several cationic side-chains which formed stabilising salt bridges with the sulphates of SDS. The complex was then characterised *in vitro*, by SAXS and CD spectroscopy. The addition of the protein led to a doubling in micelle length, with multiple SHERP molecules found to bind to the anionic head-groups in the shell of the micelle. The residues identified during the MD simulation were substituted with alanine to make a series of mutants with increasing negative charge. Significant decreases in helicity, micelle length and the numbers of protein bound occurred as negative charge increased, possibly caused by decreased affinity of the protein for the micelle causing less protein molecules to bind per micelle, leading to a decreased chance of stabilising protein-protein interactions resulting in partial folding of the protein. These results demonstrate the importance of charge-charge interactions in the disorder-to-order structural transition of SHERP, and provide structural context for future functional work on this protein.

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor, Professor Bonnie Wallace, for her continuous support during my PhD research and study. She let me take charge of the direction of my project and approach problems in my own way – without the guidance and opportunities she gave me I would not have made it.

I would also like to thank the other members of the my thesis committee, Professor David Jones and Dr Katherine Thompson, for their advice as my project progressed and for asking the hard questions that kept me focused.

I am hugely grateful to my wonderful collaborators from the University of Sao Paulo, Brazil – Professor Rosangela Itri and her group, especially Gustavo Scanavachi – for their tireless work on the SAXS data collection and analysis and for welcoming me so warmly when I visited Brazil.

I would like to thank Dr Andy Miles who taught me how to properly carry out and analyse circular dichroism spectroscopy data, as well as for collecting a large amount of the SRCD data presented in this thesis.

I would like to thank Dr Martin Ulmschneider for helping me get started with molecular dynamics and for his invaluable help and discussion as I took on more complex problems during the course of my project.

For funding my PhD, my thanks to the UK Medical Research Council. Thanks must also go to the Laboratório Nacional de Luz Síncrotron, Brazil (LNLS), ISA in Denmark, the Soleil Synchrotron in France and all the staff at these facilities for allowing me to obtain data over the course of my project. I would also like to thank University College London (UCL) for providing access to the Legion and Grace high performance computing clusters.

Thanks to everyone in the Wallace group over the course of my PhD for their knowledgeable advice, friendship and for giving me a second PhD education in the structural biology of sodium channels.

I would like to thank my family, especially my parents for supporting me and keeping me sane. My grandparents, particularly my Pop Pop who told me I should do biochemistry at University when I was still at primary school – sound advice so far.

I would like to thank my friends for their support, especially my girlfriend, Alice, for putting up with me as I wrote the thesis these last few months.

Finally, I’d like to thank the nation of Turkey for introducing coffee to Europe, Mr John James Sainsbury for founding the company that would one day produce Taste the Difference Tripe Chocolate Cookies, and Bardeen, Brattain and Shockley for inventing the transistor so Google could exist and make this process slightly less painful.

Table of Contents

List of Tables and Figures

Abbreviations

AIDS	acquired immune deficiency syndrome
APC	adenomatous polyposis coli
APBS	Adaptive Poisson-Boltzmann Solver
CD	Circular dichroism
Cdk	cyclin-dependent kinase
CKI	cyclin-dependent kinase inhibitor
COM	centre of mass
DNA	deoxyribonucleic acid
DOPC	1,2-Dioleoyl-sn-glycero-3-phosphocholine
DOPE	1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine
DOPG	1,2-Dioleoyl-sn-glycero-3-phosphoglycerol
ELM	eukaryotic linear motif
EM	electron microscopy
GADIS	genetic algorithm for design of intrinsic secondary structure
GO	gene ontology
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	lysogeny broth
LM	linear motif
LNLS	Laboratório Nacional de Luz Síncrotron
MD	molecular dynamics
MoRF	molecular recognition feature
NMR	nuclear magnetic resonance
NRMSD	normalised root mean square deviation
Nup	nucleoporin
PDB	protein data bank
PM	permutant
RDF	radial density function
REMD	replica-exchange molecular dynamics
RMSD	root mean square deviation
RMSF	root mean square fluctuation
RNA	ribonucleic acid
SAXS	small-angle X-ray scattering
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SHERP	small hydrophilic endoplasmic-reticulum associated protein
SLD	scattering length density
SLiM	short linear motif
SOC	super optimal broth

SRCD	synchrotron radiation circular dichroism
UCL	University College London
UV	ultraviolet
WT	wild type

1.0 Chapter 1: Introduction

The following introduction will give a broad outline of intrinsically disordered protein (IDP) research up to the present day, focusing on how their sequences, structure and dynamics have informed their evolution and function. The second part will introduce a small IDP from the parasite *Leishmania major*, the small hydrophilic endoplasmic reticulum associated protein (SHERP), providing a detailed look at the study and characteristics of a single IDP.

1.1 Intrinsic Disorder in proteins

When the first protein structures were solved by protein crystallography in the 1930s -50s, it was thought that a fixed three dimensional structure may be a requirement for proteins to function. Anfinsen’s Dogma states that the native structure of a protein is only determined by its sequence, and that said native structure is the unique, stable and kinetically accessible minimum of the free energy – there can be only one native structure (Anfinsen 1973). The concept of “configurational adaptability” had been put forward by F. Karush in the 1950s, where he suggested that a protein could adopt multiple configurations in the low energy native state allowing it to adopt different configurations depending on binding to partner molecules (Karush 1950). As more and more protein structures were solved, large portions of X-ray datasets were impossible to assign – this was due to regions of low electron density suggesting that some parts of proteins occupy multiple positions. It became apparent that proteins exist as conformational ensembles, with some regions less constrained than others. However, some proteins were found to exist on the extreme end of this spectrum, with very few constraints on their structure. For a long time there was little information to be found regarding these molecules; as such during the 80s and 90s they were rediscovered and renamed multiple times – examples include “floppy” (Lee et al. 1985), “rheomorphic” (Holt and Sawyer 1993) and “dancing” (Livesay 2010). It wasn’t until the early 2000s that studies were carried out showing that these flexible proteins constituted a novel and broad class of proteins with a staggeringly wide range of functions (Dunker et al. 2000). The most widely used moniker for this class of proteins nowadays is “intrinsically disordered proteins” (IDP). Intrinsic disorder can manifest on the level of secondary structure (leading to random-coil like conformations) or on the tertiary structure level.

Since their acceptance as a distinct protein class, IDPs have been shown to serve a huge range of critical signalling and regulatory functions in organisms as diverse as *E. coli* to *H. sapiens* (Dunker et al. 2000). Their unique dynamic properties allow them to perform tasks that would be impossible for the classically defined "rigid", ordered protein. However, these properties also make their study much more difficult and require some novel approaches combining computational biology, statistics and biophysics.

1.1.1 Sequence and Evolution

Bioinformatics tools have been crucial in the identification and study of IDPs. As IDPs were identified, analysis of their sequences began to shed some light on why they behave so differently to structured proteins. The observation that IDPs tend to be highly enriched in charged and polar residues and lacking in hydrophobic residues led to the development of the Charge-Hydropathy plot – a crude but effective way of differentiating between IDPs and structured proteins (Uversky et al. 2000). The lack of hydrophobic residues makes the formation of a buried hydrophobic core unlikely, leading to the extended structures often seen in IDP ensembles. This is exacerbated by the generally high net charge of the protein which can lead to unfavourable charge interactions. When hydrophobic residues are seen, this is generally in the form of motifs which recognize binding partners and facilitate the disorder-to-order structural transitions which are often critical to IDP function. Linear Motifs (LMs), also called eukaryotic linear motifs (ELMS) or short linear motifs (SLIMS) are examples of such recognition sequences that are conserved and found to fall in regions of local disorder (Fuxreiter et al. 2007). They are thought to serve as a primary contact site which, due to the greater exposure of highly disordered segments within the ensemble, make contact with an interacting partner first and can therefore drive binding of the protein.

Molecular recognition features (MoRFs) are another feature thought to provide a mechanism for partner recognition and binding. The preformed structural element hypothesis describes how conformational preferences of some IDPs for their bound structures – for example, a propensity for part of the IDR sequence to have some helical character that is also seen in the ordered structure – allows for recognition of a partner molecule and further disorder-to-order structural transitions

(Fuxreiter et al. 2004). Several varieties of MoRF have been discovered: α-MoRFs (form a-helices), β-MoRFs (form B strands), *i*-MoRFs (which form irregular structures) and complex MoRFs which have a mixture of secondary structure (Cumberworth et al. 2013). Nowadays, a variety of computational methods exist to predict MoRFs from protein sequences (Disfani et al. 2012; Malhis et al. 2016; Mészáros et al. 2009).

A large number of disorder prediction software have been developed using a variety of methods (Li et al. 2000). Some of these methods utilize machine learning techniques and homology methods (e.g. DISOPRED (Ward et al. 2004b), some physics-based methods (e.g. IUPRED (Dosztányi et al. 2005)) and there are a number of meta-predictors (e.g. PONDR-FIT (Xue et al. 2010)). Several databases of IDPs exist, most prominent being DisProt (Vucetic et al. 2005) which includes proteins with regions of disorder as well as fully disordered proteins.

The central role of IDPs in many essential functional roles in organisms across the entirety of the tree of life means that understanding the mutations that lead to a wide range of disease states is incredibly important. However, due to the unique properties of IDPs, the evolution of disordered regions differs from that of structured proteins. It has been shown that disordered sequences have a greater chance of mutating and that those changes will be non-conservative when compared to ordered sequences (Brown et al. 2010). Another study using order-based evolutionary models like BLOSUM62 (S. Henikoff and J. G. Henikoff 1992) and PAM120 (Dayhoff et al. 1978) predicted that disordered domains would not be conserved, when in actuality studies have shown that intrinsic disorder is often maintained over evolutionary time scales (Chen et al. 2006). The rate of evolution in IDPs is often faster than that in ordered sequences (Brown et al. 2002) and there is a much larger proportion of insertions and deletions in disordered regions (Brown et al. 2010).

The evolutionary differences between ordered and disordered sequences could be down to a variety of factors. Disordered proteins tend not to form many intramolecular contacts (Dunker et al. 2008) which would lead to less conservation. Some of the functions of IDPs would need little conservation to maintain function – for example regions that function as flexible linkers, where the only selective pressure is continued flexibility. However, many other features of IDRs, like post-translational modification sites, would have greater evolutionary constraints as changes to these may have a deleterious effect on important signalling pathways, though phosphorylated sites in disordered regions have still been shown to evolve faster than those found in ordered sequences in a study of mammalian proteomes (Chen et al. 2010). Tryptophan and tyrosine residues, though depleted in disordered sequences generally, were found to be highly conserved due to their importance in protein-protein interactions (Brown et al. 2010).

Evolutionary rate and the expression level of genes have also been found to be correlated by functional genomics studies (Xia et al. 2009). The expression of IDPs tends to be tightly regulated (Gsponer et al. 2008). This tight regulation makes biological sense when the signalling and regulatory roles IDPs tend towards are considered. In structured proteins, one theory for the slow rate of evolution for proteins that are highly expressed is due to the potentially high costs of a protein misfolding (Wilke and Drummond 2010). IDPs are unfolded by default; therefore misfolding is not as strong a selective pressure as it is with structured proteins.

Though disorder is found throughout the tree of life, it has been shown that IDPs are significantly more common in eukaryotes than prokaryotes. A proteomic study of six archaean, 13 bacterial and five eukaryotic genomes found that an average of 33% of eukaryotic proteins were predicted to contain long (>30 residue) regions of disorder compared to 4.2% for eubacteria and 2% for archaea (Ward et al. 2004a). It is thought that the presence of cellular compartments in eukaryotes provide greater protection against proteolysis for vulnerable IDPs with exposed backbones and indeed many of the major functions associated with IDPs are those associated with compartments like the nucleus (e.g. DNA binding). Other gene ontologies commonly associated with IDPs are those involved in the cell cycle and development – these functions are all essential for multicellular organisms but not present in single celled prokaryotes.

1.1.2 Structure and Dynamics

The sequence composition of IDPs leads to very different structural and dynamic properties compared with structured globular proteins. The most obvious difference lies in the flexibility of the protein allowing it to sample a large variety of conformations, often with transient secondary structural elements. Rather than the protein adopting one particular structure, an IDP exists as an ensemble of conformations. This means standard ways of representing and characterizing protein structure are not ideal when describing disorder.

The energetic landscape of IDPs differs greatly from that of ordered proteins. An ordered proteins folding can be described as a “folding funnel” with high energy conformations transiently sampled as the protein folds to the lowest energy, native state (Onuchic et al. 1996). Local minima also exist in this model, and represent transiently occupied transition states which temporarily trap the protein in a partially folded conformation (Leopold et al. 1992; Onuchic et al. 1996). Disordered proteins by their nature occupy a large number of conformational states, sampling these randomly. They do not fold and adopt a low energy native state without a partner or changes to their environment. Therefore the folding funnel does not apply to IDPs as strongly as it does to ordered proteins – instead the energy landscape of disorder is flatter, with many conformations of roughly equal free energy sampled (Granata et al. 2015).

Charge content has been shown to be crucial to defining the conformations adopted by IDPs (Müller-Späth et al. 2010). The lack of hydrophobic residues observed precludes the formation of a buried hydrophobic core; however IDPs show distinct preferences for collapsed or extended structures and this has been linked to the net charge of the protein. A 2010 study showed that nucleoporin (Nups) FG domains adopt conformations of varying compactness depending on charge (Yamada et al. 2010). Low charge content led to more globular conformations, while Nups with high net charge prefered highly dynamic extended configurations. This was also observed in an earlier study which simulated coarse-grained polypeptides with varying charge, hydrophobicity and length and led to the conclusion that disordered protein stability can be described as a “coil-to-globule transition in a charge/hydropathy composition space” (Ashbaugh and Hatch 2008). The distribution and patterning of charges has also been found to be of great importance to IDP conformation – sequences with opposite charges segregated lead to molten globules, whereas “well mixed” patterns of charges lead to extended conformations (Das and Pappu 2013).

Disordered protein chains are much more solvent-accessible than those in structured proteins. This accessibility, combined with an enrichment of recognition motifs for enzymes that carry out post-translational modification, explain IDPs prominent role as hub proteins (defined as proteins that bind to more than 10 partners) involved in complex signalling and regulatory pathways (Haynes et al. 2006). It also makes IDPs more susceptible to proteolysis, which has been exploited in the study of unstructured proteins with methods like FASTpp (Minde et al. 2012). The length of disordered regions has been linked to the half-life of proteins – IDPs with long >30 residue disordered terminal regions or >40 residue internal disordered regions are much more susceptible to proteosomal degradation than proteins with shorter disordered regions (van der Lee et al. 2014).

IDPs often undergo a structural disorder-to-order transition when they bind to a partner molecule (Dyson and Wright 2002). Often the disordered region in question has some transient structural propensity, as has been shown in p27 (Bienkiewicz et al. 2002) and p53 (Lee et al. 2000). These regions undergo induced folding upon interaction with the target molecule. These interactions have high specificity but low energy of association due to the energetic cost of undergoing the structural transition (Dunker et al. 2001). This allows for interactions with high reversibility, ideal for signalling proteins. The unfolded nature of IDPs also provides a greater contact area for interaction per residue. It was hypothesized that the flexible nature of IDPs allowed for increased rates of binding as the conformationally free, unfolded polypeptide chain would have a larger capture radius than an ordered protein and be able to “search” its environment for a partner – this was termed the “fly-casting mechanism” (Shoemaker et al. 2000). Although IDPs have been found to (generally) bind faster than ordered proteins, the main contributor to this is a high on and off rates due to low energy barriers to binding which allows encounter complexes to form, greatly increasing the chance of the final complex state occurring (Huang and Liu 2009).

The binding modes available to IDPs are numerous due to their flexibility. As stated above, many IDPs undergo a disorder-to-order transition upon binding to a partner. These can arise from binding to an ordered partner, as is the case with p27 and (the ordered protein) cyclinA-CDK2 (Russo et al. 1996), the association of two or more disordered proteins as found in many coiled-coils (Gazi et al. 2008), interaction with metal ions as seen in calsequestrin (Slupsky et al. 1987), binding with DNA as seen in bZIPs (Das et al. 2011) like Epstein Barr Virus ZTA (Figure 1 A) (Sinclair 2006) or binding to other molecules like lipids or detergents like SHERP (Figure 1 B) (Moore et al. 2011). These complexes are generally “static” and often ordered enough to allow structure determination by X-ray crystallography. Some IDP’s functions occur while the protein is still disordered – no structural transition takes place. These so-called “fuzzy complexes” (Figure 1 C) can be formed with ordered proteins (Permyakov et al. 2003), other disordered proteins (Sigalov et al. 2007) and lipid membranes (Sigalov et al. 2006). Rather than a structural transition taking place and a defined interface forming, fuzzy-complexes rely upon many low affinity binding sites which the IDP samples rapidly with equal probability (Uversky 2011). This provides a highly reversible way for IDPs to interact with other macromolecules.

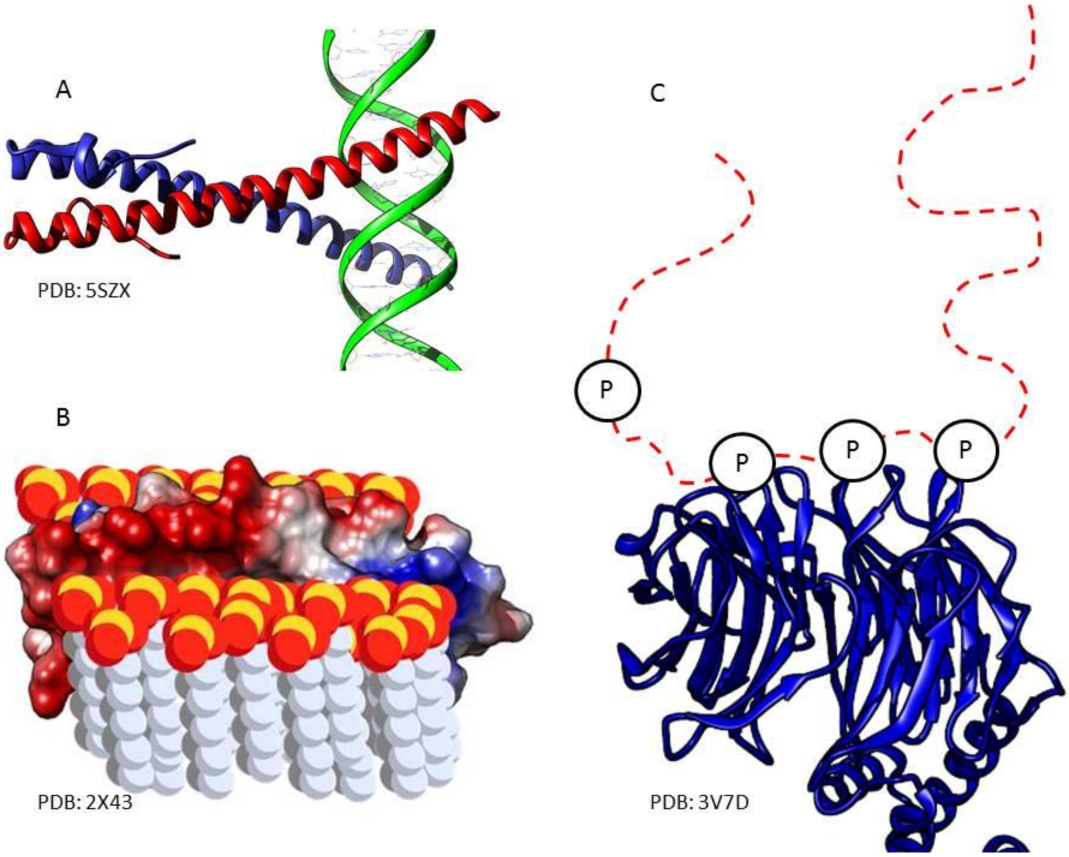


Figure 1: Examples of IDP binding modes. A Epstein Barr Virus ZTA (red and blue) bound to DNA – an example of a bZIP. B Hypothesised Binding mode of SHERP – SHERP (electrostatic rendering) folds upon binding to anionic lipids or anionic detergents like SDS. C Fuzzy complex formed by Cdc4 (blue) and Sic1 (red). Phosphorylation sites on Sic1 allow for transient contacts to be formed with Cdc4. Structures visualised using Chimera (Pettersen et al. 2004).

1.1.3 Function and Disease

A recurring theme throughout this introduction has been how ordered proteins and IDPs tend to be associated with very different functions in the body, and that their respective structural and dynamic features are what make them suitable for these roles. A study of the *S. cerevisiae* proteome found that the gene ontology (GO) terms IDPs are associated with were those important for signalling and molecular recognition e.g. “DNA binding”, “transcription factor”, “signal transduction” (Ward et al. 2004a). Many-to-one binding, tight regulation via post translation modifications, LMs and MoRFs – these properties make IDPs ideal for these kinds of functions. The GO terms least associated with IDPs include “structural” and “catalytic” – for these roles the promiscuity and plasticity of IDPs are a disadvantage. Catalytic activity, for example, is usually dependent on the presence of a fixed and highly specific catalytic site which requires a stable tertiary structure not found in IDPs.

Below is a selection of important functions associated with IDPs and how the properties of IDPs are important for these roles.

Histones are proteins that are the main components of chromatin, providing the first level of organization of DNA in chromosomes in eukaryotes. Eight histone proteins (2 H2A, 2 H2B, 2 H3 and 2 H4 subunits) combine to form the nucleosome around which DNA is coiled (Harvey and Downs 2004). Linker histones (H1) are involved in the shift in the organization of chromatin to more condensed forms. In low salt conditions, histone proteins have high levels of intrinsic disorder though in higher concentrations of salt (as would be found physiologically) histones become more ordered (Munishkina et al. 2004). Structures of histones solved by X-ray crystallography show that the protein is mostly helix – however around 30% is not ordered (Arents et al. 1991; Luger et al. 1997). The core histones (H2,H3 and H4) have a disordered N-terminal while the H1 linker histones have a disordered C-terminal – these disordered regions are involved in conformational changes of the nuclear pore particle by interactions with a range of macromolecules in the chromatin (Hansen et al. 2006). These conformational changes are involved in the large scale regulation of gene expression in the cell.

The cell cycle is the process by which cells grow and divide in eukaryotes. The cycle is controlled by cyclin-dependent kinases (Cdks) – these proteins regulate the timing for events in the cells life (Morgan 1995). Regulation of these proteins occurs in part by changes in the levels of cyclin-dependent kinase inhibitors (CKIs) which bind to specific Cdk-cyclin complexes that regulate different stages of the cell cycle – CKIs include the intrinsically disordered p27 (Lacy et al. 2004), p21 (Kriwacki et al. 1996) and p57 (Adkins and Lumb 2002) which fold upon binding to the Cdk-cyclin complex.

When something disrupts the cell cycle or something goes wrong with the cells molecular machinery, it will undergo apoptosis – programmed cell death. One of the key proteins in this process is an exemplary example of how the properties of IDPs allow for a multitude of essential functions to be carried out by one protein – p53. p53 is essential in multicellular organisms as it acts as a tumour suppressor. It demonstrates the promiscuity of IDPs, with over 1000 unique interaction partners found in its entry (ID: 113010 as of the 9/7/17) in the BioGRID database (Stark et al. 2006). At 37°C in the absence of any partners it is around 50% disordered. The C-terminal regulatory domain has been found to adopt 4 different structures upon binding to different partners – an alpha helix, a beta strand and two differently shaped coils (Oldfield et al. 2008). This structural plasticity is an example of what is termed a “chameleon sequence” – a sequence of amino acids that adopts different structures depending on the context (Minor and Kim 1996). This allows p53 to bind to many different partners. The promiscuity of p53 allows for highly tuned regulation of its activity by proteins like Mdm2 which induces structural changes in the N-terminal transactivation domain of p53 preventing it from activating or inhibiting gene expression (Vargas et al. 2003; Kussie et al. 1996).

We have seen that IDPs are involved in a wide range of essential regulatory and signalling pathways. It is therefore no surprise that IDPs are often implicated in the development of many diseases. In ordered proteins, a disease state is often associated with a protein failing to adopt its correct, functional structure, for example due to a mutation. This can lead to misfolding of the protein and loss of function. Although it has been shown that IDP sequences are more tolerant of changes than ordered proteins, they are still susceptible to mutations that lead to disease.

A computational study of a dataset of oncoproteins showed IDPs were over represented when compared to other eukaryotic proteins (Iakoucheva et al. 2002). This has also been seen in diabetes (Uversky et al. 2008), cardiovascular disease (Cheng et al. 2006) and acquired immune deficiency syndrome (AIDS) (Casu et al. 2013). IDPs have also been heavily implicated in neurodegenerative diseases like the synucleopathies, which include Parkinson's and Alzheimer's (Uversky 2003). Parkinson's is a progressive disease that affects an area of the brain called the *substantia negra*, leading to tremors, stiffness and issues with balance. It is characterized by the presence of Lewy bodies and Lewy neurites (Trojanowski et al. 1998; Takeda et al. 1998). These bodies can be up to 25 µm, made up of a core of filamentous material with filaments radiating out from the core. The predominant protein making up these fibres in Lewy bodies has been found to be a-synuclein (Spillantini et al. 1997). A-synuclein is intrinsically disordered in its monomeric state. However when it aggregates it forms fibrils with a high content of β-sheet – this is characteristic of amyloid fibrils (Serpell et al. 2000). It is these aggregates that are thought to play in a role in the pathogenesis of Parkinson's.

IDPs are associated with the development of many cancers. It has been suggested that mutations that increase order in disordered regions may interfere with the functions of MoRFs, preventing correct protein-protein interactions between the mutant IDP and its partner. The adenomatous polyposis coli (APC) tumour suppressor protein has a long 2000 residue C-terminal region which is predicted to be disordered (Minde et al. 2011). Mutations in this region have to be found to greatly increase the risk of tumours in the intestines. Some of these mutations, for example R1348W, have been predicted to lead to an increase in order in that region. Alterations to post-translational modification sites and the modification of protein-protein interaction sites have been proposed as mechanisms by which the mutations destroy APC function.

1.1.4 Conclusion

The flexible, unfolded nature of IDPs allows them to perform a vast array of functions that ordered proteins are unsuited to. As a fairly recent addition to the protein structure family, the study of these protein's structure, dynamics and function is in its infancy. However, their association to myriad diseases in humans, from cancer to diabetes, and ubiquity in nature make it clear that understanding the physics and biology behind intrinsic disorder is essential for a complete grasp of the fundamental processes that underpin biology.

1.2 Leishmania major and the Small Hydrophilic Endoplasmic Reticulum associated protein (SHERP)

1.2.1 Leishmania major

Leishmania major is one of the pathogens that cause human cutaneous leishmaniasis, a major subtropical disease infecting around 12 million people with up to 350 million people at risk according to data from the World Health Organisation (Desjeux 2004). Between 20,000 and 50,000 deaths occur each year (Lozano et al. 2012). Due to the widespread threat it poses to populations in these areas, the development of vaccines against it is of great interest; unfortunately no licensed vaccines for humans exist and treatments available can have unpleasant side effects (Singh and Sivakumar 2004). The parasite is carried and transmitted by female sand flies, which infect humans by feeding on their blood. Leishmaniasis presents itself in several different forms, ranging in severity from skin ulcers to low red blood cell counts, enlarged spleen and liver and death (Barrett and Croft 2012).

The genomes of parasites that have multiple host organisms over their life cycles are known to be enriched in IDPs (Pancsa et al. 2012). This is thought to be a consequence of their environment and “lifestyle”, as intrinsic disorder may lead to a reduced immune response from the host organism due to the difficulty of producing antibodies that can recognise these regions (Guy et al. 2015). *L. major* is no exception to this rule, with estimates of up to 70% of proteins produced by the organism having an IDR (Ruy et al. 2014).

The life cycle of *L. major* (Figure 2) proceeds as follows (Killick-Kendrick 1990): Female sand-flies inject the infectious metacyclic promastigotes when feeding on the human host (1). The metacyclic promastigotes are then engulfed by macrophages (2) and transform into amastigotes (3). These then replicate in infected cells (4) – different *Leishmania* species affect different tissues in specific ways, which leads to the varying clinical symptoms seen in different forms of leishmaniasis.

Sand-flies then become infected upon feeding on humans (5) carrying the parasite by ingesting macrophages ingested with amastigotes (6). These then differentiate into promastigotes in the sand-flies gut (7) which proceed to differentiate into the metacyclic promastigote, finally migrating to the sand-flies proboscis (8).

1.2.2 SHERP

The small hydrophilic endoplasmic reticulum protein, SHERP, is approximately 6.2 kDa with an acidic pI. It is expressed in the metacyclic promastigote stage of the parasite – the only protein to be highly expressed specifically in this stage of the life cycle (Knuepfer et al. 2001) and at a very high level (~100,000 molecules/cell). This, combined with its localization to the cytosolic facing surface of the cell membrane and the outer membrane of the mitochondria, suggests it plays an important role in the parasite’s infectivity. *In vivo*, *L. major* without the genes for SHERP and HASPB, another important *L. major* protein, cannot infect mammalian hosts (Doehl et al. 2017; Sádlová et al. 2010). However, the exact function of SHERP is still unknown.

Secondary structure predictions of SHERP suggest it will form an amphipathic alpha helix (Knuepfer et al. 2001). However, when disorder is predicted by DISOPRED2, 81% of the residues are predicted to be disordered. To characterise the secondary structure content of SHERP, synchrotron radiation circular dichroism spectroscopy (SRCD) and solution heteronuclear nuclear magnetic resonance spectroscopy (NMR) was employed to see how the protein behaved in solution, in the presence of lipids and in the presence of SDS, an anionic detergent (Moore et al. 2011).

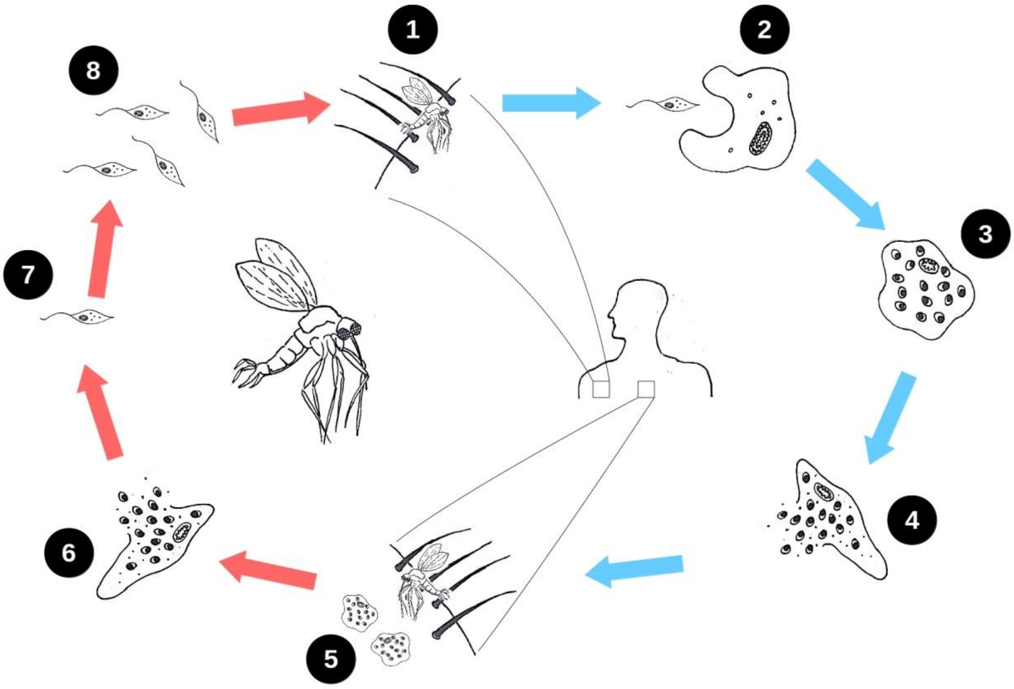
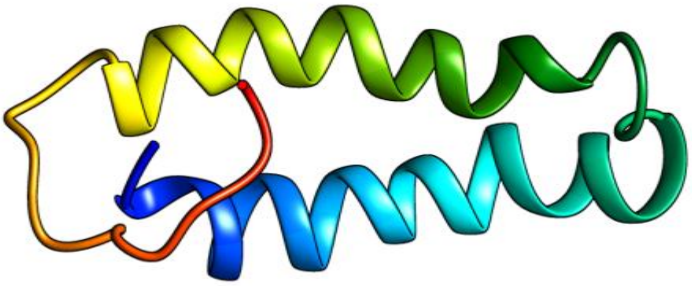


Figure 2: Life Cycle of Leishmania major. (Killick-Kendrick 1990)*: Female sand-flies inject the infectious metacyclic promastigotes when feeding on the human host (1). The metacyclic promastigotes are then engulfed by macrophages (2) and transform into amastigotes (3). These then replicate in infected cells (4) – different Leishmania species affect different tissues in specific ways, which leads to the varying clinical symptoms seen in different forms of leishmaniasis. Sand-flies then become infected upon feeding on humans (5) carrying the parasite by ingesting macrophages ingested with amastigotes (6). These then differentiate into promastigotes in the sand-flies gut (7) which proceed to differentiate into the metacyclic promastigote, finally migrating to the sand-flies proboscis (8). Red arrows indicate the stages that occur within the female sand-fly vector (5-8), blue arrows indicate stages that occur in the mammalian host (1-4).*

In solution, SHERP was found to be approximately 12% helical, with 44% of the structure being defined as “other” i.e. disordered. A combination of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) and 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) was next tested – this had no effect on the secondary structure of SHERP with the protein remaining disordered. However, in the presence of DOPC and the anionic 1,2-dioleoyl-sn-glycero-3-phosphoglycerol (DOPG) the helical content increased to ~48%. Taking spectra in the presence of SDS further tested the protein’s reaction to the presence of anionic membrane-like molecules. With SDS, helical content increased even further, to approximately 64% (Moore et al. 2011). These results strongly suggested that an anionic environment was essential for the disorder-to-order transition of SHERP and, because of this, probably its function. It is interesting to note that while the computational secondary structure predictions for SHERP were inconsistent with experimental data obtained with the protein in solution, they agreed very well with that of the protein in an anionic environment.

A structure of the protein in complex with SDS was obtained using heteronuclear multidimensional NMR spectroscopy (Figure 3). A ^{1}H-^{15}N HSQC spectrum of SHERP was also obtained in the absence of SDS, which showed a narrow ^{1}H chemical shift dispersion in the 8.0-8.6 ppm region which is characteristic of an unfolded, highly dynamic protein. The spectra in the presence of SDS confirmed the adoption of regular secondary structure by the SHERP. Analysis of chemical shift values for the protein showed that the protein had an alpha helical structure. The final solution for the structure is shown in Figure 3. The protein has a helix-turn-helix motif with flexible N and C termini.



```
1          11         21         31         41         51
MDQETRDQMK NAAAEAKDNV HDKIQELKDD VGNKAAEVRD AVSSTVESIK DKLSGGSSSR
61         71
ASSYTLEHHH HHH
```

Figure 3: NMR structure and sequence of ordered SHERP in the presence of SDS (PDB: 2X43) (Moore et al. 2011)*. Backbone is coloured from blue at the N-terminus to red at the C-terminus. Cartoon representation used to show secondary structure present. The protein adopts a helix-turn-helix conformation in the presence of SDS, resulting in an overall helicity of ~64% when theC-terminal hexa-histidine tag (not represented in the structure above) is included. Below the structure is the 73 residue sequence of the construct used throughout this thesis and also to obtain the NMR structure. The first 57 amino acids comprise the wild-type sequence from L major. The residues in green are a linker region used to connect the hexa-histidine tag to the rest of the sequence on the C-terminus. Structure visualised using Chimera* (Pettersen et al. 2004).

Anionic phospholipids have been shown to induce folding in a range of proteins including the intrinsically disordered a-synuclein (Davidson 1998) and human plasma apolipoprotein A1 (Boucher et al. 2004). These proteins also adopt amphipathic helices which interact with the lipid surface via charged side-chains – positively charged amino acids interacting with the negatively charged phospholipid head groups (Wang et al. 1997). NMR studies of apolipoprotein C-II in DOPC gave the proteins binding mode the nickname “snorkelling” due to the penetration of the helices ~3 Å into the lipid surface, forming larger assemblies of extended helices at higher concentrations of protein (MacRaild et al. 2004). Some recent studies have provided evidence that a-synuclein may form larger complexes with itself, allowing it to insert further into the membrane (Flagmeier et al. 2017). There is also some evidence that the concentration of lipid present is a determinant in the aggregation of the protein (Bodner et al. 2009). The similarities between SHERP and these proteins suggest that the protein could be prone to aggregation. It is noted that the NMR data does not preclude secondary structure being induced by oligomeric interactions between SHERP molecules – this may result in clustering of the proteins on the surface of the micelle, or even the formation of helical bundles that can penetrate the membrane surface further. This may explain differences in alpha helical content seen between the SDS spectra and those taken in DOPC/DOPG – different membrane/micelle makeup might lead to different oligomeric states, which could affect the secondary structure content.

Cross-linking studies have shown that SHERP is able to form a complex with subunit B of V-ATPase (Moore et al. 2011) which is known to be involved in the acidification of intracellular compartments in eukaryotes (Qi et al. 2007). Further experiments using surface plasmon resonance determined SHERP and V-ATPase rapidly dissociate – this suggest SHERP functions by specifically and reversibly binding to the protein, localizing it to specific parts of the cell membrane and mitochondria (Moore et al. 2011).

1.6 Summary

This project explores the effects of mutations upon the disorder-to-order transitions of a 57 amino acid protein, SHERP, from the parasite *Leishmania major* (Moore et al. 2011). In aqueous solution, SHERP is ~10% helical and mostly disordered. On addition of SDS and certain anionic lipids the protein has been shown to undergo a radical structural transition, becoming between 48% and 64% alpha helical (Moore et al. 2011).

There are several aims to this project. The first is to design a SHERP sequence that will fold into a more helical, ordered structure in aqueous solution, without the partner molecule and with minimal mutations, gaining an insight into the relationship between sequence and disorder-to-order transitions. The second is to characterise the interaction of SHERP and SDS using CD spectroscopy, SAXS and MD techniques, identifying the structural changes that occur to the micelle in the presence of protein and identify and characterise the role of charge in the interaction.

2.0 Chapter 2: Materials and Methods

2.1 Sequence based prediction of protein disorder

The sequence of SHERP was downloaded from Uniprot (Apweiler et al. 2004) in FASTA format (Uniprot ID: Q9XTN3). All single mutant sequences of SHERP were then generated by replacing the WT residues 2-57 with each of the 20 amino acids in turn, to yield 1064 unique sequences. DISOPRED2 (Ward et al. 2004b) was then run on each of these sequences to generate a per-residue disorder prediction.

The TOP-IDP disorder-propensity scale was used to investigate any links between a substitutions disorder-propensity and its predicted effect on disorder (Campen et al. 2008). This scale was obtained by a survey of 517 amino acid scales ability to discriminate between order and disorder. Using these results, a simulated annealing approach was used to find alternative scale values for each amino acid, resulting in improved performance vs existing scales. Order-promoting amino acids have more negative values, while disorder-promoting residues have more positive values.

2.2 Purification of Plasmid DNA

The SHERP gene construct was synthesized and ligated into a pEX-A2 vector backbone via Type IIS restriction enzymes Nco1 and Xho1 (Eurofins). The lyophilized plasmid was dissolved in double distilled and autoclaved H_2O to make a 100 ng/ml solution. 1µl of pEX-A2-SHERP was added to competent DH5α cells (New England Biolabs) then kept on ice for 20 minutes, before 30 seconds at 42°C then 5 minutes on ice. 200 µl of SOC overgrowth medium was then added to cells followed by 1 hr incubation at 37°C with shaking. The culture was then streaked onto a 100 µg/ml ampicillin plate and placed in a 37°C oven overnight. Colonies were cultured in 5 ml of LB media with 1:1000 dilution of 100 µg/ml ampicillin and grown overnight with shaking at 37°C. Plasmid DNA was purified and eluted in nuclease-free water with a Promega Wizard SV plus Miniprep Kit (Promega). The DNA concentration was determined using a NanoDrop ND-1000 (Thermo Scientific). pET-28a+ plasmid was transformed and purified as for pEX-A2-SHERP, substituting ampicillin for 50 µg/ml kanamycin.

2.3 Restriction Digest

Both the pET-28a+ and pEX-A2-SHERP plasmids were digested in a 50 µl reaction volume composed of 4 µl Nco1 enzyme, 4 µl Xho1 enzyme, 5 µl of CutSmart Buffer (all New England Biolabs) with the remaining 37 µl plasmid DNA at varying concentrations (see Section 3.2.1). The reaction mixture was spun briefly then incubated for 90 minutes at 37°C. 8 µl of 6x concentration Purple Loading Dye (New England Biolabs) was added to each digestion. The digestions plus 10 µl of 1 kbp-100 bp DNA ladder were run on a 1% TAE agarose gel (with 10 µl SYBERSafe (Invitrogen) added) at 100 V for 50 minutes. The gels were photographed under UV using the UVP BioDoc-IT Imagine system. Bands pertaining to SHERP gene construct and PET28a+ backbone vector were cut out of the gel and purified using the centrifugation protocol of the Promega Wizard SV gel clean up system.

2.4 Ligation

Ligation was performed in a 20 µl reaction containing 2 µl of 3 µg/µl T4 ligase (Promega), 2 µl ligase buffer (vortexed for ~1 minute) and 16 µl of digested pET28a+ and SHERP DNA in an approximately 1:6 ratio. The ligation reaction was incubated for >1 hour at RTP before incubation overnight at 14°C. 10 µl of ligation mixture was then transformed into competent DH5α cells as described previously, and plated out on 50 µg/ml kanamycin agar plates. DNA was purified from the colonies using a Miniprep kit (Promega) with the restriction digest and gel electrophoresis performed as above except with a 20 µl reaction volume to check for presence of SHERP insert, due to the small size of the gene.

2.5 Site Directed Mutagenesis

Complementary primer pairs were designed to be between 39 and 42 bases long with a Tm >78˚C. Table 1 shows the reagents and protocol used. The reaction mixture was initially heated to 95˚C for 180 seconds for the initial melting step. This was then followed by 18 rounds of 95 ˚C (melting step) for 50 seconds, 60˚C (annealing step) for 50 seconds followed by 68˚C (extension step) for 420 seconds, with a final 480 second hold at 68˚C after completion of 18 rounds. At the end of the program, the PCR product was held on ice for >5 minutes before addition of 2 µl of Dpn1 enzyme and spinning for 1 minute. The reaction was then incubated at 37˚C for 2 hours. 2 µl of this reaction was used to transform DH5a E. coli cells before streaking onto LB agar plates with 50 µg/ml kanamycin. Purified plasmid DNA from selected colonies was sent for sequencing.

Reagent	Volume Added (µl)
PfuUltra 10x Buffer	5
Template Plasmid	2
Primers	1.25 (of each primer)
DMSO	1.5
dNTPs	1
Autoclaved Double Deionised H2O	38
PfuUltra enzyme	1

Temperature (˚ C)	Time (seconds)	
95	180	
95	50	◄┐
60	50	Repeat x 18
68	420	◄┘
68	420	

Table 1: Reagents and protocol used in PCR mutagenesis reaction. The reaction mixture was initially heated to 95˚C for 180 seconds for the initial melting step. This was then followed by 18 rounds of 95 ˚C (melting step) for 50 seconds, 60˚C (annealing step) for 50 seconds followed by 68˚C (extension step) for 420 seconds. A final 480 second heating at 68˚C was performed, after which the reaction mixture was placed in ice for 5 minutes before addition of Dpn1 enzyme to digest any methylated DNA that might be present. DMSO was added to lower the Tm of the primer and template – it facilitates the amplification of GC rich regions template by binding to cytosine residues in the DNA, reducing the probability of secondary structure formation by the primer (Chakrabarti and Schutt 2001).

2.6 Expression and Purification of SHERP

1 µl of the purified SHERP-pET-28a+ construct previously described was added to 50 µl of BL21 (DE3) competent cells and held on ice for 20 minutes. The cells were then heat-shocked at 42˚C for 30 seconds then returned to ice for 5 minutes. 200 µl of SOC media was added to the cells which were then incubated at 37˚C for 1 hour with shaking. The cells were then streaked on a LB-agar plate containing 50 µg/ml kanamycin and incubated at 37˚C overnight. A colony from this plate was selected and used to inoculate 60 ml of LB media containing 50 µg/ml kanamycin before incubation at 37˚C with shaking overnight.

LB medium containing 30 µg/ml kanamycin was inoculated with a 1:100 dilution of the overnight growth culture, for a total of 8 flasks containing 750 ml of LB media. The cells were grown at 37˚C with shaking to an A_{600} of 0.6 – 0.8. Isopropyl B-D-thio-galactosidase (IPTG) was added to a final concentration of 1 mM and cells were incubated for a further 4 hrs. Cells were harvested by centrifugation at 5000 g and 4˚C for 15 minutes. Supernatant was discarded and pellets frozen at -80˚C until required.

Pellets were thawed at room temperature before resuspension with 1.5% of the original culture volume in 50 mM K_2HPO_4/KH_2PO_4, 300 mM NaCl, 5 mM imidazole at pH 7. Lysozyme was added to a final concentration of 0.75 mg/ml. The cells were incubated at room temperature for 25 minutes before lysis by sonication (3 x 30 s with 30 s intervals on ice) at 60% intensity. The solution was centrifuged at 15000 g at 4˚C for 25 minutes. The supernatant was filtered (0.45 µm) and applied to TALON resin in a Bio-Rad disposable column. The column was washed three times with 5x bed volume (7.5 ml) 50 mM K_2HPO_4/KH_2PO_4, 300 mM NaCl, 5 mM imidazole at pH 7. Protein was eluted using 50 K_2HPO_4/KH_2PO_4, 300 mM NaCl and 300 mM imidazole at pH 7, and elution fractions were analysed using SDS-PAGE. SHERP-containing fractions were pooled. A final size exclusion purification step using a G.E. Superdex 75 30/100 column with elution in 50 mM sodium phosphate buffer at pH 7.4 was performed with a 0.5 ml fractionation step, and all fractions found to contain protein were analysed with SDS-PAGE and SHERP containing fractions pooled.

2.7 Circular dichroism spectroscopy – background

Polarised light can be described as the sum of two perpendicular, linearly polarised electromagnetic waves, generally referred to as vertically and horizontally polarised light (Wallace and Janes 2009). Combining vertically polarised and horizontally polarised light waves that are in phase results in a light wave that is linearly polarised at 45 degrees. In order to stay linearly polarised, the components have to be in phase. If the one of the component waves is out of phase with the other by a quarter-wave the resulting light wave is a helix and is known as circularly polarised light. This circularly polarised light can be either left- or right-handed, and it is chiral – left- and right-handed circularly polarised light waves are non-superimposable mirror images of each other (Figure 4). This chirality is key to CD's usefulness in studying biological macromolecules

Chiral molecules, or enantiomers, interact with left and right handed circularly polarised light differently in a wavelength dependent fashion. Due to the different absorption of each "handedness" of light that occurs in chiral media, the result of superposing these two circularly polarised light waves is a linearly polarised wave that rotates, and would trace an elliptical shape with its electric vectors looking down the beam (Figure 5). The difference in absorption gives us a value for the circular dichroism of a sample as a function of wavelength.

Many biological molecules display chirality. 19 of the 20 amino acids are chiral, as are the higher order structures formed by proteins, DNA and RNA. The most common application of CD spectroscopy in biology has been the analysis of protein secondary structure.

2.7.1 Information obtained from CD spectroscopy

Far-UV (190-250 nm)

The far ultraviolet (UV) region can be used to obtain information about the secondary structure of proteins. α helix, β sheet, unfolded proteins and poly-proline II all have distinct spectral signatures in the far-UV (<260nm) wavelength region (Figure 6) due to electronic transitions that occur in the

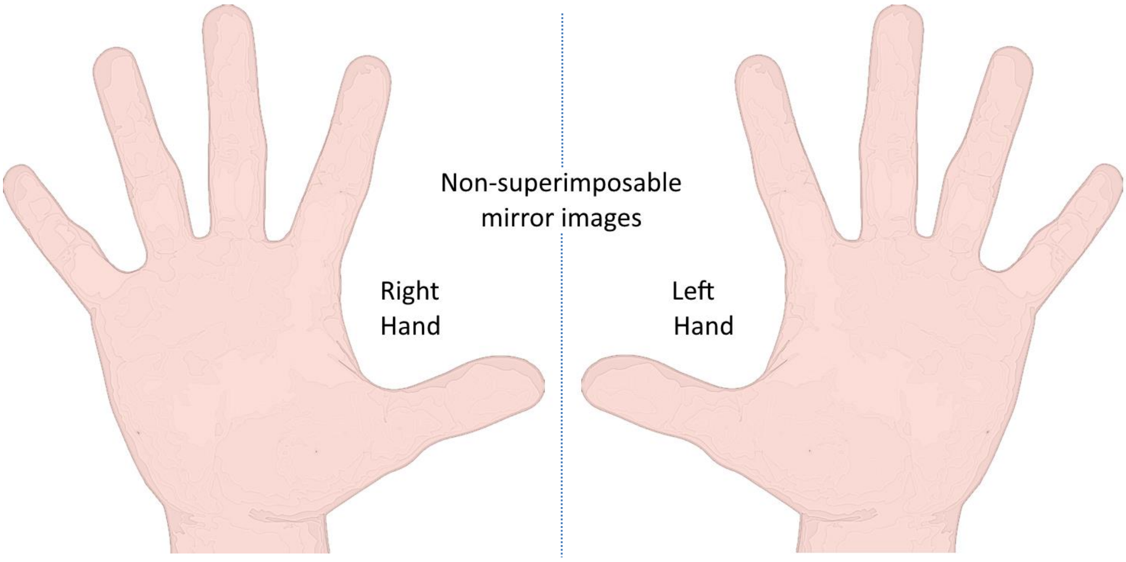


Figure 4: An example of chirality. Human hands are chiral. Comparing the right and left hand of an individual (to all intents and purposes mirror images of each other) and they are non-superimposable – in other words it is distinguishable from its mirror image.

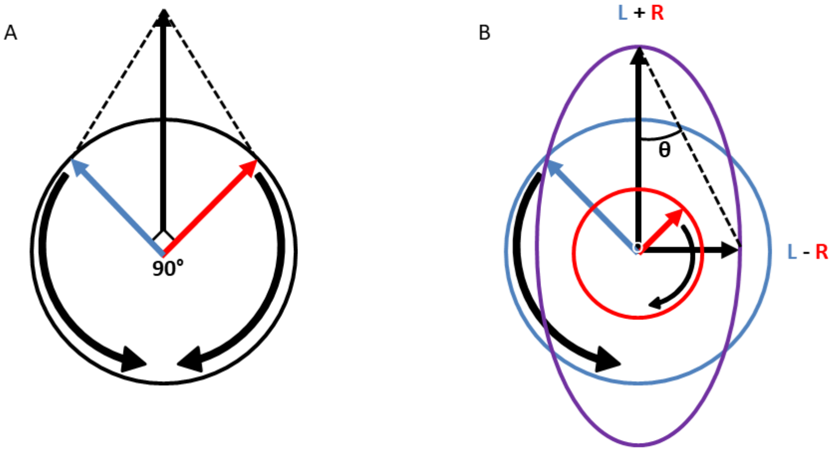


Figure 5: Circularly polarised light and CD. ***A*** *Circularly polarised light is composed of two perpendicular planar waves of equal magnitude 90° out of phase with one another. If the wave was propagating towards an observer, the electric vector of the wave would appear to be rotating.* ***B*** *If light composed of both left- (blue) and right-handed (red) circularly polarised light passes through a chiral medium the left- and right-handed light are absorbed differently and that difference (ϴ) is called circular dichroism, giving rise to elliptically polarised light (purple).*

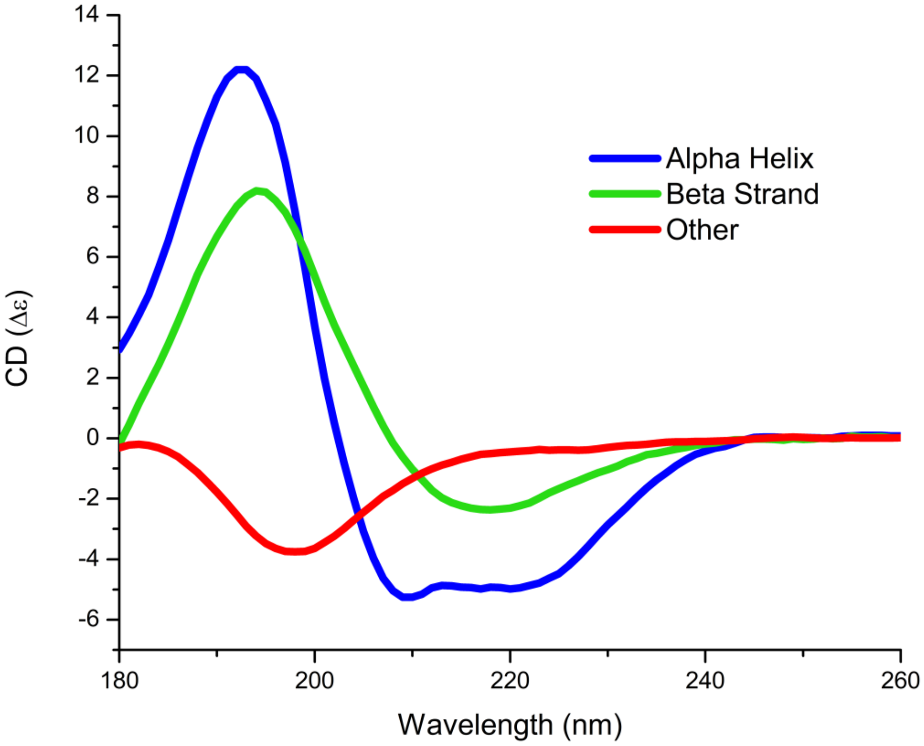


Figure 6: Examples of far-UV CD spectra. CD spectra of predominantly alpha helical (blue), beta sheet (green) and "other" (red) proteins (Abdul-Gader et al. 2011; Lopes et al. 2013)*. Alpha helical spectra are characterised by minima at 208 nm and 222 nm, and a maxima at 191 nm, Beta sheet spectra have a variety of shapes due to the range of topologies sheets can assume, but a "typical" anti-parallel sheet has a negative band at 218 nm and a maxima at 196 nm, "Other" or disordered spectra are typified by a minima at ~200 nm. Data accessed from PCDDB* (Whitmore et al. 2017) *on 15/08/17.*

peptide backbone which is a chromophore in this region. For alpha helices this results in a positive signal at 191 nm and negative signals at 208 nm and 222 nm. β sheets show a negative band at 218 nm and a positive one at 196 nm (although tertiary structure arrangement i.e. parallel or antiparallel sheets can have an effect on the location of these signals), and "other" (typically referring to disordered regions) has a negative peak at ~200 nm (Kelly and Price 2000).

Near-UV (250-350 nm)

In the near-UV, the dominant chromophores in protein samples are aromatic side-chains and disulphide bonds. This region is sensitive to changes in tertiary structure – for example, a protein with secondary structure but very little tertiary structure (e.g. a single, extended alpha helix) would have almost zero signal in this region, while a well folded protein would have a large CD signal. This is due to folding of the protein placing aromatic residues into chiral environments (Kelly and Price 2000). Information can also be gleaned on the environment of various non-protein cofactors and ligands that may be present in the protein e.g. haem groups. The near-UV spectra of a protein can act as a "fingerprint" for a protein, which is of great use to researchers interested in ensuring a protein is folded correctly.

Visible range (>350 nm)

Many non-protein cofactors and prosthetic groups like haem, flavin etc. exhibit CD in this range (Kelly and Price 2000). Typically the group on its own does not have a CD signal. However when it is placed in the chiral environment of a protein, CD effects can be seen.

2.7.2 Analysis of protein CD spectra

The CD spectrum of a protein is the summed contributions of these structural elements. The simplest method for getting structural data from CD spectra is to use this fact and calculate a weighted linear combination of theoretical "pure" spectra of each type of secondary structure. Unfortunately there are no standard reference spectra of pure secondary structures, and other complications exist – for example, CD is dependent on alpha helix length, and the tertiary arrangement of secondary structure elements in relation to one another can also change the spectra, most clearly seen in the large variety of spectra resulting from different forms of beta sheet (Miles et al. 2003). Therefore, although deconvolution methods vary in their approach, they generally rely on a reference dataset of protein CD spectra with known secondary structure content (Whitmore and Wallace 2008).

The secondary structure of the reference proteins is generally obtained from X-ray crystal structures, although other sources (e.g. bioinformatics sequence based predictors) could be used. The proteins in the reference dataset are chosen so they cover the large range of folds and secondary structure found in the PDB. Generally, a deconvolution method selects basis spectra from the reference dataset. These basis spectra are then combined to fit the experimental CD spectra being analysed, and secondary structure content is estimated from the weighted contributions of each reference spectra making up the reconstructed data. The accuracy of the result obtained is determined by calculation of the normalised root mean square deviation (NRMSD) between the reconstructed and experimental data. Obviously CD spectroscopy can only measure secondary structure content as an ensemble property – one cannot determine which residues are which secondary structure, just the proportion of the sample which corresponds to each type.

CD spectroscopy also proves a powerful tool for monitoring changes in protein structure due to its environment. Changes due to temperature, pH, denaturants or interacting ligands can be monitored, with spectra collected in minutes and single wavelength kinetics possible on a millisecond scale. A common application of CD is to check a mutant protein has folded correctly, or check the effect of drugs on protein structure and stability.

2.7.3 Sample conditions and CD Spectroscopy

Careful preparation of the sample, and selection of sample conditions, is essential for accurate and reliable CD spectra. The protein must be as pure as possible in order to prevent signals from contaminants affecting the results. Protein concentration must be measured accurately – this will allow the path length of the cell to be chosen correctly to allow for the best signal to noise ratio possible (Greenfield 2006), and is essential if the analysis will involve deconvolution of the spectra using a tool like DichroWeb (Whitmore and Wallace, 2004).

Buffer choice is highly important, as some reagents absorb light in the UV region (Greenfield 2006). Chloride ions are present in a large number of commonly used salts and buffers. Unfortunately they absorb very strongly in the far-UV, and are unsuitable for UV-CD experiments. Instead NaF salts can be used as a replacement. Imidazole (commonly used in Ni affinity chromatography to elute the protein) absorbs in the UV very strongly and should be removed by buffer exchange or dialysis before taking spectra. These issues should also be kept in mind when performing denaturation experiments with chemicals like urea – the lower wavelength regions of the spectra will be unusable for analysis. The wavelength cut-offs of a range of commonly used buffers in a 1 mm path length cell are shown in Table 2.

2.7.4 Synchrotron Radiation circular dichroism

CD spectroscopy is a powerful tool, but (as with many spectroscopic techniques) it is limited by the brightness of the light source used. By using a synchrotron light source with much higher flux than conventional xenon arc lamps, the signal-to-noise ratio is greatly improved allowing for smaller amounts of sample to be used and much faster measurements to be made. It also provides consistent flux over a much larger wavelength, allowing data collection in aqueous samples to ~168 nm. The improved signal-to-noise ratio allows smaller changes in a proteins structure (due to environment or mutation) to be resolved. The lower wavelengths in the vacuum UV region (<190 nm) contain considerable information about protein structure only attainable by SRCD (Miles and Wallace 2006). For example, the strong signals produced by alpha helical residues can make it hard to discern the signal caused by beta sheets in mixed alpha/beta proteins. However at low wavelengths the signal of alpha helix and beta sheets have different signs, allowing for more accurate determination of secondary structure content.

Buffer Constituent	**Wavelength Cut-off (nm)**
Distilled Water	<185
100mM Ammonium Citrate	220
150mM Ammonium Sulphate	190
100mM MES	205
100mM Pipes	215
100mM Sodium Chloride	195
10mM Sodium phosphate	<185
100mM Sodium phosphate	190
Phosphate buffered saline	200
100mM Tris-HCl	200
Acetonitrile	185
DMSO	252
Ethanol	195
Hexafluoroisopropanol	<185
Methanol	195
Trifluoroethanol	<185
4M GdnHCl	210
4M Urea	210

Table 2: Commonly used buffers and reagents and their wavelength cut-offs. Values refer to their use in UV-CD experiments in a 1 mm path length cell. Table reproduced from the Vanderbilt University guide to CD spectroscopy. ("CD sample preparation", 2014).

2.8 Specifics of CD spectroscopy methods used in this thesis

SRCD spectra were measured on beamline CD1 at the ISA synchrotron, Aarhus, Denmark. Protein concentrations were measured in triplicate using a Nanodrop 1000 immediately prior to data collection using an extinction coefficient of 1490 M^{-1} cm^{-1}. The extinction coefficient was calculated using the ProtParam webserver (Gasteiger et al. 2005). All samples were in 50 mM Na phosphate buffer at pH 7.4. For samples with detergent, 10% SDS was added to the sample for a final concentration of 0.1% SDS. The samples (2-5 mg/ml) were loaded into a 0.0024 cm path length demountable cylindrical Supracil cell (Hellma) or for samples at 1-2 mg/ml a 0.00507 cm path length cell. Spectra were measured over the wavelength range 280 nm to 170 nm using a 1 nm step size and a dwell time of 2 s. The wavelength cutoff for these samples was determined by the simultaneous measurements of the high tension voltage spectra. High tension voltage is roughly proportional to absorbance, so when it rises above a certain threshold (which varies due to the machine you are using) it's an indication that the sample is saturated and not enough light will reach the detector and meaningful values will not be obtained at those wavelengths. Three repeat scans were averaged and subtracted from 3 averaged baselines of 50 mM Na phosphate buffer at pH 7.4 with or without 0.1% SDS. The resulting spectrum was calibrated against a spectrum of camphorsulfonic acid (Miles et al. 2003), before scaling to units of delta epsilon, using a mean residue weight of 113.

Conventional CD spectra were collected on an Aviv 400 CD spectrophotometer. Sample preparation and collection was performed as for SRCD spectra.

Processing was carried out using the CDTool software (Lees et al. 2004). Secondary structure analyses were carried out using the DichroWeb analysis server (Whitmore and Wallace 2004). Values from the algorithms CONTINLL (Provencher and Glöckner 1981; van Stokkum et al. 1990), SELCON (Sreerama and Woody 2000) and CDDSTR (Sreerama and Woody 2000) (using reference set 4 for samples in presence of 0.1% SDS and set 7 for samples in absence of SDS) were averaged and standard deviations calculated from the results of these 3 methods. For all methods, a goodness-of-fit parameter, the NRMSD was calculated as a measure of the correspondence between calculated structure and experimental data (Mao et al. 1982).

Thermal melts for the K28F, D29W, A12E and S43F mutants were collected by Dr Andrew Miles. All other spectra were collected by the author.

2.9 Small-angle X-ray Scattering (SAXS) - background

Small-angle X-ray scattering is an analytical method that records the elastic scattering of X-rays by a sample with particles in the nm-range. The intensity of X-rays scattered by a sample at small angles (generally from 0.1-10°) are measured - this can give information about the structure and size of macromolecules like proteins, and the composition of partially ordered materials (like micelles). It is able to resolve structural information between 5-25 nm. The first SAXS observations were reported in the late 1920s (Krishnarmurti 1930), and by 1955, Guinier and Fournet had demonstrated that information on the organization and structure of disordered and partially ordered systems could be obtained using SAXS, as well as information on particle size and shape (Guinier and Fournet 1955).

X-rays are electromagnetic waves with a wavelength 0.01 to 10 nm, with energies ranging from 0.1 to 100 keV. For biological applications, X-rays with wavelengths typically <0.3 nm are usually used, as this wavelength is similar to the size of atoms and atomic bonds they are perfect for looking at the structure of proteins and other molecules – maximum diffraction occurs when the wavelength is equal to the distance between particles as per Bragg’s Law. X-rays, like all electromagnetic radiation, can be described as both a particle (a photon) and a wave.

2.9.1 X-ray Scattering

In order to get information on the sample using electromagnetic radiation, an interaction between it and incident radiation must occur. In the case of X-rays shone on a sample, 3 different events can occur: some X-rays will pass through the sample without interacting; some will be absorbed and transferred into another form of energy; and some of the incoming radiation will be scattered in all directions by the atoms in the sample.

In the case of SAXS, it is the intensity of scattered light that is measured. In particular, we are interested in Rayleigh scattering – the elastic collision of photons with electrons in the sample (Strutt 1899). This collision causes the electrons to oscillate at the same frequency as the incident X-ray photons, leading to the emittance of radiation with the same frequency. The energy and frequency of the photon is conserved – just the direction of travel is changed. The radiation emitted can therefore interfere at the detector leading to an interference pattern – it is from this that structural information about the sample can be deduced.

This interference can be both constructive or destructive depending on whether they are in or out phase upon reaching the detector due to the observation angle (the angle between the incident beam direction and the observation direction; See Figure 7), and the relative positions and distances of the atoms in the sample. The variations of intensity that can be seen in the scattering pattern are therefore characteristic to these distances and positions.

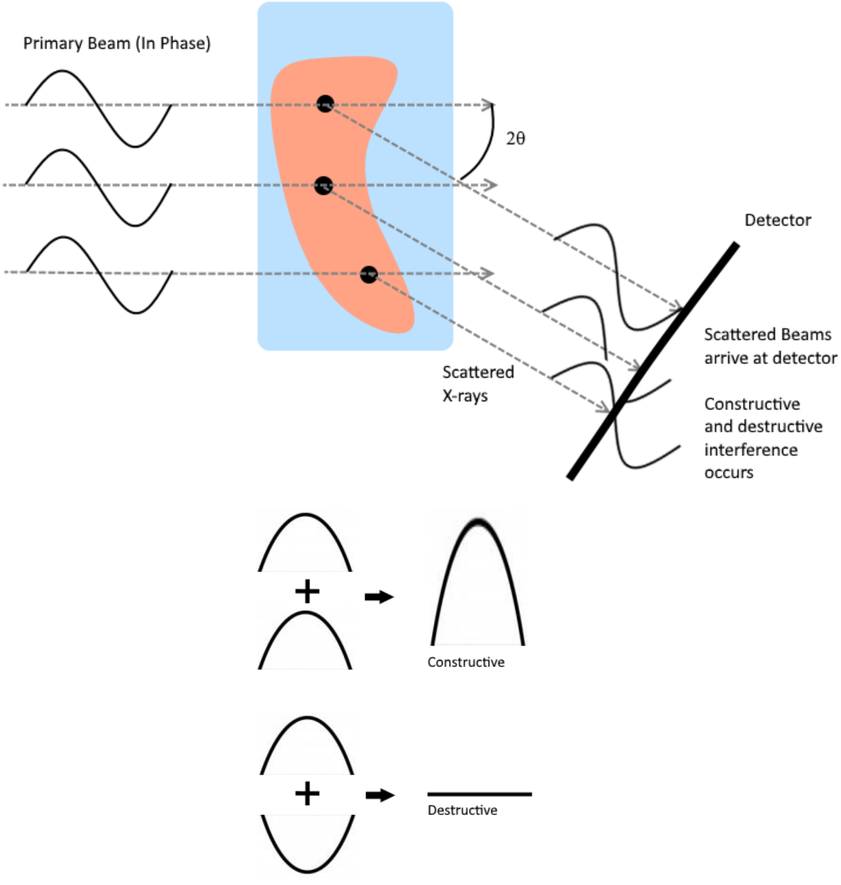


Figure 7: Scattering of X-rays results in a scattering pattern. The orientations and distance of electrons in a sample affects the intensity of light hitting the detector due to the constructive or destructive interference of the scattered waves. Constructive interference occurs when the waves are in phase leading to a larger signal, whereas destructive interference occurs when the waves are out of phase leading to a smaller one. This interference leads to the scattering pattern observed at the detector. 2ϴ is the angle between the incident X-ray beam and the detector, used to obtain the scattering vector q.

A scattering pattern can be described as a function of q, described by the equation below. q is the length of the scattering vector and generally has units of 1/nm – a scattering pattern describes the structure in reciprocal space.

$$q = \frac{4\pi}{\lambda} \cdot sin(\theta)$$

2.9.2 Form Factor P(q)

Usually, the object we are studying in biology is a particle made up of many atoms, like a protein. The scattering of a single particle can be described as the squared sum of all the wave amplitudes at the detector – this gives an interference pattern that oscillates in a way characteristic of the shape of the particle.

The form factor only relates to one particle if:

- The sample is monodisperse (i.e. all the particles have the same size and shape)
- The sample is dilute (i.e. the distance between particles, *r*, is large in comparison to the wavelength of the incident radiation)

In a dilute, monodisperse sample, the experimental scattering pattern corresponds to the form factor P(q) multiplied by the number of particles that are illuminated. In a polydisperse sample (where particles have different sizes and shapes), the average scattering pattern is described by the sum of the form factors of all the particles sizes in the solution – this has to be dealt with in a different fashion, for example by the fitting of mathematical models using computer software like GENFIT (Spinozzi et al. 2014).

2.9.3 The Structure Factor S(q)

When a sample is not dilute (i.e. the distances between particles are of a similar magnitude to the distances found between atoms inside the particles), the interference pattern contains contributions from interactions with neighbouring particles. This is termed the structure factor, S(q). This multiplies with the form factor, changing the shape of the SAXS profile obtained for the sample. At small q-values (~0 - 0.5 nm^1), the structure factor highlights changes in intensity due to concentration – an increase in intensity is due to attractive interactions between particles (often a sign of aggregation in a sample) while a decrease indicates repulsive interactions. In a dilute system, S(q) can be approximated as 1.

2.9.4 Data Interpretation

What information can be obtained from the scattering curve? The equation below shows the different terms that make up the SAXS profile obtained from an experiment.

$$\Delta I(q) = I_0 \cdot (\Delta\rho)^2 \cdot {V_1}^2 \cdot P(q) \cdot S(q)$$

The first three terms are constant – they give information on particle volume, contrast and concentration. The form factor, P(q), has information on the shape and the internal density distribution of the particles in the system. Analysis of this can also give information about whether the system is monodisperse or polydisperse/polymorphous. The structure factor, S(q), gives information on particle-particle interactions. Inter particle distances can be obtained from this factor.

2.9.5 Kratky Plots

The Kratky plot provides an excellent way to evaluate how folded or compact a sample is. It can be derived from the scattering curve and is plotted as a $q^2I(q)$ vs q. Folded domains have peaks that are roughly parabolic, with the position of the peak giving some indication of the size of the molecule (peak at higher q meaning a larger particle). An unfolded sample (e.g. an unfolded or disordered polypeptide) lacks this peak and is instead linear with respect to q at higher q values. Figure 8 shows a variety of Kratky plots originating from proteins of differing foldedness.

2.9.6 Applications in Biology

SAXS ability to give information about the size and shape of a molecule in solution can be of great use for biological samples. It can be used to get information about samples that are hard or

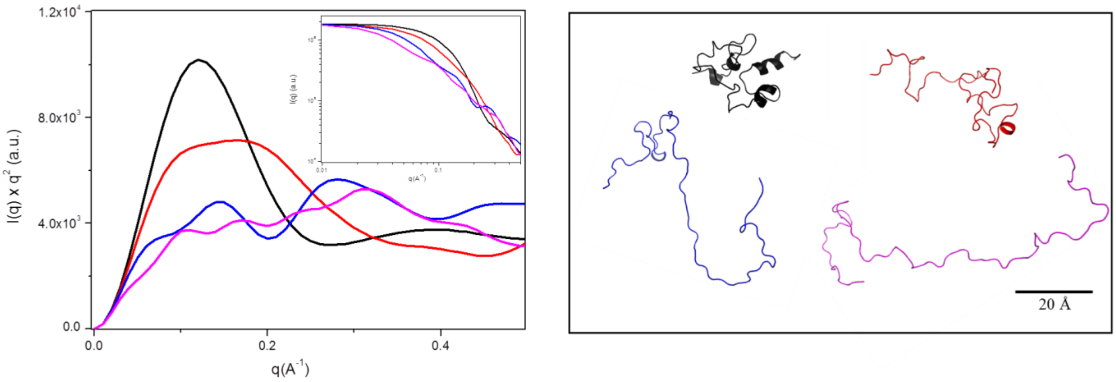


Figure 8: Kratky plots calculated from a variety of SHERP conformations obtained from the REMD simulation discussed later in this work (Section 4.2.1), from compact (in black) to extended and disordered (in pink). Compact, globular structures exhibit parabolic plots as seen in black. As a structure becomes more extended, the plot becomes "flatter" as seen for the red, blue and pink plots. Structures visualised using Pymol (Schrödinger, LLC 2010).

impossible to study by X-ray crystallography – it also has the benefit of allowing data to be collected in near physiological conditions. It can be used to study proteins and systems too large for solution NMR. However, it is perhaps most useful when used in conjunction with other methods. Combining SAXS and electron microscopy (EM) has allowed the building of atomic resolution models of large protein complexes which have eluded crystallization (Pattanayek et al. 2011). SAXS data can provide restraints for modelling of conformational ensembles – whether these are produced by clustering of statistical coils or used to guide full molecular dynamics simulations. This has been used to characterize the structure and dynamics of IDPs which do not tend to crystallize (Oroguchiet al. 2011). SAXS can be used to mitigate the main problems faced by solution NMR – the size of the molecule being studied. For example, NMR structures of individual components of large complexes can be fitted to a low resolution structure of the same complex obtained from SAXS (Pattanayek et al. 2011).

Several programs exist to process SAXS data. The program SCATTER (Förster et al. 2010) is commonly used and allows the user to perform a wide variety of analyses (Guinier, Porod, Kratky) and obtain a variety of statistics about the sample (like the Rg). It also allows the determination of the P(r) function – the curve from this analysis can be used to determine the general shape of the particle, and can be passed to programs like DAMMIN and DAMMIF which attempt to model the shape of the sample using the SAXS profile (Svergun 1999). Another software suite with a large variety of tools for analysing and visualising scattering data is the ATSAS package (Petoukhov et al. 2012). This includes the Ensemble Optimisation Method (EOM) (Tria et al. 2015) program which generates a large pool of random conformers of the protein of interest, then via a genetic algorithm picks the ensemble of these conformers that best fits the experimental SAXS data – this is particularly useful in visualizing disordered regions in proteins that are hard to investigate by other methods. It can be useful to fit theoretical, mathematical models to experimental data to obtain values for parameters that describe the system. Programs like GENFIT allow for the fitting of a large variety of models that cover a huge range of possible systems, and also allow for combinations of multiple models which is particularly useful if one suspects a system is polydisperse (Spinozzi et al. 2014).

2.10 Specifics of SAXS methods used in this thesis

2.10.1 Data collection

SAXS data was collected at the SAXS1 beamline of the National Synchrotron Light Laboratory (LNLS) Campinas, Brazil in collaboration with Professor Rosangela Itri and Gustavo Scanavachi. Protein samples all had a concentration of 3 mg/ml, and SDS was at 3.5 mM for the SDS alone and protein-SDS samples. All measurements were taken at 22°C. Samples were placed between two mica windows with a 1 mm spacer placed perpendicular to the incident X-ray beam (λ=1.488 Å). Scattered X-rays were recorded with a 2D Pilatus detector. The q range available was from 0.0012 Å-1 to 0.45 Å-1 ($q=(4pi/\lambda)sin\Theta$ where 2Θ is the scattering angle). By recording the X-ray beam intensity decrease during the experiment, the scattering curves were normalised. Buffer solution scattering curves were subtracted from each sample's SAXS curves.

P(q) and I(q) functions for PDB atomic structures were calculated using the SASMOL software package.

2.10.2 Modelling of SAXS data

For protein in the absence of SDS, the sample was modelled as a Gaussian chain . This can be analysed using the method of Debye (Debye 1947). The P(q) calculated from this model depends only on the Rg and is given by the following equation:

$$P(q) = \frac{2(\exp(-u) + u - 1)}{u^2}$$

$$where\ u = q^2 R_g^2$$

To model of the SAXS curves of SDS alone, a two layered spherocylinder model was used (Figure 9 A). The equation for the scattering amplitude, A(q), of this model is described below.

$$A_{sc}(q) = 2\pi L \frac{sin\left(\frac{1}{2}q||L\right)}{\frac{1}{2}q||L}$$

$$\times \sum_{k=1}^{2} \left(\rho_{k,cyl} - \rho_{k+1,cyl}\right) R_{k,cyl}^{2} \frac{J_1\left(q \perp R_{k,cyl}\right)}{q \perp R_{k,cyl}}$$

$$+ 4\pi \sum_{k=1}^{2} \left(\rho_{k,cap} - \rho_{k+1,cap}\right) R_{k,cap}^{3}$$

$$\times \int_0^1 J_0\left(q \perp R_{k,cap} X\right) \frac{sin\left(\frac{1}{2} q|| R_{k,cap} \sqrt{1 - X^2}\right)}{\frac{1}{2} q|| R_{k,cap} \sqrt{1 - X^2}} dX$$

Where L is the length of the cylindrical domain, $R_{k,cyl}$ and $\rho_{k,cyl}$ are the outer radius and the scattering length density (SLD) of the *kth*-cylindrical shell, with $R_{k,cap}$ and $\rho_{k,cap}$ the same measurements for the *kth*-hemispherical caps.

The SLD of solvent ρ_0 is defined so $\rho_0 \equiv \rho_{3,cap} \equiv \rho_{3,cyl}$. $q||$ is the component of scattering vector q parallel to the cylindrical axes and is equal to $q\cos\beta$, and $q\perp$ is the component perpendicular to the cylindrical axes and is equal to $q\sin\beta$, where β is the Debye-Bueche size.

The form factor can be obtained from the A(q) with the following expression:

$$P(q) = < A^2(q) > = \int_0^{\pi/2} d\beta sin\beta [A(q)]^2$$

For SHERP in the presence of SDS, a three layered spherocylinder model was used (Figure 9 B). This model had the same two layers as the two layered model relating to the paraffinic core (inner) and head group region (outer), extended to include a third layer (out) that consists of the hydration shell surrounding the micelle. The three layered spherocylinder model was then combined in a linear fashion with the Debye model for SHERP in solution and two layered spherocylindrical model for free micelle to obtain the final theoretical fit.

All these models were implemented in the GENFIT software which was used to fit the models to the SAXS curves (Spinozzi et al. 2014). GENFIT operates by minimising the reduced χ^2 function between the theoretical scattering profile of the model and the experimental curve, with the output giving values for a set of predefined parameters describing the system.

The SAXS data was obtained as part of collaboration with Professor Rosangela Itri and her group from the University of Sao Paulo, Brazil. Data for the 3KR1 and 4K mutants were collected by Professor Rosangela Itri and Gustavo Scanavachi at the LNLS, Brazil. All other data was collected by the author at the LNLS, Brazil. Modelling of the data with GENFIT was performed by Gustavo Scanavachi.

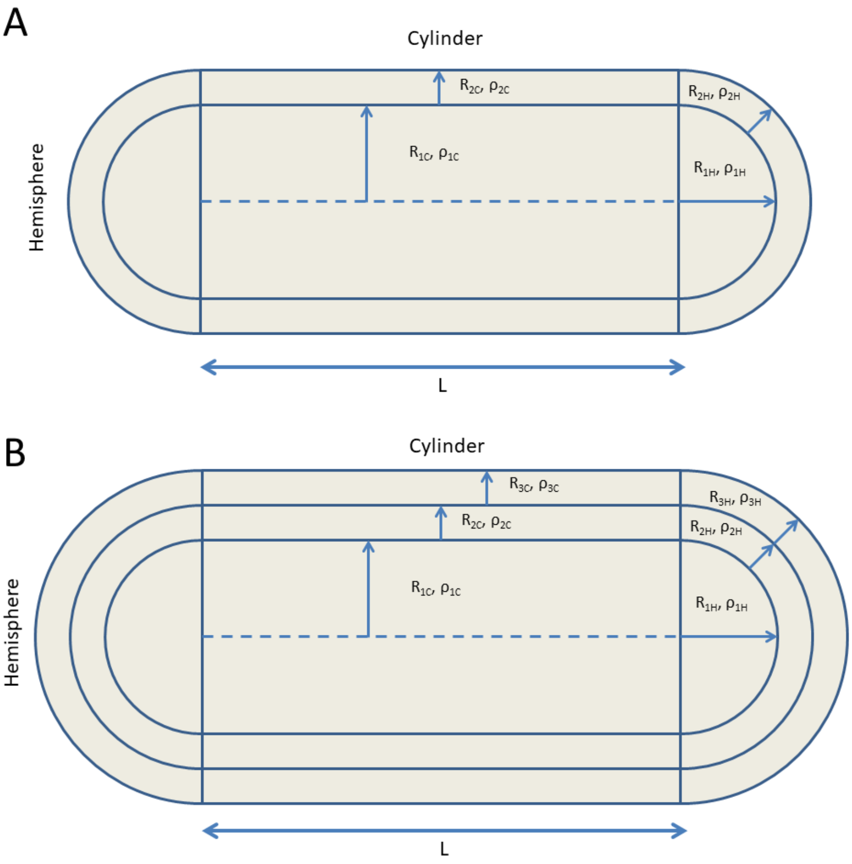


Figure 9: Models used for fitting of SAXS data. A The two layered spherocylinder model used for SDS without protein. B The three layered spherocylinder model used for protein samples in the presence of SDS, which includes a term for free protein and for free micelle (using the two-layer spherocylinder model shown in A). For both models, L is the length of the cylindrical domain, RkC and ρkC are the outer radius and the scattering length density (SLD) of the kth-cylindrical shell, with RkH and ρkH the same measurements for the kth-hemispherical caps.

2.11 Molecular Dynamics - background

Molecular dynamics simulation is a computational method that provides information on the conformational changes of molecules by calculating their time-dependent behaviour (Alder and Wainwright 1959). It allows the study of a huge range of processes that occur in biological systems, including protein stability, protein folding and molecular recognition.

It is particularly useful in the study of targets that are difficult to characterise *in vitro* – for example intrinsically disordered proteins which are not amenable to techniques like X-ray crystallography – or to allow greater insight into mechanistic details that are hard to observe in the lab.

2.11.1 Classical mechanics

Newton's second law shows that force acting on an object is equal to the product of its mass and acceleration. Molecular dynamics simulations use this equation to determine the positions, velocities and accelerations of each particle in a system over time – this is called a trajectory. From the trajectory, time averaged values of macroscopic properties can be calculated (Alder and Wainwright 1959).

Assuming constant acceleration, by integrating the expression for acceleration with respect to time the velocity of a particle can be determined. The same can be done again to the velocity to obtain the position of the particle. Acceleration can be obtained as the derivative of the potential energy with respect to the position. Combining these equations results in the equations used in MD integration algorithms. These assume that the information on the particle can be approximated by a Taylor series expansion. Below, the Verlet algorithm is shown (Verlet 1967):

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2$$

Where r is the position, a is the acceleration (the second derivative with respect to time) and t the time. From this equation we can see that MD simulations are deterministic – this means one only needs the initial positions of the atoms, their velocities and their acceleration to calculate these positions at all other times. The initial positions are often obtained from experimental structures such as an X-ray crystal structure of a protein. Initial velocities are often obtained randomly from a Gaussian or Boltzmann distribution that conforms to the temperature of the simulation. As force can be described using the gradient of the potential energy, the initial acceleration can be derived from this.

The potential energy is a function of the atomic positions of all the particles in the system, and must be solved numerically due to the lack of an analytical solution. Finite difference methods are employed, where discrete time steps are used to solve the function. This gives an approximation of the true solution, and thus several errors are associated with these methods. These can be truncation errors, which are inherent in the method due to the truncation of the Taylor expansion the methods are usually based on. The other common type of error is due to the finite number of digits used by computer arithmetic – round-off errors. Truncation errors can be minimised by decreasing the time step used, and round-off errors can be ameliorated by use of 64 bit precision.

2.11.2 Ensembles

To mimic experimental conditions during simulations it is important to control certain variables. Depending on which variables you decide to keep constant, different statistical ensembles can be produced (Gibbs 1902).

The NVE ensemble, also known as the microcanonical ensemble, controls the number of particles, the volume of the simulation and the total energy of the system. Temperature and pressure are not constant. This ensemble is a poor choice for equilibration as the desired temperature cannot be achieved when no temperature control is present. However it is suitable for data collection.

The NVT ensemble, also known as the canonical ensemble, controls the number of particles, the volume of the simulation and the temperature. Energy and pressure are not controlled. This ensemble is appropriate for simulations in vacuum or without boundary conditions, as the lack of boundary conditions means pressure is not defined.

The NPT ensemble controls both temperature and pressure, allowing volume and energy to vary over time. This allows for simulations that accurately mimic experimental conditions. It can also be used to equilibrate systems to the correct temperature and pressure before data collection in the NVT and NVE ensembles.

2.11.3 Force fields

To accurately model the forces acting upon an atom in a molecular system, the interactions between particles must be defined mathematically. In most MD applications these interactions are modelled empirically to create a mathematical model for the potential energy of the system called the force field. The force fields are empirical as the terms have been derived from experimental studies of molecules.

The force field generally consists of six types of interactions:

$$U = VLJ + VEL + VBS + VAB + VPT + VIT$$

The first two terms in the potential are called the non-bonded interactions. The first term, *VLJ*, describes the potential energy between two non-bonded atoms based on their separation, accounting for both attractive forces (van der Waals and London dispersion) and repulsive forces (due to Pauli repulsion). It is often referred to as the Lennard-Jones potential (Jones 1924).

$$V(r)=4\epsilon[(\sigma/r)12-(\sigma/r)6]$$

Where:

V is the intermolecular potential between the two atoms or molecules.

ϵ is the well depth and a measure of how strongly the two particles attract each other.

σ is the van der Waals radius, which gives a measurement of how close two nonbonding particles can get.

r is the distance of separation between both particles (measured from the centre of one particle to the centre of the other particle).

The interactions described by the Lennard Jones Potential are weak and short ranged.

The second term, *VEL*, defines the potential energy due to electrostatic interactions according to Coulomb's Law:

$$F = k_e \frac{q1q2}{r^2}$$

Where *q1* and *q2* are the signed magnitudes of the particle charges, *r* is the distance between the particles and k_e is coulombs constant (8.99 x 10^9 N m^2 C^{-2}). Electrostatic interactions are stronger and have a much longer range than van der Waals interactions.

VBS, *VAB*, *VPT* and *VIT* model the contributions due to covalent bonding between atoms, and are referred to as the bonded interactions. *VBS* models the elastic interaction between two covalently bonded atoms, *VAB* the "angle-bending" energy between three atoms forming a stable angle, *VPT* and *VIT* the proper and improper torsional energies due to proper and improper dihedrals formed between four atoms. These terms do not permit covalent bonds to break, so cannot be used to simulate chemical reactions

A large number of force-fields are used in research today. All-atom force fields give parameters for every atom type in the system and treat each atom discretely (e.g. CHARMM36 (Best et al. 2012), OPLSAA (Jorgensen et al. 1996), Amber (Cornell et al. 1995)). United-atom force fields treat methyl groups as a single interaction centre, combining the carbon and hydrogens into one particle (e.g. OPLSUA (Jorgensen et al. 1996)). Further simplification of the atomic representation leads to coarse-grained force fields (e.g. MARTINI (Marrink et al. 2007)) which combine multiple heavy atoms into a single interaction site, allowing for longer simulations at lower computational cost.

2.11.4 Programs and Utilities

Several program suites exist for preparation and execution of MD simulations. GROMACS (Berendsen et al. 1995; Van der Spoel et al. 2005) is perhaps the most commonly used and well known MD software suite. It provides a large set of tools for preparation of systems before simulations, allows highly parallel high performance calculations to be run, has support for GPUs which can greatly increase simulation speed, allows the use of a large number of force fields including coarse grained models and a comprehensive toolkit for analysis of trajectories once a simulation is complete. Other packages include LAMMPS (Plimpton 1995), CHARMM (Brooks et al. 1983) and NAMD (Kalé et al. 1999).

2.12 Specifics of molecular dynamics methods used in this thesis

2.12.1 Molecular Dynamics – Unfolding simulations in aqueous solution

The NMR structure for SHERP in SDS (PDBID: 2X43) was obtained from the PDB, consisting of ten models (Moore et al. 2011). Models were assessed for energetic stability using FoldX (Schymkowitz et al. 2005) and the most energetically stable, model three, was chosen for use in the simulation. GROMACS (Berendsen et al. 1995; Van der Spoel et al. 2005) was subsequently used to generate a system using the OPLS/AA force field (Jorgensen et al. 1996) and TIP3 explicit solvation model (MacKerell et al. 1998). Na^+ were added to ensure the system had zero charge. The system underwent a two-step equilibration. The first step was 10 picoseconds with a 0.005 ps time-step with strong restraints on all bonds in the protein. A second 500 ps long equilibration (0.002 ps time-step) was performed. The equilibrated system was then unrestrained and a production run at 323 K was carried out for 1 µs with a 0.002 ps time-step. For the first equilibration step, the Berendsen weak-coupling algorithm was used for both temperature and pressure coupling. For the second equilibration step and production run the V-rescale algorithm was used for temperature coupling and the Berendsen for pressure coupling.

The PM starting structure was generated by changing the residue name column in the PDB file for model 3 of the NMR structure before minimisation by HIPPO (J.P. Ulmschneider and M. B. Ulmschneider 2009) – the program repairs the altered residue and the mutant structure is produced. The modified starting structure was then subject to the same simulation regime as above.

2.12.2 Conventional MD simulations of SHERP-SDS systems

A system containing the NMR solved ordered structure of wild-type SHERP and 80 SDS molecules (Table 4) with explicit TIP3 solvation and neutralising ions was generated using the CHARMM-GUI webserver (Jo et al. 2008), and a 7 step equilibration was performed with restraints applied to the individual components of the system (Protein backbone, side-chains and the SDS sulphur atoms). These restraints were gradually relaxed over the course of the equilibration (Table 3). A short production run of 30ns was performed using the GROMACS package (version 5.1) (Berendsen et al. 1995). A 0.002 ps time step was used, and the V-rescale thermostat and Parrinello-Rahman barostat (Parrinello and Rahman 1981) were used for temperature and pressure coupling. The system split into two micelles, a 51 SDS micelle bound to the protein and a separate, smaller 29 SDS micelle. The protein-bound 51 SDS micelle was used as a starting structure for a longer simulation of the system. As before, the SHERP-Micelle molecule was solvated with TIP3 water and ions were added to neutralise charge. The system was equilibrated with a 0.001 ps time-step for 250 ps before a 200 ns production run was performed. For all simulations the CHARMM36 force field (Best et al. 2012) was employed.

2.12.3 Replica Exchange MD (REMD) simulations of unfolded SHERP in solution

20 replicas with the following temperature regime was chosen for REMD: 280.00 K, 290.43 K, 301.17 K, 312.23 K, 323.60 K, 335.29 K, 347.34 K, 359.74 K, 372.49 K, 385.68 K, 399.19 K, 413.11 K, 427.45K, 442.20 K, 457.36 K, 472.98 K, 489.04 K, 505.59 K, 522.62 K, and 540.14 K. These temperatures were chosen using the webserver available at http://folding.bmc.uu.se/remd/ (Patriksson and van der Spoel 2008). A short 100 ps simulation was run prior to production to ensure the replicas exchanged with a probability of 0.2-0.5. A 73 amino acid straight chain of SHERP including the hexahistidine tag was generated as the starting structure for each replica. Using GROMACS (Berendsen et al. 1995; Van der Spoel et al. 2005), each replica was then equilibrated to the required temperature for 250 ps using the generalised Börn implicit solvent model (Still et al. 1990) and CHARMM36 force field (Best et al. 2012). After equilibration the REMD simulation was started with exchanges attempted every 2 ps. Each replica was simulated for 50 ns with a time-step of

0.002 ps for a total of 1 μs across 20 replicas. Temperature coupling was achieved using the V-rescale thermostat. Pressure coupling was not employed due to the use of implicit solvent and no periodic boundary conditions. To check whether conformational space was exhaustively sampled, the REMD simulation was repeated with the same conditions and ensemble properties compared, and both simulations were clustered using the *kmeans* method to see if the clusters obtained were similar.

2.12.4 Analysis of molecular dynamics simulations

For the unfolding calculations, REMD and SHERP-SDS simulations, analysis of the radius of gyration, phi/psi angles, root mean square deviation (RMSD), root mean square fluctuation (RMSF) and potential energy were performed using utilities included in the GROMACS suite of applications (Berendsen et al. 1995). Secondary structure calculations were performed with the VMD timeline implementation of STRIDE (Heinig and Frishman 2004; Humphrey et al. 1996).

The radial distribution function (RDF) of a system describes how, on average, particles in a system are radially packed around each other. The GROMACS utility *gmx rdf* was used. RDFs were calculated for detergent head groups S atoms, hydrophobic tail carbon atoms, protein atoms, ions and water molecules with respect to the COM.

Equil. Step	Integrator	Temperature/Pressure Coupling	Time step (ps)	No. Steps	Restraints (backbone)	Restraints (side-chains)	Restraints (SDS S Atom)
1	steep	n/a	n/a	5000	4000	2000	1000
2	md	Berendsen/None	0.001	25000	4000	2000	1000
3	md	Berendsen/None	0.001	25000	2000	1000	0
4	md	Berendsen/Berendsen	0.001	25000	1000	500	0
5	md	Berendsen/Berendsen	0.002	50000	500	200	0
6	md	Berendsen/Berendsen	0.002	50000	200	50	0
7	md	Berendsen/Berendsen	0.002	50000	50	0	0

Table 3: Equilibration regime for all atom simulations of protein and SDS systems. The backbone and sidechain restraints are applied to the x,y and z coordinates of the atoms. The restraints on the SDS sulphur atom only apply to the z coordinate allowing unrestricted movement in the x-y plane.

Atom Name	Charge	Mass
S (SO_4^-)	1.330	32.0600
OS1 (SO_4^-)	-0.280	15.9994
OS2, OS3, OS4 (SO_4^-)	-0.650	15.9994
C1 (CH_2 adjacent to SO_4^-)	-0.280	12.0110
C2 – C11 (CH_2 in tail)	-0.180	12.0110
C12 (Terminal CH_3)	-0.270	12.0110
All hydrogen	0.090	1.0080

Table 4: CHARMM36 Parameters for SDS used during GROMACS simulations.

3.0 Chapter 3: Effect of mutations on the disorder of SHERP in solution

3.1 Introduction

In order to examine the effects of sequence differences on SHERP secondary structure content, site specific mutations were designed and characterised with the aim of decreasing the disorder content of the protein in solution. Residues 2-57 were mutated to each amino acid *in silico* and the resulting per-residue disorder profiles were predicted using DISOPRED2 (Ward et al. 2004b). Four mutants were chosen as a result of this analysis and were cloned, purified and expressed, along with the wild-type (WT). Their secondary structures were determined by SRCD spectroscopy. MD simulations of the WT and a mutant with a large predicted decrease in disorder were also performed, along with a SAXS experiment of the WT in solution, with the aim of determining whether the protein was in an extended confirmation in solution, or adopted a molten globule conformation.

Another construct with multiple mutations was designed to introduce *i→i4* salt bridges while maintaining amino acid composition in order to stabilise the type of alpha helical structure that may transiently form in the disordered state. This mutant, referred to henceforth as the “permutant” (PM) was also characterised by SRCD spectroscopy and MD.

3.2 Results

3.2.1 Cloning of SHERP gene construct

The initial transformation and growth of both pEX-A2-SHERP and pET28a+ were successful. Purification of the plasmid DNAs resulted in yields of ~47 ng/µl and ~67 ng/µl respectively. Difficulties were encountered during the restriction digest and ligation steps. Initial attempts at digestion used 10 µl of plasmid in a 50 µl reaction, the 27 µl difference made up with ddH2O. When the digestion products were run on an agarose gel, the bands for the SHERP insert were very faint and the ~150 bp bands expected from the digestion of pET28a+ (DNA between the Xho1 and Nco1 sites) was not visible at all. Purification of the DNA from the gel yielded negligible and impure DNA product. The digest was repeated again, increasing the volume of pEX-A2-SHERP plasmid DNA in the reaction to 27 µl and pET28a+ DNA to 20 µl. A stronger band was observed for the SHERP insert. Approximately 6 ng/µl of pET28a+ DNA was purified from the gel, but a negligible amount was retrieved for the SHERP insert. A third digest was performed using 37 µl of both plasmids, with similar results when purified. All ligations attempted with the above purified DNA failed to transform, with no colonies growing.

Although a small amount of pET28a+ was purified from the gel, SHERP's small size (203 bp) resulted in faint bands which were hard to identify and purify. It became apparent the concentration of plasmid DNA used in the restriction digest was not high enough. To that end, a larger scale MidiPrep was performed and higher DNA concentrations (67 ng/µl pEX-A2-SHERP and 124 ng/µl pET28a+) with greater purity were obtained. The restriction digests were performed again with 37 µl of pEX-A2-SHERP and 25 µl of pET28a+ (Figure 10A). The SHERP insert can be seen much more clearly and the ~140 bp fragment removed from pET28a+ was also visible – suggesting that much more DNA had been digested. Purification of the gel-bound DNA gave 16.8 ng/µl of SHERP and 173 ng/µl of pET28a+ backbone. A transformation of 10 µl of ligation reaction into DH5α cells resulted in a number of colonies. Plasmids from eight colonies were digested with Nco1 and Xho1 and analysed with gel electrophoresis, and all contained the SHERP insert (Figure 10B). Three of these plasmids were sent for sequencing and all were found to contain the SHERP insert in the correct position.

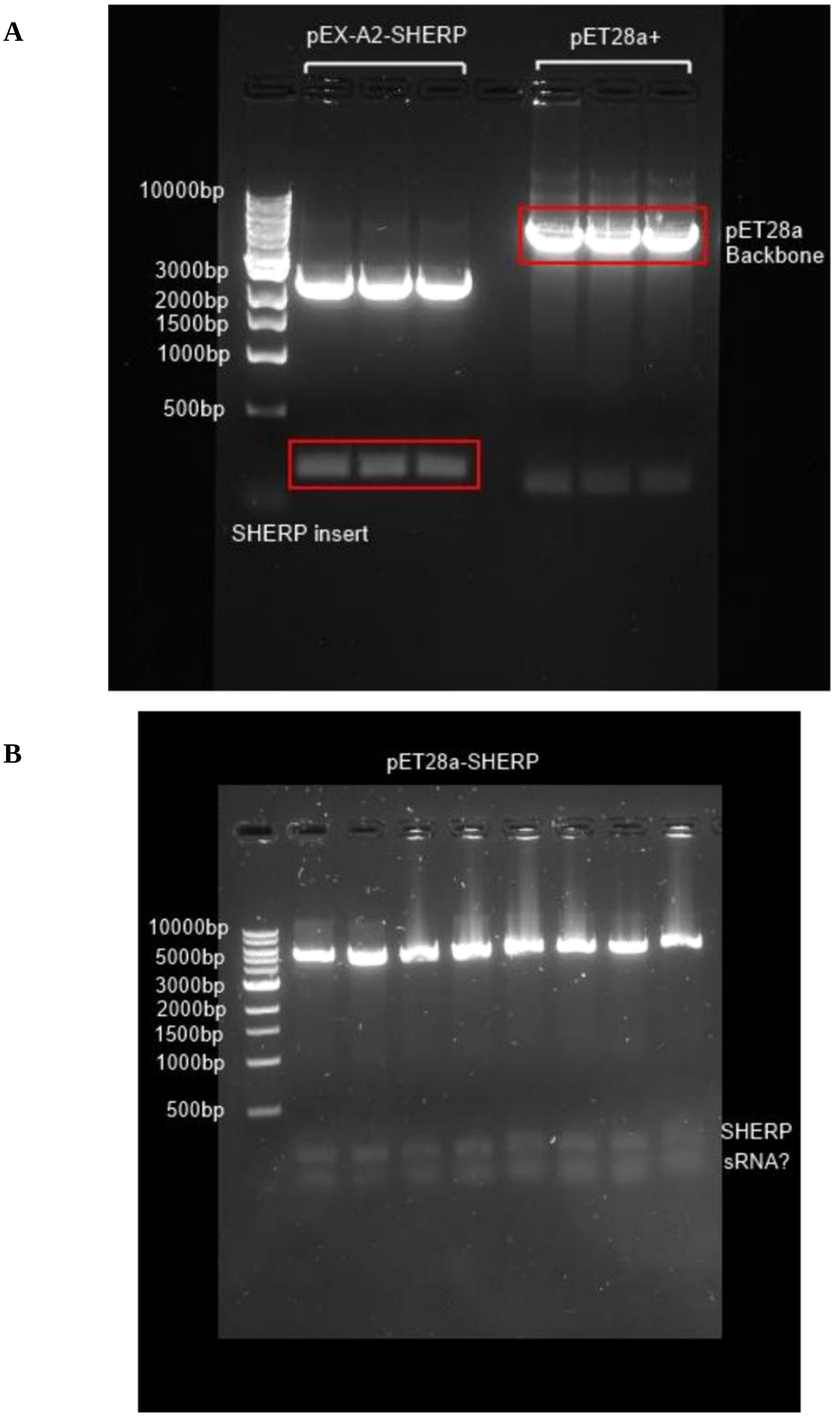


*Figure 10: Cloning and Purification of pET28a+-SHERP. **A** Restriction digest of pEX-A2-SHERP and pET28a+. Lane 1 corresponds to 10 kb DNA ladder (New England Biolabs). Lanes 2-4 correspond to a pEX-A2-SHERP and lanes 6-8 correspond to pET28a+, both digested with Xho1 and Nco1. The ~200 bp bands of SHERP and the ~5000 bp bands of the pET28a+ backbone (both in red boxes) were cut from the gel and purified. **B** Gel of digest with Xho1 and Nco1 of pET28a-SHERP construct showing that SHERP was successfully ligated into plasmid. DNA was run on 1% agarose TAE gel at 100 V for 50 minutes.*

3.2.2 Protein expression and purification

The SDS-PAGE gel of the fractions resulting from the initial affinity chromatography step of the purification showed that almost all the contaminants present were removed (Figure 11A). The elution fractions containing the protein were pooled and concentrated before the final size exclusion chromatography step. The SDS-PAGE gel shows the protein is clearly separated from the remaining contaminant, allowing concentration of pure protein for experiments (Figure 11B).

3.2.3 Design of permutant Sequence

A rational design approach was taken for the selection of which mutations to make for the permutant (PM) experimental construct. This is a mutant sequence where residues were swapped or shifted so as to maintain the same overall amino acid composition. Charged residues (E, D, K and R) exposed to solvent in the NMR ordered structure (PDBID: 2X43) were identified visually within the helical regions of the protein using Pymol (Schrödinger 2010). Suitable amino acid swaps were found that would lead to potential $i \rightarrow i+4$ salt bridge formation (Figure 12). The amino acid substitutions made in the first alpha helix were identified as: K10N, N11K, N19D, D18K, and K17N (Figure 13).

However, suitable permutations in the second helix were harder to identify, so several locations where charged residues could be introduced to lead to potential salt bridges were identified along with the single swap found. These mutations were A35E, A41K, V42K, V46E and E47V (Figure 13).

3.2.4 Single mutant's effect on disorder

Single mutant sequences of SHERP (Uniprot ID: Q9XTN3) were generated by replacing residues 2-57 with each of the twenty amino acids in turn, to yield 1064 unique sequences. DISOPRED2 (Ward et al. 2004b) was then used to evaluate the predicted disorder of each of these sequences by generating a per-residue disorder profile.

.

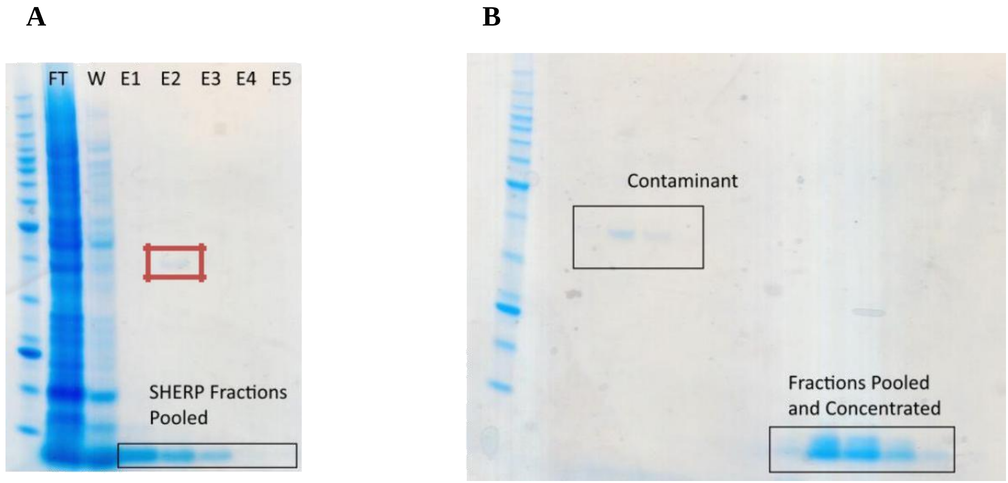


Figure 11: Purification of SHERP. ***A*** *SDS PAGE gel showing fractions from affinity chromatography step. FT = flowthrough, W = wash step, E1 – E5 elution steps. The sample was quite clean after this step, with a single contaminant in the red box. SHERP fractions (black box) were pooled for size exclusion chromatography (SEC).* ***B*** *SDS PAGE showing fractions SEC. Contaminant was larger than SHERP so clean SHERP fractions were isolated and concentrated.*

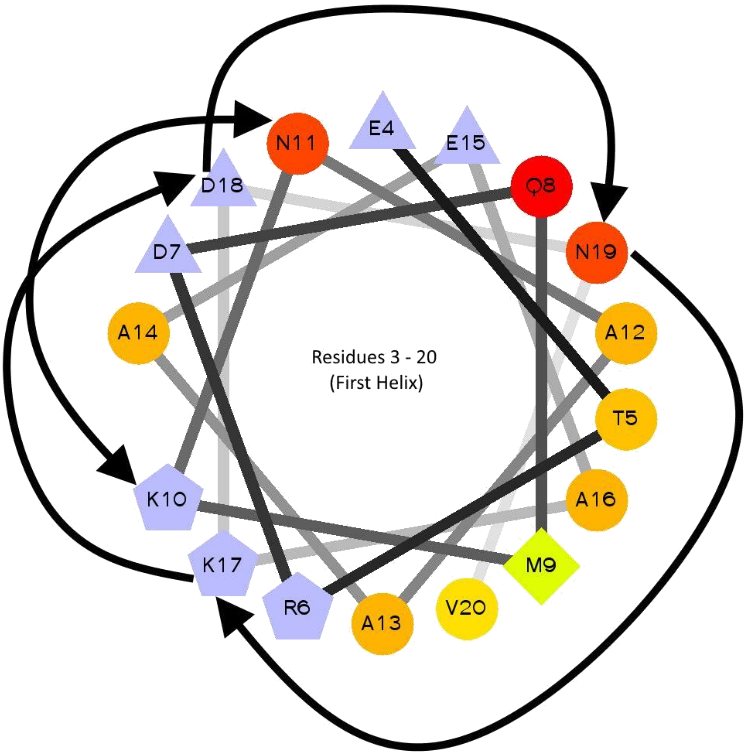


Figure 12: Permutants in the first helix of SHERP. A helical wheel representation of residues 4-20 of SHERP illustrating which residues were swapped in this region to introduce potential i-i4 salt bridges. Positively charged residues were shown in blue, negatively charge residues in red, polar residues in orange and hydrophobic residues in yellow.

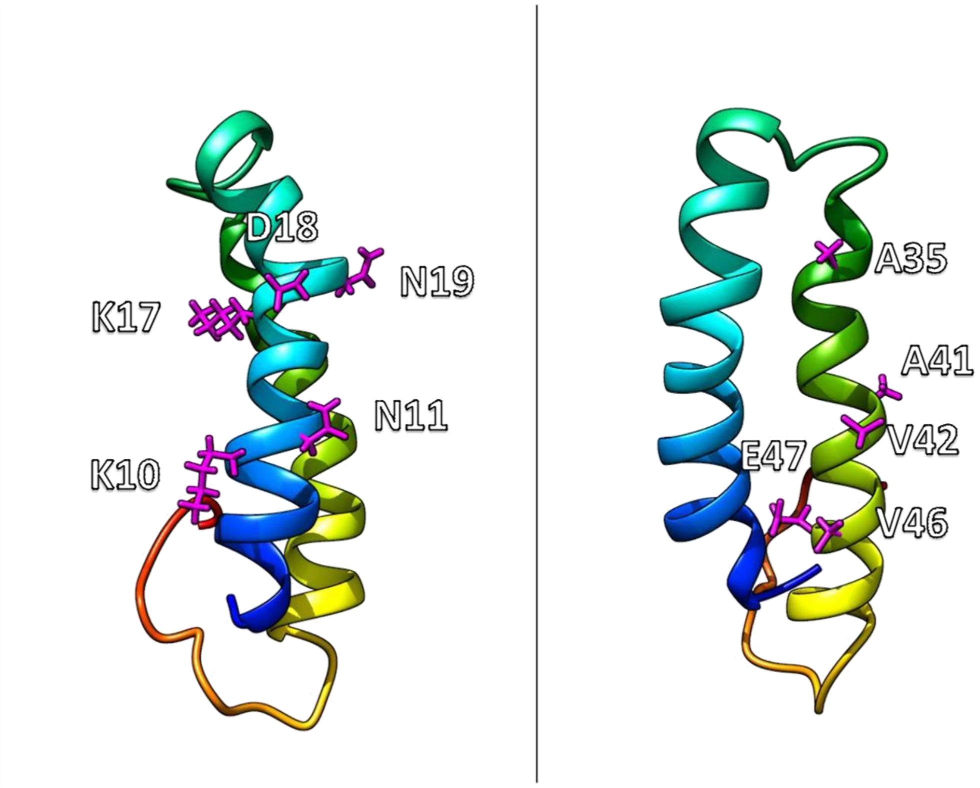


Figure 13: Locations of the residues mutated in the PM construct (magenta). The mutations were as follows: : K10N, N11K, N19D, D18K, K17N, A35E, A41K, V42K, V46E and E47V. The backbone is coloured from blue at the N-terminus to red at the C-terminus. Cartoon representation is used to show secondary structure present. The structure was visualised using Chimera (Pettersen et al. 2004).

The point mutations' predicted effects on disorder were found to be both position- and amino acid type-dependent (Figure 14). For all amino acid substitutions, no changes in disorder were predicted for the first 20 positions in the sequence with the exception of Trp substitutions, which showed small decreases for residues 18-19. Larger decreases in disorder were seen between residues 20 and 40 for substitutions of Cys, Phe, Ile, Leu, Val, Trp and Tyr - the largest changes being Trp substitutions at positions 29 and 31 (30 and 31 disordered residues respectively). Small increases in disorder were seen for Ala, Gly, His, Lys, Asn, Pro, Glu, Arg, and Ser between residues 38 to the C-terminus of the protein.

To investigate the link between the mutant's disorder-propensity and its predicted effect on disorder, the TOP-IDP disorder propensity scale (Campen et al. 2008) was used (Table 5). The mutations predicted effect on disorder correlates weakly with its disorder-promoting propensity (Figure 15). The weak correlation may have been due to the position of the mutation having a large effect on the prediction. From the plot it can be seen that amino acids with a low TOP-IDP score had broader distributions and decreased disorder the most, while those with neutral or positive TOP-IDP scores had much smaller effects on the protein's predicted disorder. WT SHERP was predicted to be highly disordered (81%) and has a very high proportion of polar and charged residues (63%) considered disorder-promoting on the TOP-IDP scale, which is characteristic of disordered and extended proteins. It could be that the introduced disorder-promoting residues were often like-for-like substitutions and had small effects, whereas the introduction of an order-promoting residue in place of a disorder-promoting one led to larger predicted effects.

Four mutants were chosen for characterisation based on the DISOPRED2 predictions – K28F, D29W, S43F and A12E (Figure 14 [circled in green]) and their positions on the structure are shown in Figure 16. DISOPRED2 predicted large decreases in disorder for the K28F and D29W mutants, and a smaller decrease for the S43F mutant. No change in disorder was predicted for the A12E mutant. SRCD spectra of the mutants and WT in the absence of SDS were collected. For all the single mutants and the WT, the spectra have the characteristic minima of disordered proteins at 200 nm at 20˚C (Figure 17) and 90˚C (Figure 18).

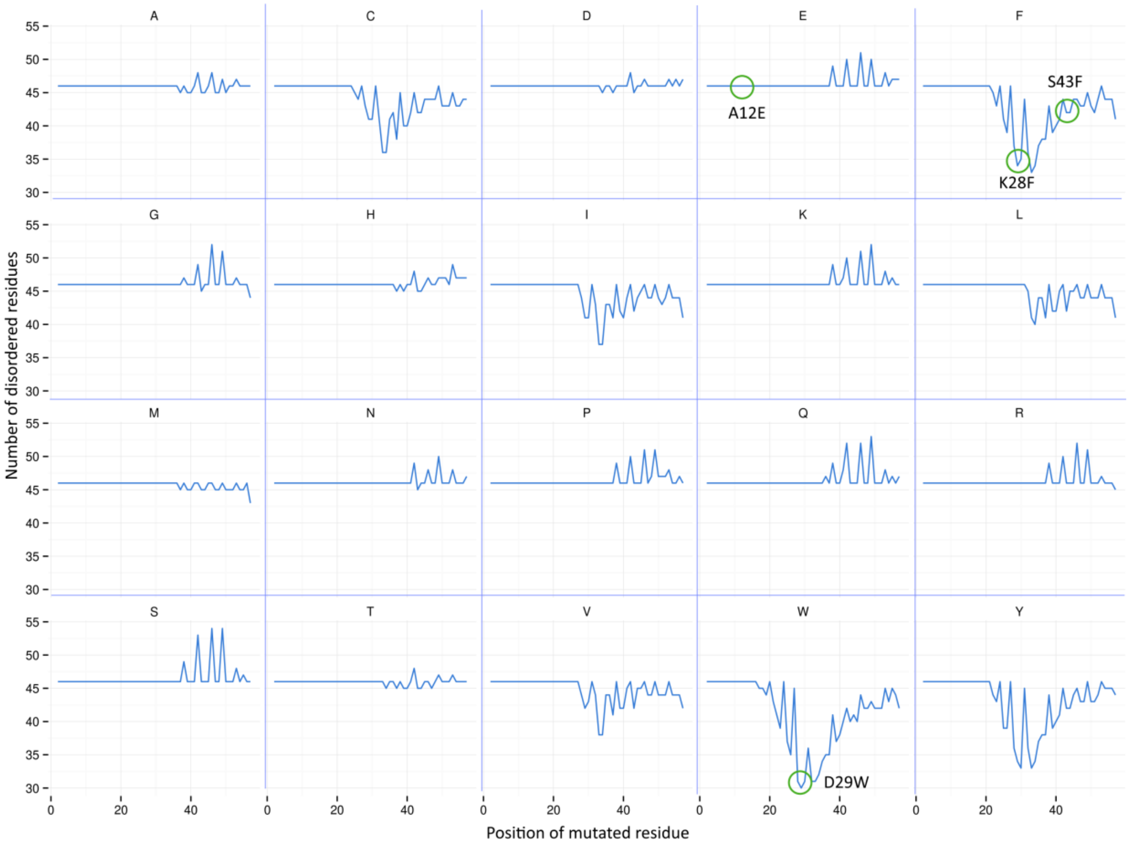


Figure 14: Disorder predictions for all single mutants of SHERP. For each of the 20 amino acids, the DISOPRED2 predicted number of disordered residues at each position is shown. For all amino acid substitutions, no changes in disorder were predicted for the first 20 positions in the sequence with the exception of Trp substitutions, which showed small decreases for residues 18-19. Larger decreases in disorder were seen between residues 20 and 40 for substitutions of Cys, Phe, Ile, Leu, Val, Trp and Tyr - the largest changes being Trp substitutions at positions 29 and 31 (30 and 31 disordered residues respectively). Small increases in disorder were seen for Ala, Gly, His, Lys, Asn, Pro, Glu, Arg, and Ser between residues 38 to the C-terminus of the protein. The single mutants chosen for expression and characterisation are shown with the green circles.

Disorder-promoting amino acid residues

A	**G**	**R**	**D**	**H**	**Q**	**K**	**S**	**E**	**P**
0.06	0.166	0.180	0.192	0.303	0.318	0.586	0.341	0.736	0.987

Order-promoting amino acid residues

W	**F**	**Y**	**I**	**M**	**L**	**V**	**N**	**C**	**T**
-0.884	-0.697	-0.510	-0.486	-0.397	-0.326	-0.121	0.007	0.02	0.059

Table 5: TOP-IDP amino acid disorder propensity scale (Campen et al. 2008). *The TOP-IDP scale was developed by analysing the ability of 517 different amino acid scales to discriminate between order and disorder, using a simulating annealing approach to identify improved, alternative scale values. Negative values indicate more order-promoting residues, and positive values indicate more disorder-promoting residues.*

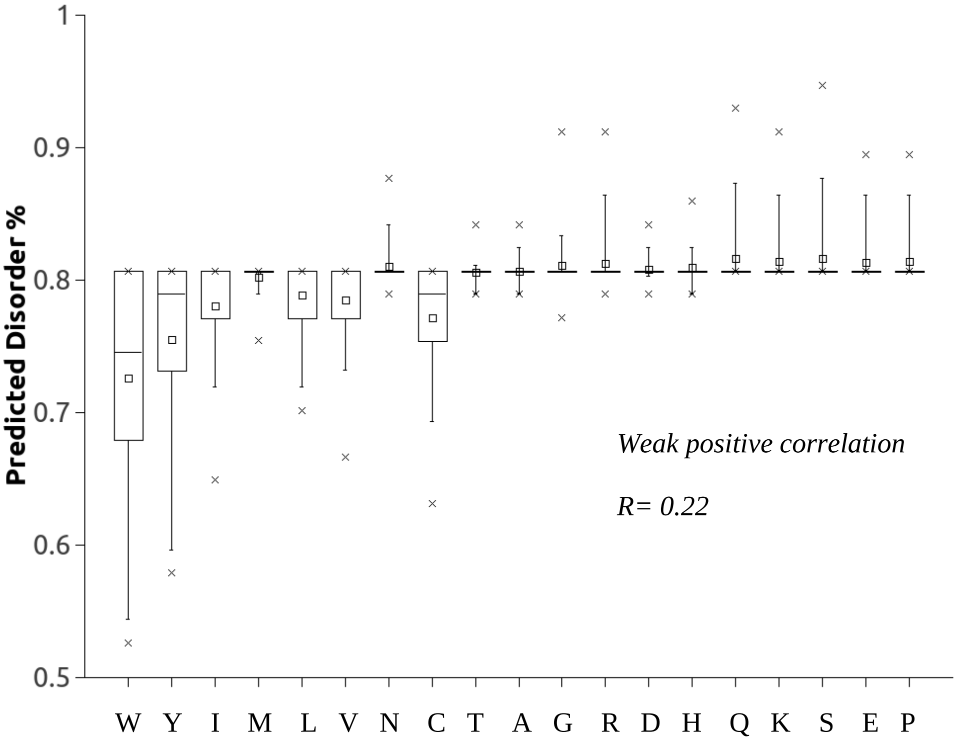


Figure 15: Predicted disorder score distribution with respect to amino acid substitution. Boxplots of DISOPRED 2 disorder predictions were shown for each amino acid substitution. Amino acids were ordered according to the TOP-IDP disorder-promoting propensity scale from least to most disorder-promoting (left to right). Pearson correlation coefficient between predicted disorder and score on the TOP-IDP scale is also shown (r = 0.22).

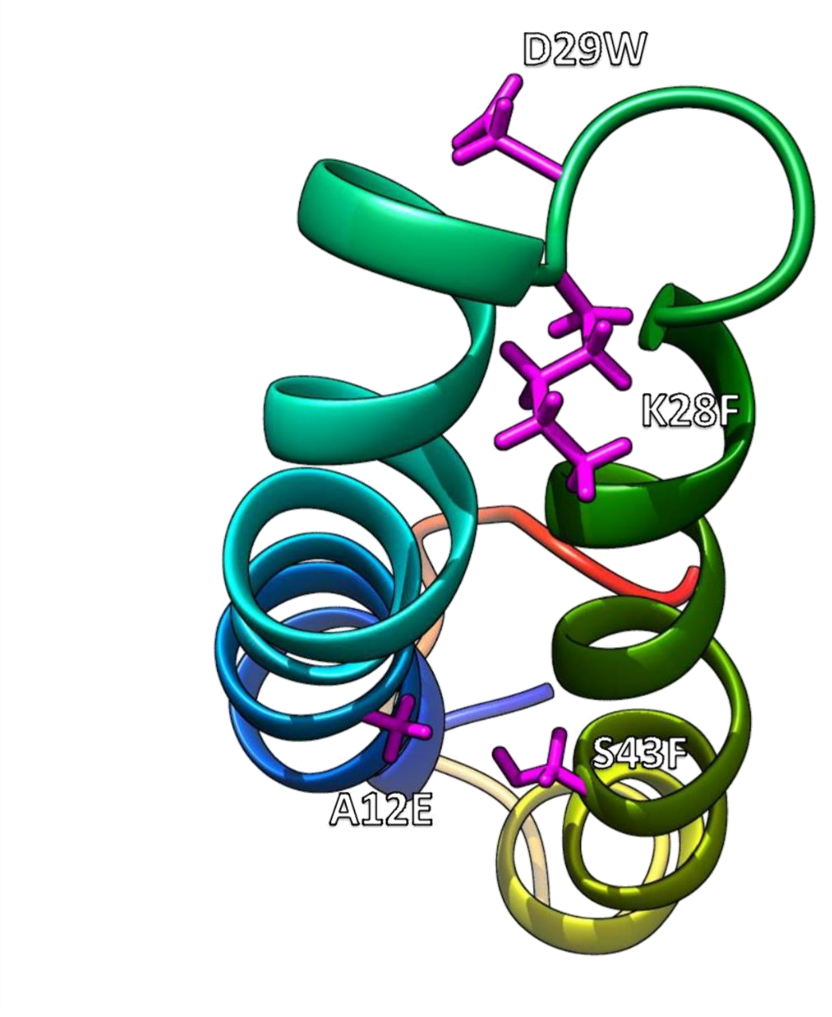


Figure 16: Locations of the single mutants selected for characterisation (magenta). The sidechain shown is the original, native residue before mutation. Backbone is coloured from blue at the N-terminus to red at the C-terminus. Cartoon representation used to show secondary structure present. Structure visualised using Chimera (Pettersen et al. 2004).

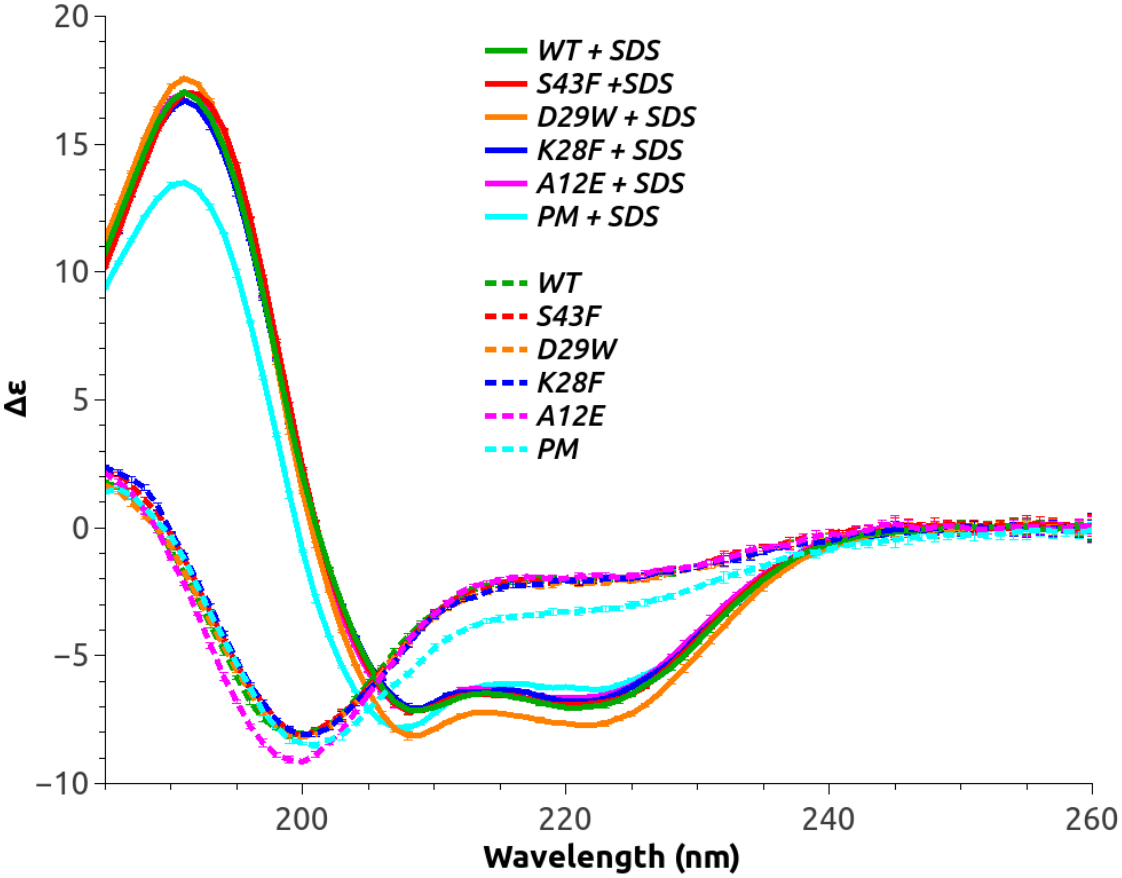


Figure 17: SRCD spectra of WT, single mutants and PM in 50mM Na phosphate (pH 7.0) in absence (dashed) and presence (solid) of 0.1% SDS at 20˚C. A small increase in helicity can be seen for the D29W mutant (orange, solid) vs the WT in the presence of SDS, as can an increase in helicity of the PM in the absence of SDS (cyan, dashed) vs the WT.

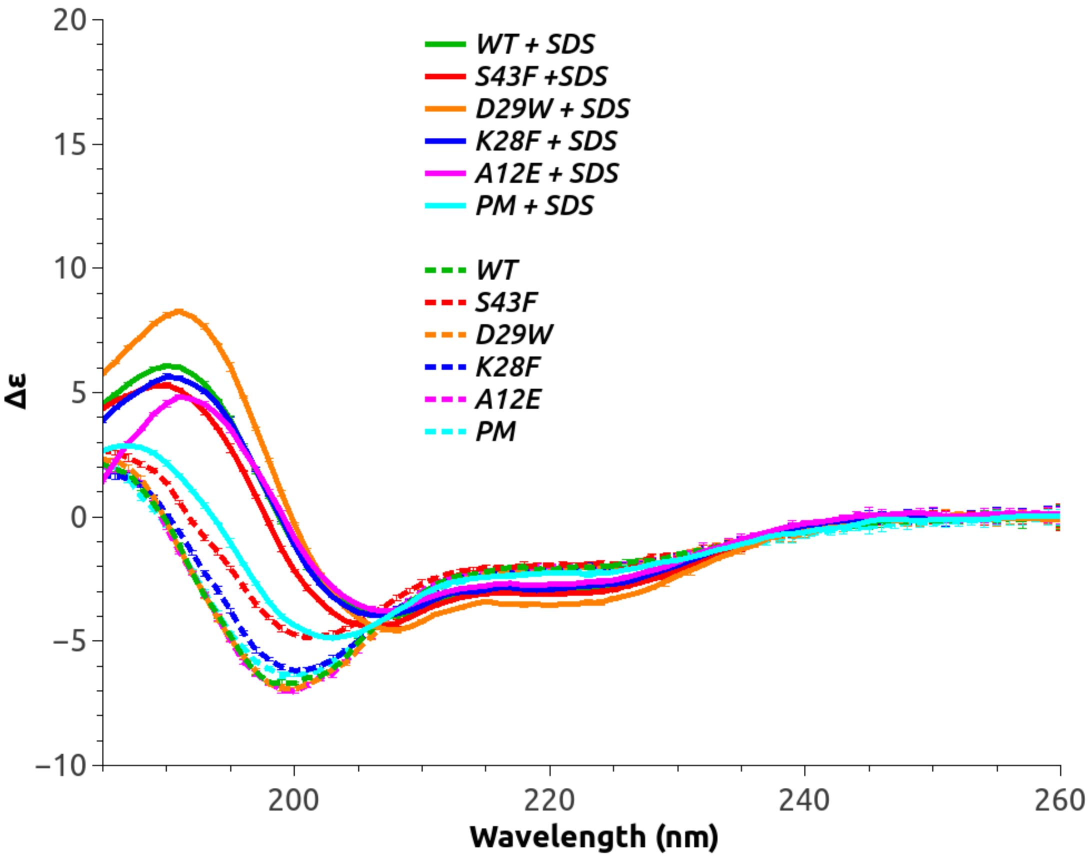


Figure 18: SRCD spectra of WT, single mutants and PM in 50mM Na phosphate (pH 7.0) in absence (dashed) and presence (solid) of 0.1% SDS at 90˚C. The D29W mutant is more stable than any other construct in the presence of SDS, maintaining a greater proportion of helicity(orange, solid). The S43F mutant in the absence of SDS gains secondary structure upon heating (red, dashed).

The DichroWeb (Whitmore and Wallace 2004) analysis of these constructs at 20˚C in the absence of SDS (Table 6) showed the K28F, D29W and S43F mutants were not significantly different to the WT. At 90˚C no significant changes were observed for the WT, K28F and A12E mutants compared to the results at 20˚C from the DichroWeb analysis (Table 6) or relative CD at 200 nm (Figure 19A), suggesting that no further unfolding takes place. The D29W mutant showed a decrease in helical content and an increase in other – it became more disordered as the temperature increased. This loss of stability ran contrary to the DISOPRED2 results which predicted this mutant would lead to the largest increase in order. The S43F mutant however has a smaller minima at 200nm at 90C and a smaller decrease in other was obtained (WT 75% ± 4 vs S43F 63% ± 2). This could be due to increased beta sheet content (rather than helix) at higher temperatures (Table 6) – it may seem unusual that a protein would gain secondary structure as temperature increases, but this behavior has been observed before with other IDPs (Gast et al. 2003; Uversky et al. 2001).

The DISOPRED2 predictions would not be expected to be able to predict the behavior of the mutants in the ordered state, although lower disorder might indicate greater stability in the ordered state. In the presence of 0.1% SDS at 20˚C, the SRCD spectra for all samples showed the peak at 191 nm and minima at 208 nm and 222 nm typical of highly alpha helical proteins (Figure 17). This was confirmed by the DichroWeb analysis that showed all samples were 66-73% helical with the largest increase in helicity being attributed to the D29W mutant – 73% compared to the WT's 66% (Table 6). This mutant was predicted to lead to one of the largest decreases in disorder. The helical content of the WT is in excellent agreement with the solution NMR and SRCD spectroscopy results in the literature (Moore et al. 2011). However, at 90˚C some differences were observed between the WT and single mutants (Figure 18). The thermal melt (Figure 19B) shows no large differences in the decreases in relative CD between the WT, A12E and K28F mutants, and the DichroWeb analysis shows they all retain ~33-34% helicity. However, the smaller reduction in relative CD at 191 nm for

Temperature	Sample	Helix	Strand	Turn	Other	NRMSD
20C	WT	8 ± 1	6 ± 7	8 ± 9	78 ± 17	0.00 ± 0.03
	WT + SDS	66 ± 2	5 ± 1	8 ± 1	20 ± 3	0.01 ± 0.01
	12E	7 ± 0	6 ± 7	7 ± 8	80 ± 16	0.00 ± 0.03
	12E+SDS	68 ± 2	4 ± 1	7 ± 1	21 ± 2	0.01 ± 0.01
	28F	12 ± 4	6 ± 8	8 ± 9	73 ± 22	0.00 ± 0.03
	28F + SDS	71 ± 4	4 ± 0	6 ± 2	20 ± 2	0.01 ± 0.01
	29W	13 ± 6	7 ± 8	7 ± 8	73 ± 22	0.00 ± 0.03
	29W + SDS	73 ± 3	3 ± 1	6 ± 1	18 ± 2	0.01 ± 0.01
	43F	12 ± 5	9 ± 11	9 ± 11	69 ± 27	0.00 ± 0.03
	43F + SDS	66 ± 3	6 ± 2	8 ± 1	20 ± 3	0.01 ± 0.01
	PM	29 ± 14	9 ± 12	11 ± 6	51 ± 32	0.00 ± 0.04
	PM + SDS	69 ± 3	4 ± 2	8 ± 3	19 ± 2	0.03 ± 0.03
90C	WT	7 ± 1	6 ± 6	7 ± 8	80 ± 15	0.01 ± 0.03
	WT + SDS	34 ± 1	16 ± 2	21 ± 1	29 ± 1	0.05 ± 0.05
	12E	8 ± 0	5 ± 7	7 ± 8	80 ± 16	0.01 ± 0.03
	12E+SDS	33 ± 1	18 ± 1	20 ± 2	29 ± 0	0.03 ± 0.02
	28F	12 ± 6	7 ± 8	7 ± 8	75 ± 20	0.01 ± 0.03
	28F + SDS	34 ± 1	17 ± 2	20 ± 2	29 ± 1	0.04 ± 0.04
	29W	8 ± 1	5 ± 7	6 ± 8	80 ± 17	0.01 ± 0.02
	29W + SDS	44 ± 2	12 ± 1	17 ± 2	27 ± 1	0.02 ± 0.01
	43F	17 ± 0	10 ± 1	10 ± 0	64 ± 1	0.01 ± 0.03
	43F + SDS	34 ± 1	16 ± 2	21 ± 1	29 ± 1	0.05 ± 0.05
	PM	11 ± 5	9 ± 11	11 ± 8	69 ± 25	0.01 ± 0.03
	PM + SDS	18 ± 1	22 ± 0	22 ± 3	37 ± 3	0.04 ± 0.04

Table 6: DichroWeb analysis of WT, single mutants and PM construct CD spectra at 20°C and 90°C in the presence and absence of 0.1% SDS. The secondary structure and stability of each mutant was similar to the WT with the following exceptions: The D29W mutant which showed increased helicity and stability in the presence of 0.1% SDS; the S43F mutant which showed an increase in secondary structure content as the temperature increased; and the PM which displayed a greater % helicity than the WT at 20°C in the absence of SDS.

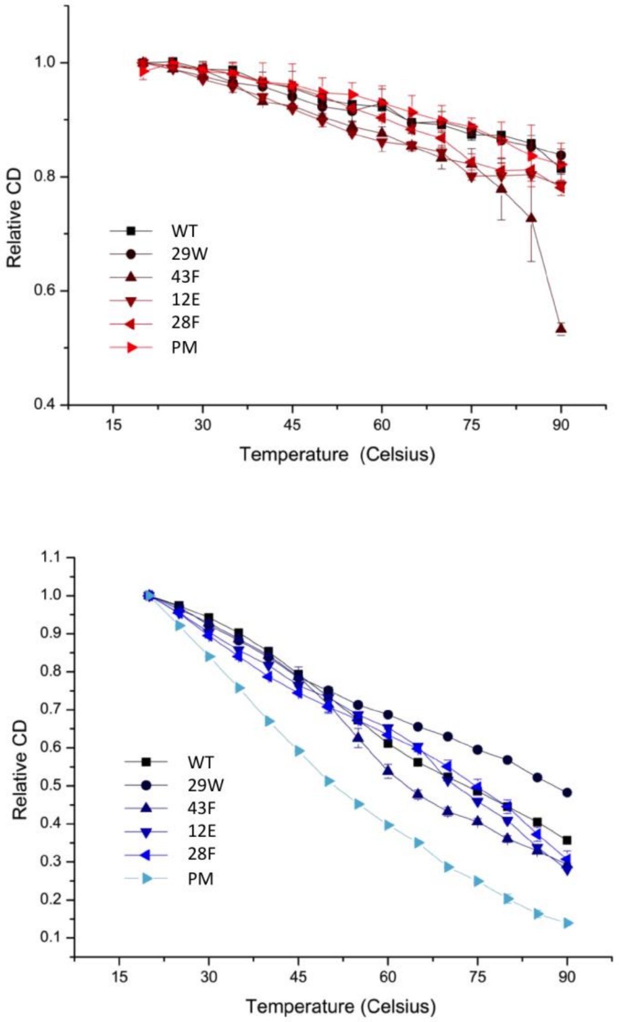


Figure 19: SRCD Thermal melt - Relative CD. A Relative CD at 200 nm for the wild-type, single mutants and PM in the absence of 0.1% SDS. B Relative CD at 191 nm for the wild-type, single mutants and PM in the presence of 0.1% SDS. From 20˚C to 90˚C in 5˚C increments the absolute CD at the aforementioned wavelengths was observed before normalisation by division by the largest CD signal observed at that wavelength, yielding relative CD signals between zero and one.

D29W compared with the WT suggests it has increased thermal stability and this is confirmed by the DichroWeb analysis showing helicity is retained at 44%.

The D29W mutant was predicted to decrease disorder content of the protein, so it might be possible that even though disorder was not reduced for this mutant in the absence of SDS some properties used by DISOPRED2 to estimate disorder might be linked to the increased stability and helicity observed when D29W was bound to SDS – perhaps the introduction of the highly hydrophobic Trp on the turn of the protein allows better hydrophobic packing with the tails of the SDS micelles. However, the K28F mutant was predicted to reduce disorder by a similar degree to the D29W mutant (Figure 14), but no increased stability compared to the WT was seen *in vitro* even though the location and properties of the amino acid introduced were similar – therefore it seems more likely that any relationship between disorder score and thermal stability/helicity of the D29W mutant was just coincidental.

The inability of DISOPRED2 to accurately predict the effect of single mutations on SHERP may have been due to a number of factors. First and foremost, DISOPRED2 was not designed to detect the small changes a single mutation may have on a protein sequence, so it might not be sensitive enough to accurately predict the effects. The utility of sequence based disorder predictors come from large scale proteomic studies where the inherent inaccuracies of the methods were mediated by the large number of predictions made (Iakoucheva et al. 2002; Ward et al. 2004a). Another issue could be DISOPRED2's homology approach. The program uses PSIBLAST (Altschul et al. 1997) to obtain a multiple sequence alignment, and this evolutionary information is used as an input for prediction of disorder along with a variety of other data about the sequence e.g. hydrophobicity. SHERP does not have any homologues – a BLASTP query of the non-redundant protein database finds only 3 matches with an *E*-value > 0.01 as of July 2017. The top two were SHERP proteins from other *Leishmania* species and the third is a match to a section of a histone protein. The lack of homologues meant that the sequence profile produced by PSIBLAST is not very information rich and couldn't contribute much to the analysis.

A method published recently in “Protein Engineering, Design and Selection” attempted to develop a computational approach to the problem of targeting specific alpha helicity profiles in intrinsically disordered proteins called “Genetic Algorithm for Design of Intrinsic secondary Structure” or GADIS (Harmon et al. 2016). A genetic algorithm generates sequences which were then simulated using Monte Carlo simulations with the ABSINTH implicit solvation model and force-field (Vitalis and Pappu 2009). These were then scored, a second simulation is run again before new parent sequences were selected and the cycle is repeated until the target helicity is reached. This method was demonstrated to be very successful at designing sequences with specific secondary structure while maintaining amino acid composition and important residues involved in binding. However it requires a large amount of computing power for just one prediction – for the 34 residue protein they use as an example, it takes 48 h to complete a single simulation on a quad core Nehalem processor, and the method requires many steps of two simulations with hundreds of sequence variants. Therefore, even though the sequence-based approach using DISOPRED2 does not give satisfactory results, the best high-throughput method available is impractical without access to large amounts of computational resources.

To investigate the WT’s structure in solution, a SAXS experiment in the same conditions to SRCD was performed with the protein in solution. Guinier analysis (Figure 20A) of the resulting I(q) plot produced an Rg value of ~26 Å and a maximum intraprotein distance of 100 Å – this suggests that SHERP in solution was in an extended and dynamic conformation. The Kratky plot of the same data (Figure 20B) supported this conclusion as the plateau seen was typical of an extended protein chain.

In order to further investigate the mutant’s effect on disorder, a 1 μs unfolding simulation of the WT was performed. These simulations started with the folded protein in solution in the hope that differences in secondary structure retention and other structural properties could be seen as the protein unfolded. The WT lost secondary structure over the course of the simulation, with ~15% helical content by 1 μs (Figure 21), in satisfactory agreement with the CD spectroscopy analysis.

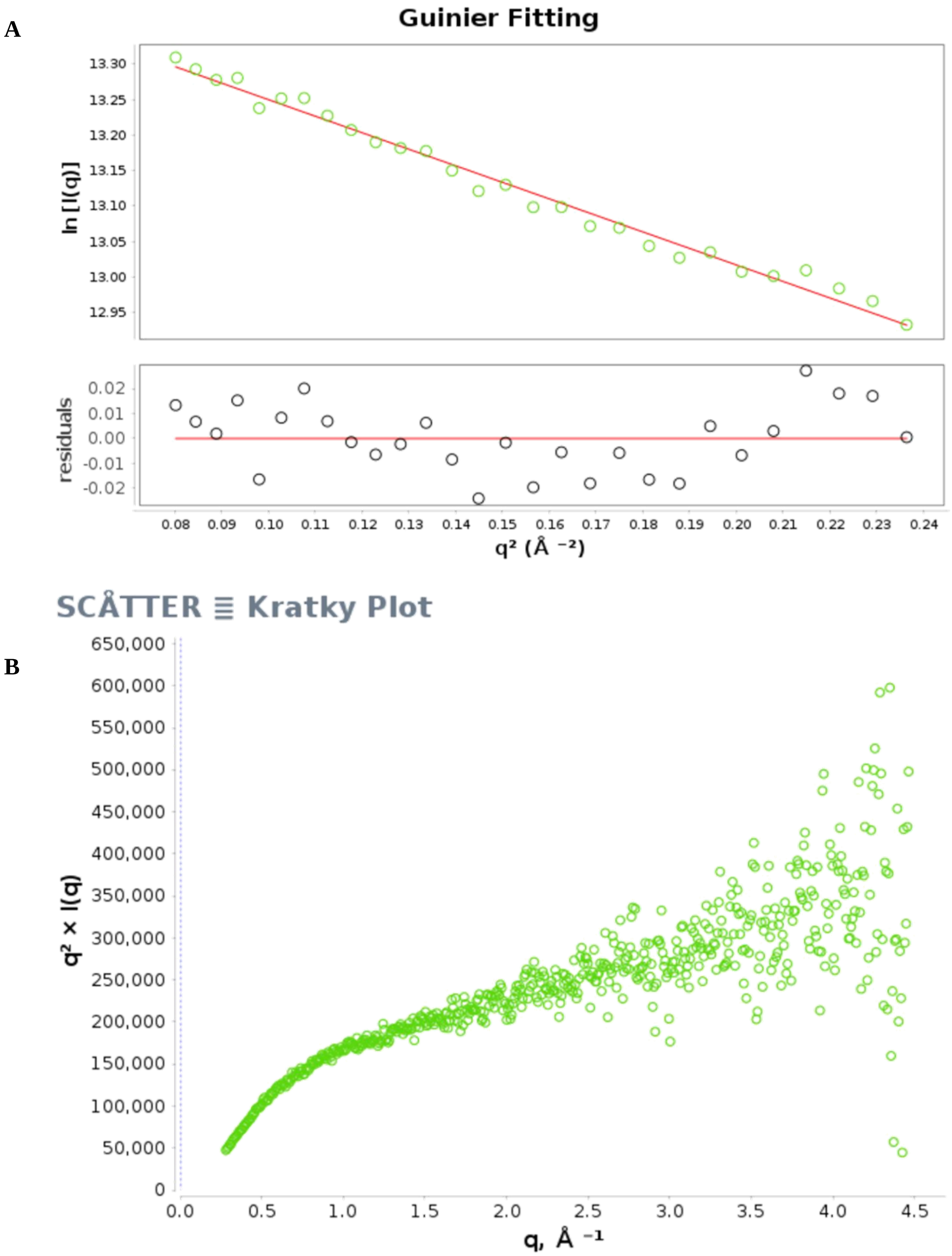


Figure 20: SAXS of WT in absence of SDS. A Guinier fitting of SAXS data. This analysis resulted in an estimated Rg of 26 Å. B Kratky representation of SAXS profile. Plots generated by SCATTER (Förster et al. 2010).

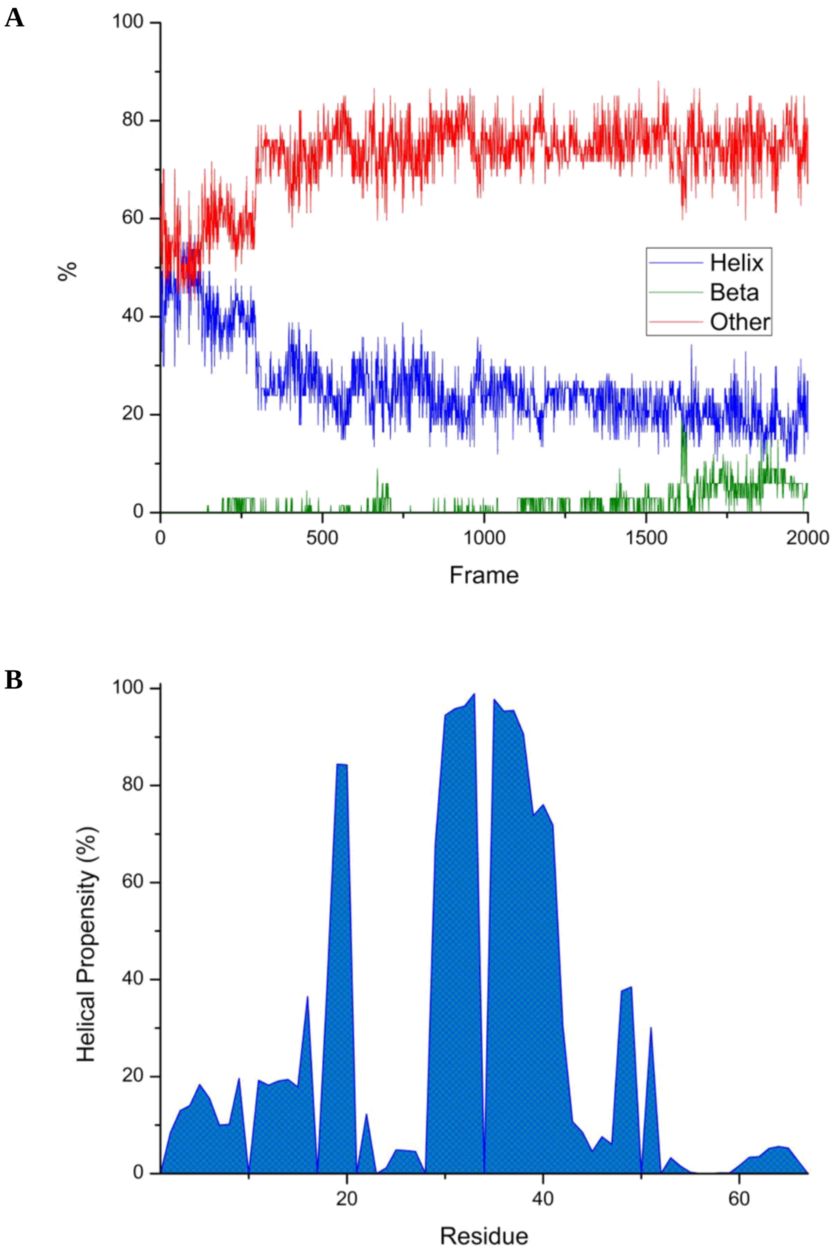


Figure 21: Secondary structure during wild-type SHERP unfolding simulation. ***A*** *Secondary structure composition per frame. Plot shows changes in helix (blue), beta sheet/strand (green) and "other" (red) during the simulation.* ***B*** *Plot of per residue helical propensity. Helical propensity is calculated as the percentage of frames the residue is defined as helical by STRIDE* (Heinig and Frishman 2004)*.*

Although for the WT the secondary structure analysis supports the results seen by SRCD, the Rg and SASA measurements (Figures 22A and 22B) contradict those found by SAXS. The protein became more compact over the course of the simulation, with Rg and SASA decreasing significantly. The SAXS results, on the other hand, show the protein is much less compact in the disordered solution state compared to the ordered state (Figure 20). This suggests that a conventional MD simulation was not a viable method for characterizing SHERP at the timescales that were accessible with the computing resources available, and starting from the ordered structure may have led to the protein conformation becoming stuck in a local energy minimum. An enhanced sampling method like REMD, starting from an unfolded protein chain, may have been more suitable for this system as it would have allowed more efficient and complete sampling of the conformational landscape. This was taken into consideration in subsequent projects, with REMD used to characterize the disordered state of the WT in Chapter 4.

3.2.5 The PM's effect on disorder

An unfolding simulation of the PM was performed under the same conditions as the WT. The PM showed greater helicity than the WT simulation, with an average of 41.3% and a final value of ~40% at 1 μs (Figure 23A). The per-residue helical propensity (Figure 23B) showed two regions of >90% helical character corresponding to the two helices present in the WT ordered structure, suggesting that *i→ i4* salt bridges were successfully added to the structure and were acting to stabilize the helices.. Indeed, a comparison of the salt bridges formed by the WT and PM show a large increase in both the number of unique possible salt bridges formed (33 for the WT vs 136 for the PM) and the average number of salt bridges per analyzed frame (Figures 24 to 25), with 1.95 salt bridges per frame for the WT and 8.51 per frame for the PM. Although it appears that the mutations did accomplish the goal of introducing salt bridges, and they did lead to increased stability of alpha helical structures, the simulation also displayed the same decreasing Rg and SASA seen in the WT (Figures 22A and 22B), suggesting it is not accurately reflecting the behavior of the protein *in vitro*.

The DichroWeb analysis of the PM SRCD data showed it had ~21% more helical character compared to the WT in solution at 20˚C (Table 6). The presence of extra *i→i4* salt bridges in the PM

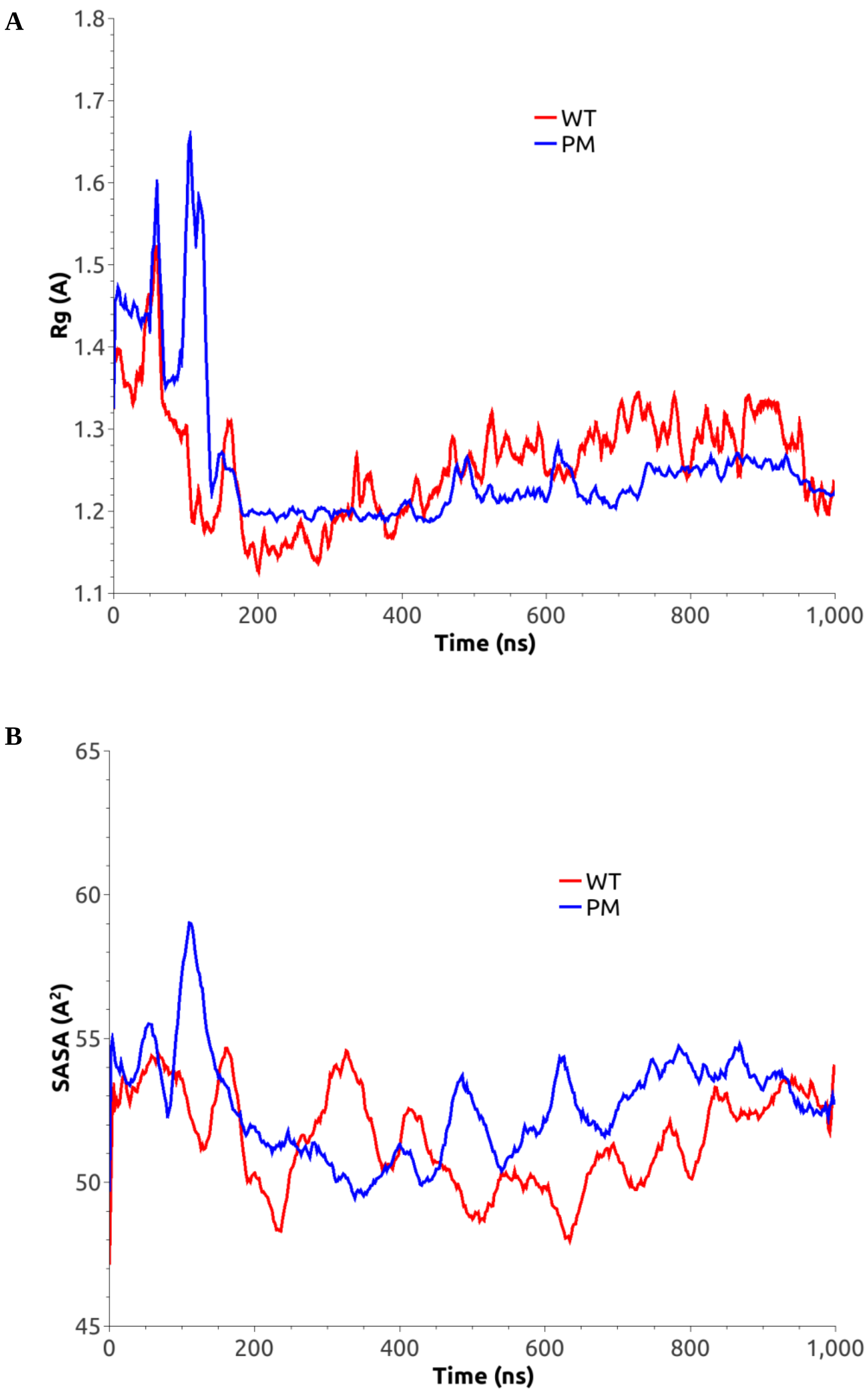


Figure 22: Rg and SASA changes during unfolding simulations. ***A*** *Rg over time. Plot shows changes in PM (blue)) and WT (red) during the simulation.* ***B*** *SASA over time. Plot shows changes in PM (blue) and WT (red) during the simulation. Both the WT and PM became more compact versus the NMR structure as the simulation progressed.*

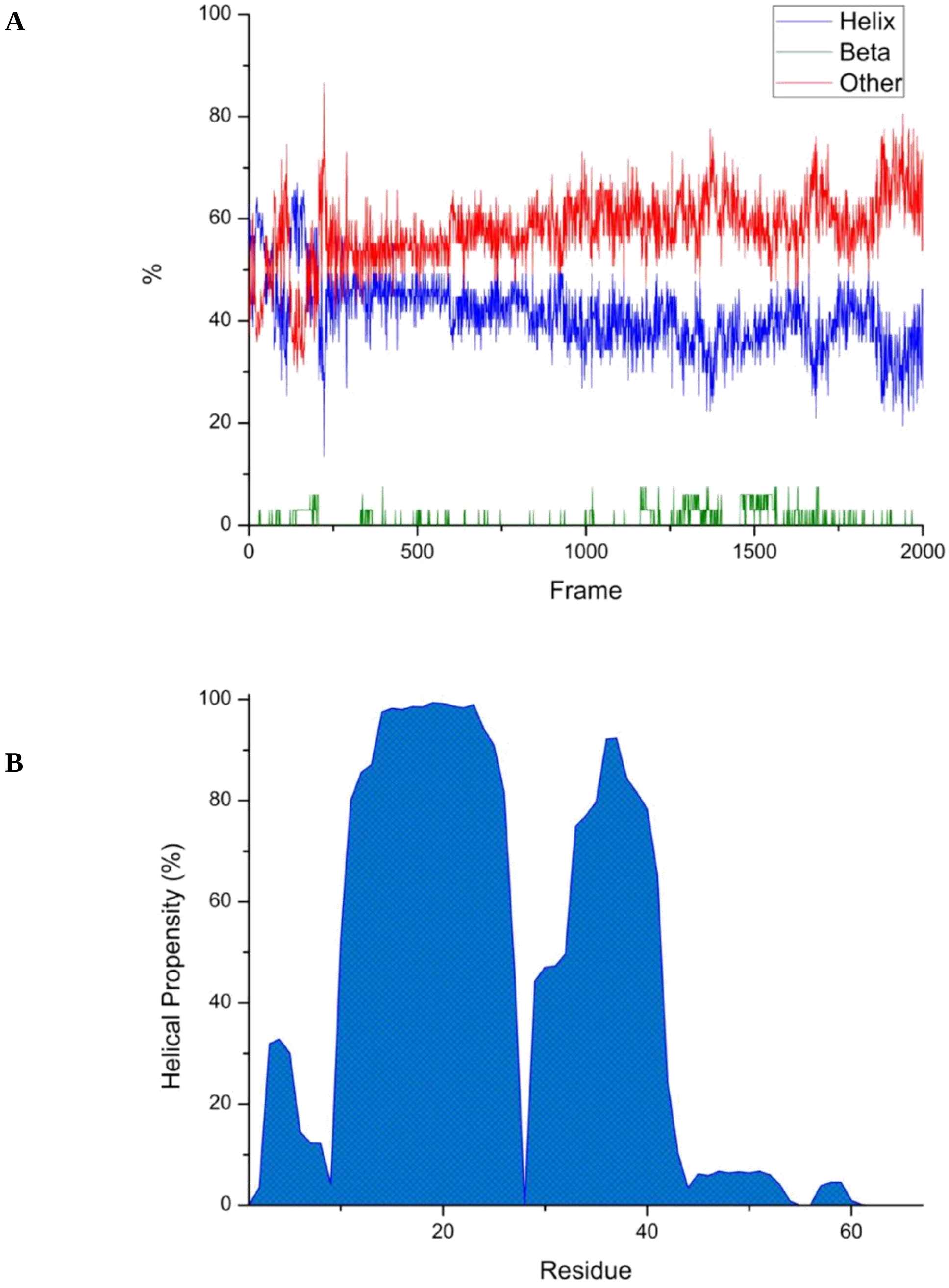


Figure 23: Secondary structure during PM SHERP unfolding simulation. ***A*** *Secondary structure composition per frame. Plot shows changes in helix (blue), beta sheet/strand (green) and "other" (red) during the simulation.* ***B*** *Plot of per residue helical propensity. Helical propensity is calculated as the percentage of frames the residue is defined as helical by STRIDE (Heinig and Frishman, 2004).*

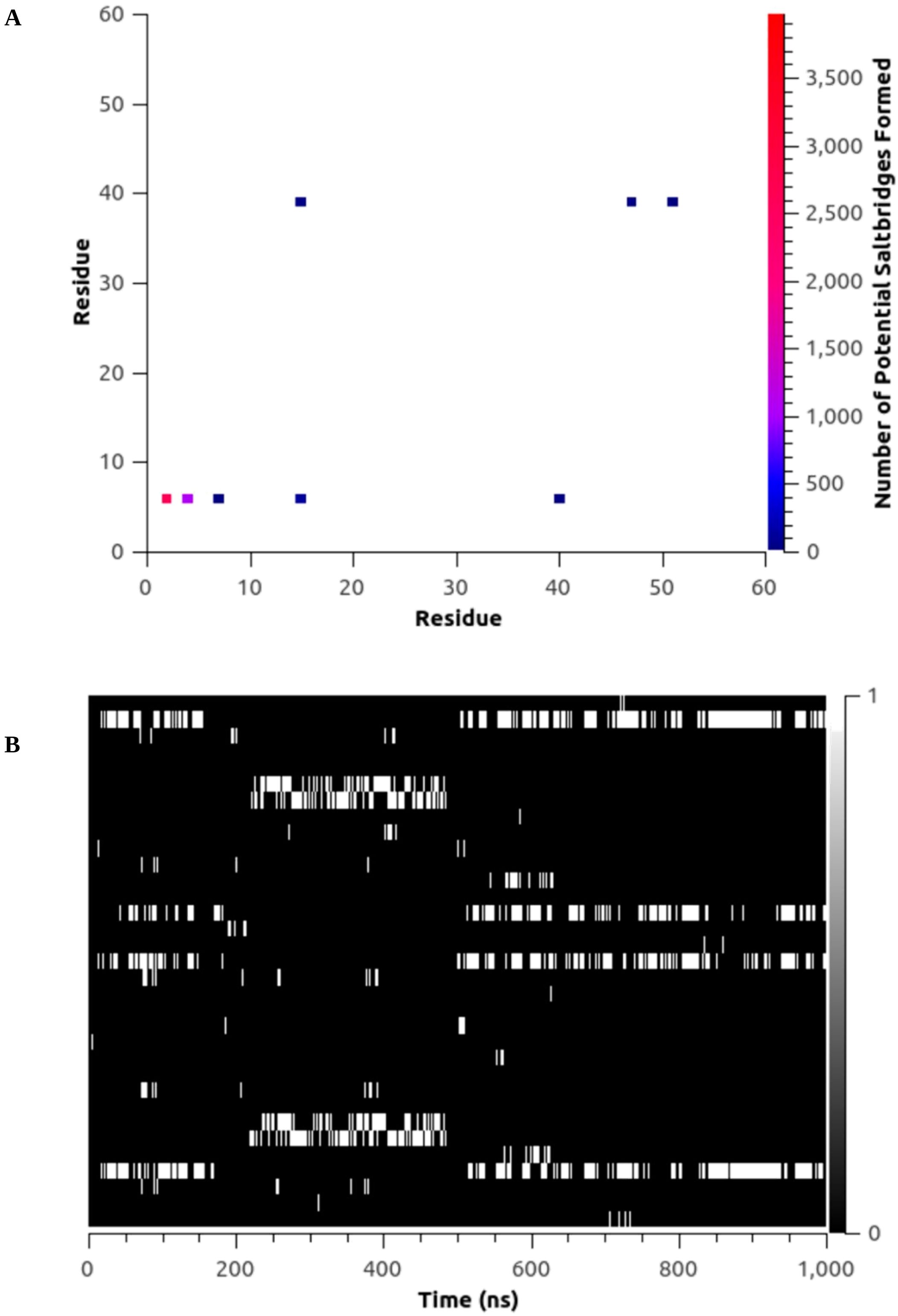


Figure 24: Salt bridge formation in the WT simulation. ***A*** *Heatmap of possible salt bridges formed as residue-residue pairs, red = high number of salt bridge, blue = low number of salt bridges.* ***B*** *Heatmap showing salt bridge formation over time, white = salt bridge, black = no salt bridge. Each "track" from left to right corresponds to a pair of residues that form a salt bridge at least once during the simulation, to illustrate salt bridge persistence over time.*

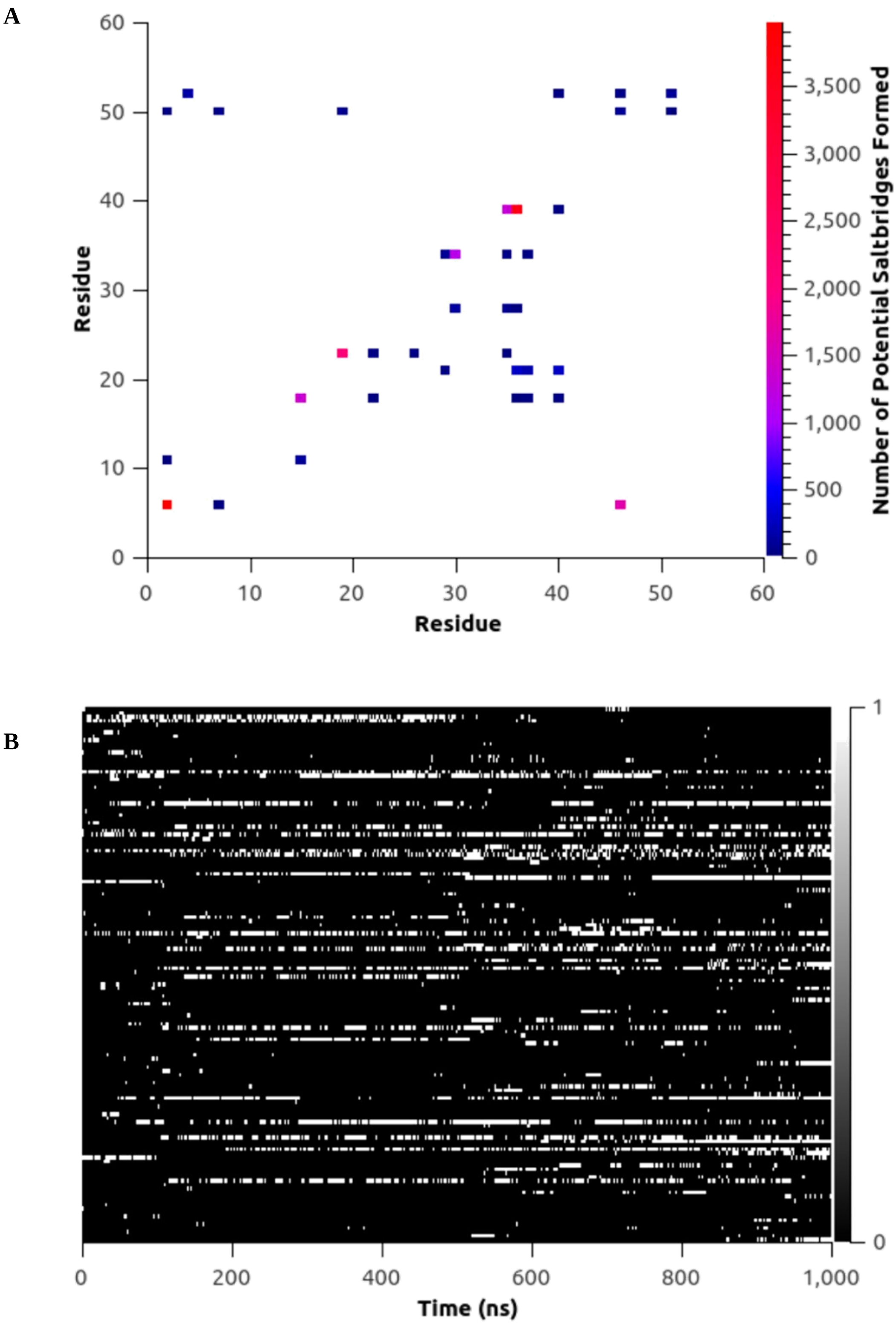


Figure 25: Salt bridge formation in the PM simulation. ***A*** *Heatmap of possible salt bridges formed as residue-residue pairs, red = high number of salt bridge, blue = low number of salt bridges.* ***B*** *Heatmap showing salt bridge formation over time, white = salt bridge, black = no salt bridge. Each "track" from left to right corresponds to a pair of residues that form a salt bridge at least once during the simulation, to illustrate salt bridge persistence over time.*

may have accounted for the increased secondary structure at 20˚C. In the presence of SDS the PM had slightly higher helicity compared to the WT (69% ± 3 vs 66% ± 2). However the PM had the lowest thermal stability of all the constructs tested, with an almost 90% drop in CD at 190 nm at 90˚C (Figure 19B). DichroWeb analysis showed the protein retained only 18% helicity compared to the WT's 34%, less than observed for the PM in the absence of SDS at 20˚C, although increases in sheet and turn meant the proportion of "other" did not increase as much compared to the WT. In the presence of SDS the additional salt bridges engineered into the sequence may not have been beneficial to the stability of the protein, as many of the residues altered were positively charged lysines and arginines that may play an important role in the interaction with the anionic head-groups of the detergent – this might have altered the proteins interaction with the SDS in a negative fashion and account for the decreased helicity seen at higher temperatures.

3.3 Conclusions

The disorder-to-order structural transition of WT SHERP was characterized by SRCD, with the protein undergoing a large increase in helicity in the presence of SDS. SAXS results indicated the protein is an extended chain in solution. Several mutant constructs identified by DISOPRED2 to have differing effects on disorder content were characterized. Increased secondary structure content was observed at 20˚C for the S43F mutant in the absence of SDS, which was predicted by DISOPRED2 to lead to a small decrease in disorder. However, the predicted decreases in disorder for the K28F and D29W mutants were not supported by the experimental data (although the D29W mutant showed a significant increase in helicity and thermal stability in the presence of SDS). The PM mutant showed increased helicity in the absence of 0.1% SDS at 20˚C, possibly due to the successful introduction of salt bridges into the structure - the PM simulation also showed greatly increased occurrence and occupancy of salt bridges, leading to increased helicity compared to the WT and retention of the helix-turn-helix motif seen in the WT ordered structure. However, although helicity values similar to those seen using CD spectroscopy were observed during the unfolding MD simulations for the WT and PM, they were not able to replicate the Rg values measured *in vitro* for the WT, becoming more compact versus the extended conformation seen by SAXS, suggesting that future work should employ enhanced sampling methods like REMD to better sample the proteins conformational space.

4.0 Chapter 4: The role of charge in the disorder to order transition of SHERP

4.1 Introduction

As previously discussed, SHERP undergoes a disorder to order structural transition in the presence of anionic lipids and detergents. However, the nature of the interaction between SHERP and its partner molecules is still unknown. Previous studies have shown that only anionic lipids/detergents induce secondary structure changes in SHERP – no transition was seen when DOPC/DOPE unilamellar vesicles were added to the protein, whereas large structural changes were seen with DOPC/DOPG or SDS (Moore et al. 2011). It can therefore by hypothesised that positively charged residues on the protein are responsible for the interaction and, due to the strong link between IDP's disorder-to-order transitions and their functions, essential for the proper functioning of the protein. Several questions exist regarding the structure of the SHERP-SDS complex. Does the protein bind to the surface of the micelle among the head-groups, in a similar manner to apolipoproteins (Wang et al. 1997; Lin and Oram 2000; MacRaild et al. 2004) or does it insert into the hydrophobic core of the micelle/membrane in a similar manner to membrane penetrating peptides (Eiríksdóttir et al. 2010)? Does it bind as a monomer, or is there a possibility that higher order oligomeric states are adopted as seen in many membrane binding peptides e.g. alpha synuclein (Dikiy and Eliezer 2012)? Figure 26 shows an electrostatic surface of folded SHERP in the presence of SDS. Positively charged regions (in blue) can be seen on the surface of the narrower faces of the protein (top-right and top–left) – it is likely these regions interact with the headgroups of SDS. A region of the protein surface with lower charge density can be seen in the bottom-right view – perhaps this part of the protein faces towards the micelle center, interacting with the hydrophobic core. Finally a highly negatively charged face can be seen in the bottom-left image – this could face away from the micelle , towards the solvent. The negative net charge of SHERP (-4) may allow it to bind without impacting the overall charge of the membrane or micelle, maintaining membrane charge asymmetry which is important for cellular function(McLaughlin and Murray 2005; Yeung et al. 2008).

In this chapter, the disordered ensemble of wild-type SHERP was characterised both *in vitro* and *in silico*, and simulations of a SHERP and SDS micelle system were performed. From the information gained from the simulations, cationic residues in SHERP that were observed to be important in the interaction of SHERP and SDS were substituted with alanine and the resulting mutants characterised by SRCD spectroscopy and SAXS.

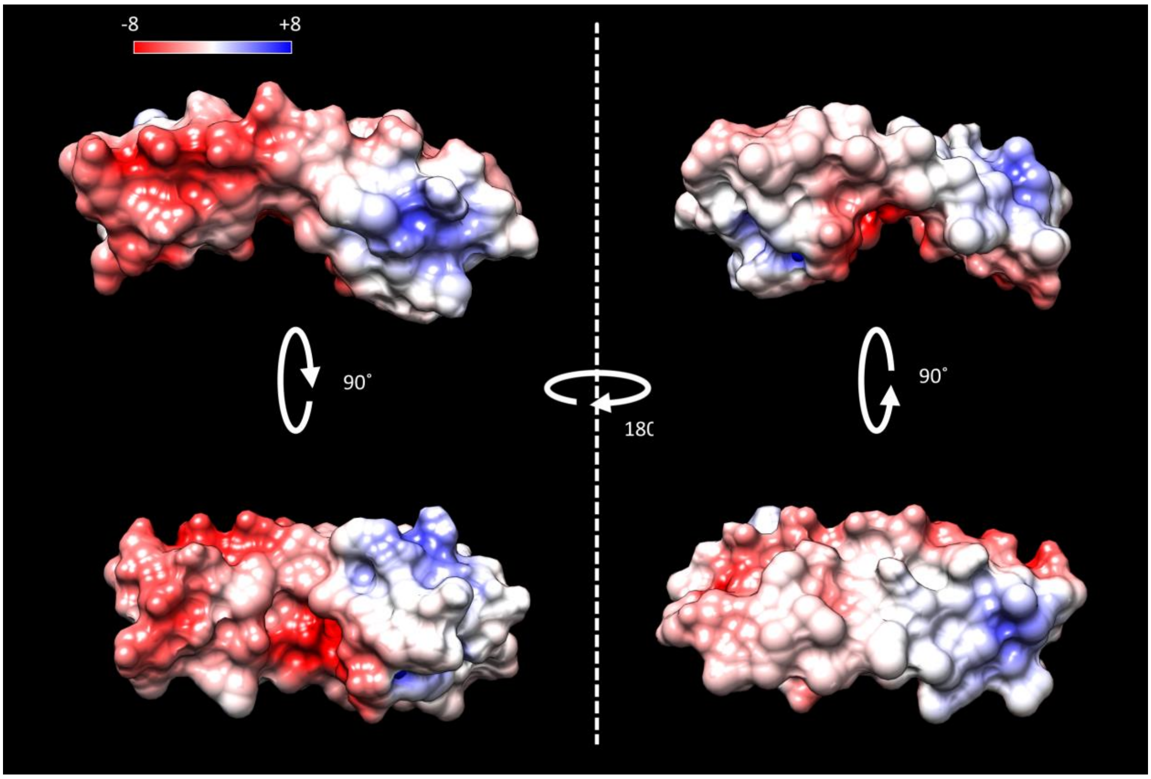


Figure 26: Electrostatic surface of folded SHERP in the presence of SDS. Pictured is the electrostatic surface of SHERP as calculated by APBS (Jurrus et al. 2018)*. Four angles of the structure are shown, and their relations to one another are indicated by the circular arrows and angles below. Positively charged regions are in blue, negatively charged regions are in red. Positively charged regions thought to be involved in interactions with SDS head-groups can be seen on the "flanks" of the protein in the top left and right images. The bottom right shows a plane of the protein with little charge – this is hypothesised to face into the centre of the micelle, making contact with the hydrophobic core. The bottom left shows a mostly negatively charged surface which would be exposed to the solvent in the hypothesised binding model. The image above was produced using Chimera* (Pettersen et al. 2004)*.*

4.2 Results

4.2.1 The disordered ensemble of wild-type SHERP in the absence of SDS

It was shown previously (8.2.4) that, in the absence of SDS, SHERP lacks ordered secondary structure, with CD spectroscopy yielding a helicity of ~8% and an "other" of 80% (Table 6). The preliminary SAXS data collected also showed that SHERP appeared to be extended in solution, with an Rg of ~26 Å and maximum distance of ~100 Å (Figure 20). IDPs cannot be described with a single structure - instead a large number of conformations are required to accurately describe the polydisperse sample.

Therefore, in order to gain more insight into the structural ensemble formed by SHERP in the absence of SDS, a molecular dynamics simulation of the unfolded protein was performed.

In section 3.2, unfolding simulations of SHERP and the PM were performed. Although these captured the differences in secondary structure content seen in the two samples, they did not accurately model the extended conformation seen by SAXS, instead forming a more compact, molten-globule arrangement. Therefore a new approach was needed to produce a more accurate ensemble for the protein in its disordered state. A commonly used technique for obtaining conformational ensembles of dynamic proteins is REMD (Sugita and Okamoto 1999). Multiple simulations are run simultaneously at different temperatures. At a given time interval, the potential energies of two temperature-adjacent replicas are evaluated and an exchange of the replica from one temperature to the other (and vice versa) is accepted or rejected based on a probability obtained from this. The chance of two replicas exchanging increases as the difference in temperature between two replicas decreases – the overlap between the frequency distribution of particles in a system over possible states (in the form of a Boltzmann distribution) increases. This allows REMD to efficiently sample both high and low energy conformations, which can greatly reduce computation time when compared with conventional MD.

Twenty replicas of 50 ns each were performed, for a total of 1 μs. As we were interested in the properties of the ensemble as a whole, these trajectories were combined and analysed together. To check that the distribution of energies sampled by each replica overlapped (to allow replica exchange to occur) the potential energy was calculated using the *gmx energy* tool of GROMACS and the results were plotted as a histogram (Figure 27). From the histogram it can be seen the distributions overlap confirming that the temperature range used was suitable for REMD. The REMD simulation was also repeated to check whether enough of the protein's conformational space had been sampled to fully describe the ensemble, and average ensemble properties and errors were obtained from the two trajectories.

First, the secondary structures of each frame were analysed and compared to the results obtained from CD spectroscopy. The helical propensity of the protein (calculated as the percentage of time a given residue was helical according to STRIDE (Heinig and Frishman 2004)) was calculated (Figure 28), and no part of the structure had a greater than 20% chance of being helical at any point in the simulation. The propensity plot was fairly flat, with no region appearing to have a particular preference for adopting a helical conformation. The average helicity of the ensemble was 9.4% ± 0.6, which is in excellent agreement with the SRCD spectroscopy results (8% ± 2, Table 6).

To investigate whether the extended nature of the protein was properly captured, the Rg for each frame was calculated. The range of Rgs found in the two simulations was large – from 10.4 Å to 52.2 Å – showing that SHERP is highly dynamic in solution, sampling a wide range of conformations. The average Rg value was 22.25 Å ± 0.06, in satisfactory agreement with our previous SAXS results (8.2.4). A free energy surface was produced using the Rg and RMSD of each frame of one of the simulation (Figure 29). It can clearly be seen that two main populations were observed – a sharper, narrow peak centred ~14 Å and a broader peak on ~26 Å. These two populations were also observed when the simulation was repeated (Figure 30).

To better see the differences between the two populations observed, structures from both REMD simulations were clustered by their Rg values using the k-means method with $k = 2$. For both runs,

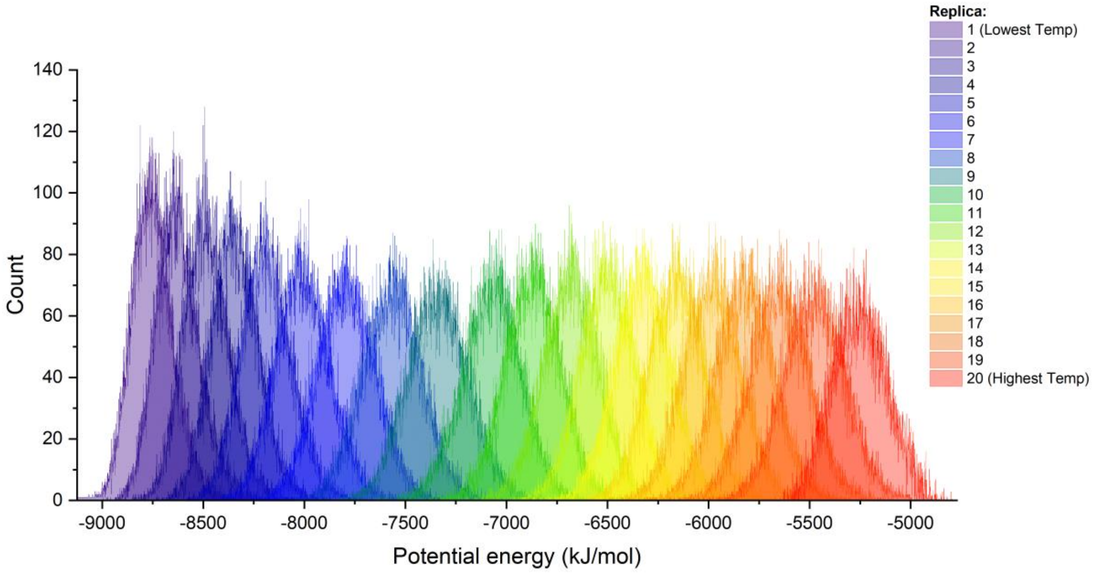


Figure 27: Histogram of potential energies for each REMD replica. Shown are the GROMACS calculated potential energies (kJ/mol) for the 20 replicas from the lowest temperature, replica 1 (in violet) to the highest temperature, replica 20 (red). The energy distributions of each replica overlap with adjacent replica's distributions, showing that the temperature intervals chosen were suitable for REMD simulation.

two distinct clusters were found corresponding to the two peaks seen in Figure 30. The clusters found in both simulations were very similar with small standard deviations in their average ensemble properties (Table 7), which again suggested that sufficient coverage of conformational space had been achieved to properly describe the disordered SHERP ensemble. The high Rg cluster (mean Rg = 28.4 Å ± 0.36) had a broad range of Rg values, and low helix and sheet content (4.6% ± 0.5 and 0.4% ± 0.09 respectively). This population was highly disordered with no residues with greater than 10% helical propensity over the course of the simulation. In contrast, the smaller Rg cluster (mean Rg of 14.9 Å ± 0.06) had a narrower distribution of Rg values and higher secondary structure content. Sheet content remained low (1.8% ± 0.29) but helical content increased by around 10% to 15.6% ± 0.09.

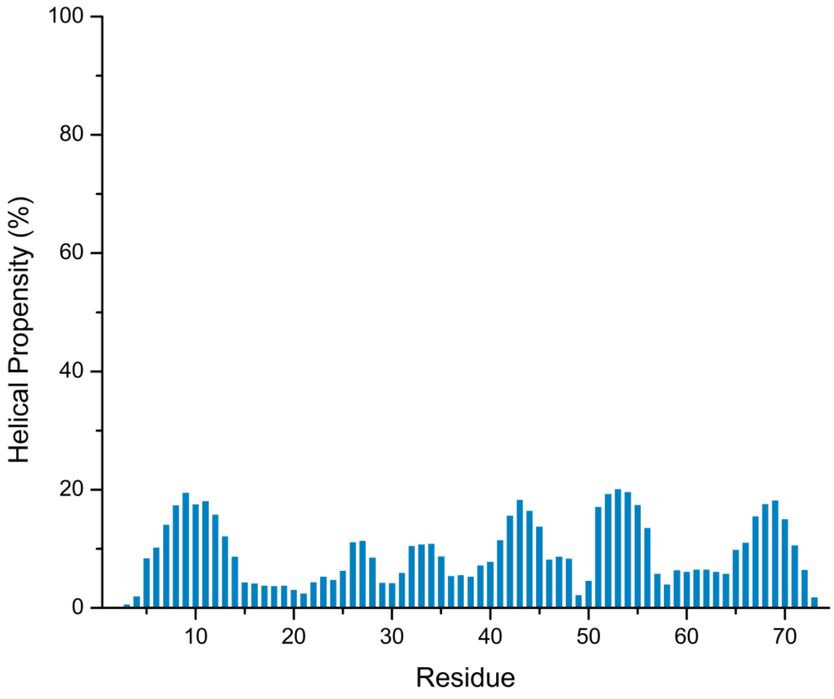


Figure 28: Per residue helical propensity of WT during REMD simulation. Helicity values calculated using STRIDE (Heinig and Frishman 2004)*. The protein has little helical character in the absence of SDS, with no residue helical >20% of the simulation duration.*

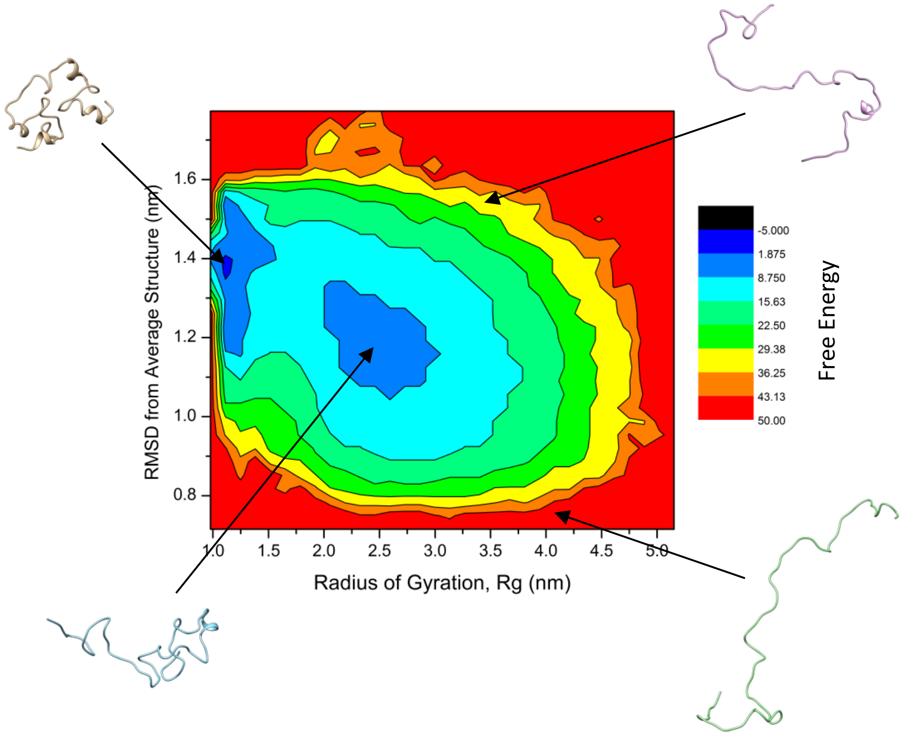


Figure 29: Free energy surface of SHERP in the absence of SDS. The free energy surface was calculated from the Rg and RMSD values observed during the REMD simulation. Blue indicates a highly populated region, red the least populated. Four structures representative of specific areas of the plot are shown. Structures visualised using Chimera (Pettersen et al. 2004). *FES calculated using the GROMACS utility g_sham* (Van der Spoel et al. 2005).

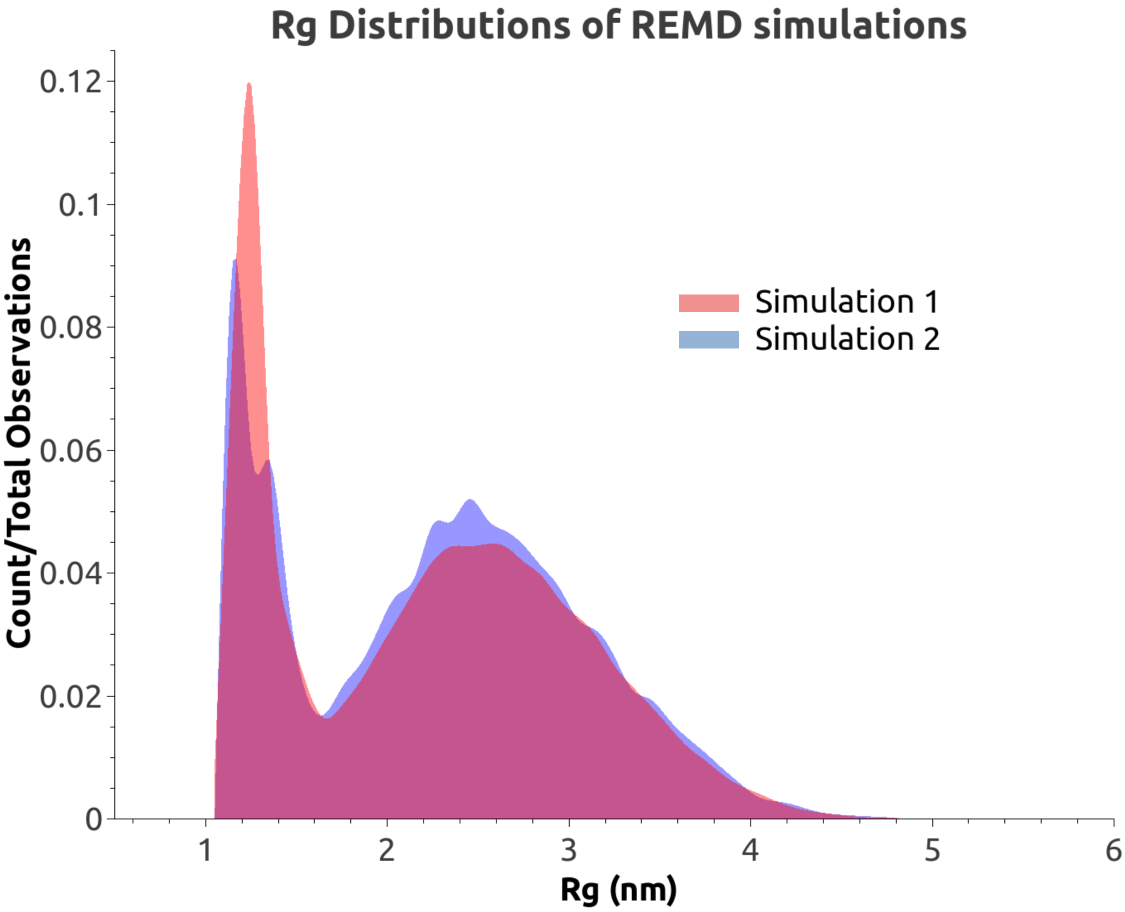


Figure 30: Histogram of Rgs for both REMD simulation runs. The two populations with a mean Rg of 14.9 Å ± 0.06 *and 28.4 Å ± 0.36 are present in both ensembles generated, suggesting the simulations had converged. Clustering of the ensembles by Rg revealed that these two populations had distinct secondary structure makeup (Table 7).*

Low Rg Cluster

	Mean		Minimum		Median		Maximum	
Rg (A)	14.9	± 0.06	10.7	± 0.20	13.4	± 0.08	21.6	± 0.15
Helix (%)	15.7	± 0.09	0.0	± 0.00	15.1	± 1.37	57.5	± 2.74
Sheet (%)	1.8	± 0.29	0.0	± 0.00	0.0	± 0.00	17.1	± 2.05
Other (%)	82.7	± 0.26	42.5	± 2.74	83.6	± 1.36	100.0	± 0.68

High Rg Cluster

	Mean		Minimum		Median		Maximum	
Rg (A)	28.4	± 0.36	21.6	± 0.15	27.5	± 0.38	50.2	± 0.35
Helix (%)	4.6	± 0.52	0.0	± 0.00	2.1	± 2.05	51.4	± 4.79
Sheet (%)	0.4	± 0.09	0.0	± 0.00	0.0	± 0.00	18.5	± 4.79
Other (%)	95.0	± 0.48	47.9	± 4.11	97.9	± 2.05	100.0	± 0.68

Table 7: Statistics from two clusters obtained from kmeans clustering of REMD simulations by Rg. The lower Rg cluster had a higher helical %, with some structures approaching the helical content and Rg of the WT in the presence of SDS.

The helical propensity of residues in this cluster was also higher, as might be expected, with residues 8 – 15 forming a patch with 25-30% helical propensity (Figure 31). Perhaps this smaller population with lower Rg values plays a role in the interaction of SHERP and SDS, providing transiently helical regions that could bind to the micelle and drive folding of the protein.

In order to make sure our previous preliminary SAXS results were correct, SAXS experiments were performed on the WT in the absence of SDS (Figure 32). To directly compare the REMD with the experimental data, 67 structures from the simulation (Figure 33), spanning the Rg values found, had their theoretical intensities back-calculated and fitted using a weighted linear combination method to the experimental profile (Figure 35; Table 8). This fitting suggested that the experimental I(q) could be represented by an ensemble consisting of ~72% unfolded structures (Rg > 22 Å), ~10% partially folded structures (12 < Rg < 22A) and finally ~18% compact, molten-globule like structures. This agreed well with the Kratky plot that showed the protein is extended and unfolded in solution (Figure 20), and the Rg was 28 Å when each REMD model was weighted by their relative contribution and averaged. The higher proportion of folded structures vs partially folded/unfolded structures also mirrored the two populations observed by REMD, and was also supported by the six highest contributing models (Figure 34) which have five “unfolded” and one “folded” members. We concluded that the values obtained from the REMD and SAXS analyses agreed with each other satisfactorily, as they were all within the resolution possible with SAXS.

4.2.2 The interaction mode of wild-type SHERP with SDS

The major aim of this chapter was to ascertain the mechanism by which SHERP binds to SDS micelles, and the organization of the resulting protein-micelle complex. From CD spectroscopy experiments, we knew that SHERP helicity increased to ~66% on addition of 0.1% SDS (Table 6). We also knew from the solution NMR solved ordered structure (Moore et al. 2011) that it adopts a helix-turn-helix conformation, with a helical content comparable to that obtained by CD (64%). However, whether it inserts into the hydrophobic core of the micelle, binds within the head-groups or adsorbed to the surface of the micelle was unknown, as was the number of SHERP that bound to each micelle and the structural changes induced in the micelle by binding.

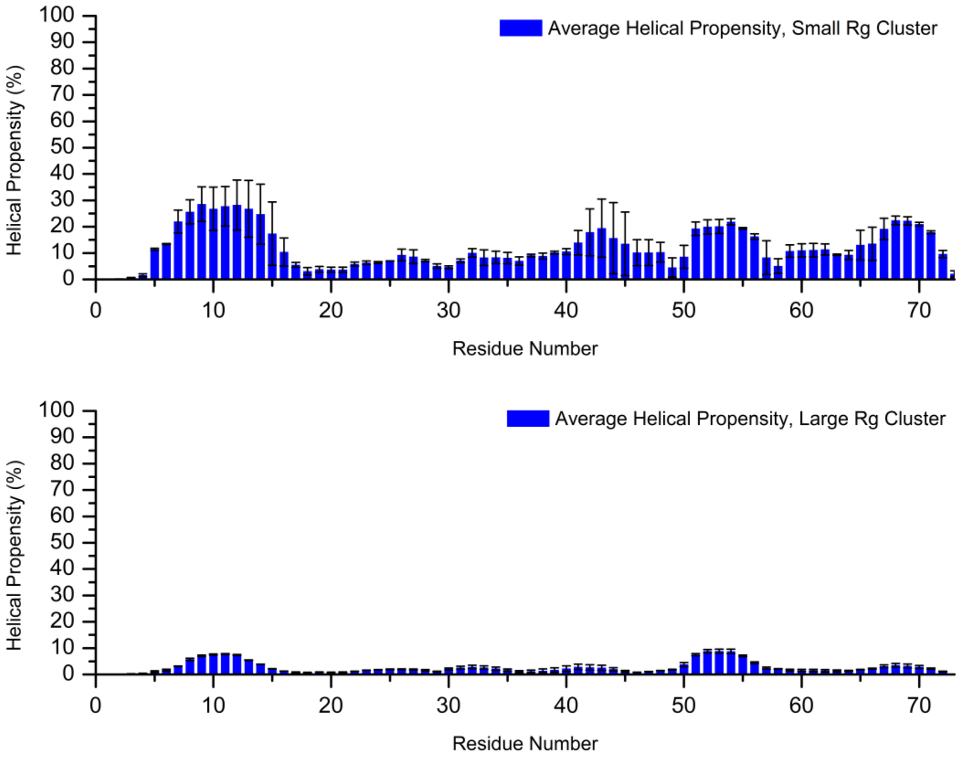


Figure 31: Per residue helical propensity of the clustered REMD simulation data. The top plot is for the low Rg cluster, the bottom plot for the high Rg cluster. The average helical propensity at each residue (obtained from the two simulations) is shown as blue columns, with error bars indicating the standard deviation. The small Rg cluster had a much higher propensity for helicity than the large Rg cluster – the formation of secondary structure elements will result in a more compact conformation so this was expected.

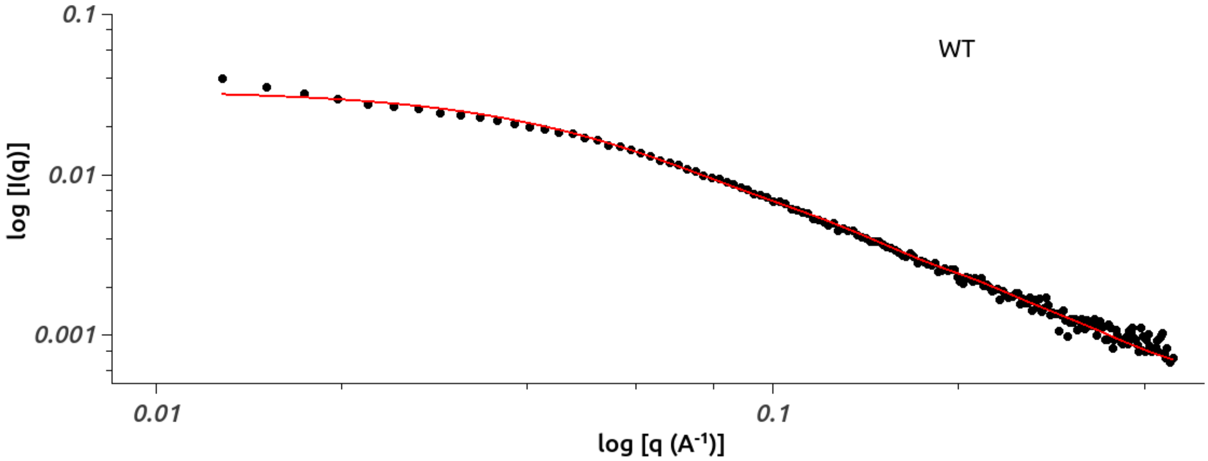


Figure 32: Weighted PDB fitting of WT in the absence of SDS. The experimental profile of WT in the absence of SDS is shown (black circles). In red is the profile resulting from fitting of 67 representative REMD structures to the experimental data. The 67 structures and their % contribution to the fitting can be seen in Figure 33 and Table 8.

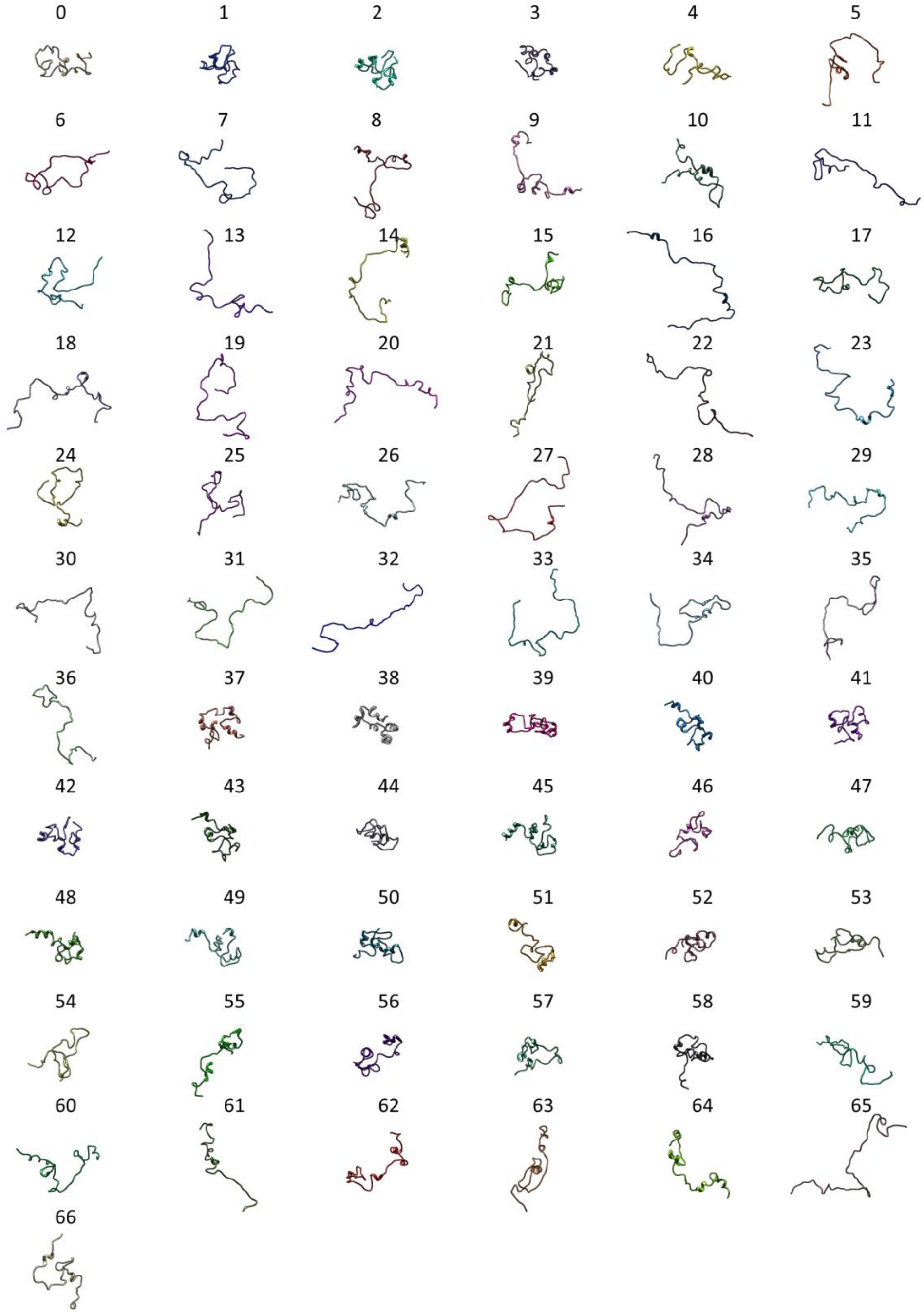


Figure 33: Structures used for fitting of SAXS data. Pictured above are the 67 PDB structures sampled from the WT REMD simulation and used for linear fitting of the WT and mutant SAXS profiles in the absence of SDS. Details on Rg and contribution to the fit of each samples profile can be found in Table 8 by looking up the number above the structure of interest. All structures were visualised using Chimera (Pettersen et al. 2004).

ID	Rg (Å)	WT(%)	2K (%)	3K (%)	3KR1 (%)	4K (%)
0	15.09	0.0 ± 0.2	0.0 ± 0.0	0.5 ± 1.3	0.0 ± 0.0	0.0 ± 0.0
1	13.42	0.0 ± 2.3	3.7 ± 1.7	3.4 ± 1.5	0.0 ± 0.0	0.0 ± 0.0
2	13.47	6.7 ± 1.2	6.3 ± 0.8	3.6 ± 1.3	0.0 ± 0.0	0.0 ± 0.0
3	13.59	0.0 ± 1.9	0.0 ± 2.0	0.0 ± 1.7	0.0 ± 3.0	5.8 ± 1.0
4	31.90	0.0 ± 0.1	0.0 ± 0.0	0.2 ± 1.1	0.0 ± 0.0	0.00 ± 0.0
5	23.31	0.0 ± 0.2	0.0 ± 0.0	0.0 ± 0.1	2.1 ± 1.2	2.4 ± 2.5
6	27.27	0.2 ± 0.1	0.0 ± 1.5	2.5 ± 1.0	0.0 ± 0.0	0.0 ± 0.0
7	24.13	4.0 ± 1.9	0.0 ± 1.5	0.0 ± 2.3	0.0 ± 0.0	0.0 ± 0.0
8	25.05	0.3 ± 0.8	5.2 ± 2.1	2.9 ± 1.8	6.2 ± 3.5	3.8 ± 1.5
9	23.58	0.0 ± 1.2	4.9 ± 2.2	0.2 ± 0.2	0.0 ± 0.0	0.0 ± 0.0
10	26.08	0.0 ± 0.1	0.0 ± 0.0	0.1 ± 0.2	6.4 ± 2.4	4.0 ± 1.2
11	29.22	0.1 ± 1.9	6.0 ± 2.4	3.9 ± 2.3	0.0 ± 0.0	0.0 ± 0.0
12	20.64	2.5 ± 2.7	0.0 ± 1.3	2.8 ± 1.7	9.6 ± 2.4	5.9 ± 1.0
13	29.53	6.4 ± 2.7	6.3 ± 2.9	3.6 ± 1.7	0.0 ± 0.0	5.8 ± 2.4
14	32.37	2.8 ± 1.9	6.1 ± 0.9	2.3 ± 1.3	0.0 ± 0.0	0.0 ± 0.0
15	25.97	0.0 ± 1.7	2.6 ± 2.1	0.1 ± 2.3	0.0 ± 0.0	0.0 ± 2.4
16	34.44	6.7 ± 1.2	6.1 ± 1.1	4.5 ± 0.9	0.0 ± 0.0	0.0 ± 0.0
17	21.84	0.0 ± 1.6	0.1 ± 0.0	4.0 ± 1.7	3.7 ± 2.9	2.9 ± 2.1
18	28.06	3.7 ± 1.8	0.1 ± 1.9	0.1 ± 1.5	0.0 ± 0.0	0.6 ± 2.5
19	24.85	0.0 ± 1.0	3.4 ± 1.4	0.0 ± 1.3	9.5 ± 2.2	5.9 ± 1.2
20	36.50	0.0 ± 0.2	0.0 ± 0.0	0.0 ± 0.1	0.0 ± 0.1	0.0 ± 1.4
21	25.49	0.2 ± 1.7	0.0 ± 1.9	1.2 ± 1.6	0.0 ± 2.0	0.0 ± 1.5
22	40.89	6.5 ± 1.2	6.2 ± 0.7	4.6 ± 0.6	0.0 ± 0.0	0.0 ± 0.0
23	29.60	6.6 ± 2.5	6.2 ± 2.5	3.5 ± 1.4	0.0 ± 0.1	0.0 ± 1.7
24	22.00	0.0 ± 2.4	0.0 ± 0.1	3.5 ± 1.6	9.2 ± 4.4	5.8 ± 1.3
25	23.18	0.0 ± 1.4	0.0 ± 0.0	3.7 ± 2.3	6.2 ± 2.8	3.3 ± 1.6
26	24.02	1.8 ± 1.2	0.0 ± 1.5	0.3 ± 1.3	0.0 ± 1.8	0.0 ± 1.2
27	27.64	4.6 ± 2.9	4.3 ± 2.0	0.2 ± 2.0	7.7 ± 4.1	4.7 ± 2.7
28	29.57	0.3 ± 2.7	0.3 ± 2.2	0.0 ± 1.4	0.0 ± 2.9	0.0 ± 2.0
29	23.37	5.2 ± 2.1	0.1 ± 1.4	1.3 ± 1.8	9.5 ± 2.5	5.9 ± 1.0
30	29.54	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.1	0.0 ± 0.2	1.4 ± 2.4
31	29.14	2.6 ± 1.8	0.2 ± 1.9	0.6 ± 1.2	7.6 ± 3.1	5.3 ± 1.4
32	47.40	6.7 ± 1.4	6.0 ± 0.8	4.6 ± 0.6	0.0 ± 0.0	0.0 ± 0.0
33	30.32	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
34	24.33	0.0 ± 1.8	0.0 ± 1.9	3.6 ± 1.2	7.4 ± 4.3	4.6 ± 1.3
35	28.42	4.1 ± 2.2	3.4 ± 3.0	0.8 ± 1.0	9.4 ± 2.3	5.5 ± 2.8
36	29.79	0.3 ± 2.2	1.6 ± 1.5	2.8 ± 1.8	0.0 ± 0.0	0.0 ± 2.3
37	13.99	0.0 ± 0.1	0.0 ± 1.2	1.3 ± 1.7	0.0 ± 0.0	0.0 ± 0.0
38	15.24	0.0 ± 1.6	0.0 ± 0.0	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
39	16.13	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.9	0.0 ± 0.0	0.0 ± 0.0
40	14.87	5.6 ± 0.5	6.2 ± 1.0	4.5 ± 0.7	0.0 ± 0.0	0.0 ± 0.0
41	13.09	0.0 ± 0.0	0.0 ± 0.0	2.6 ± 1.0	0.0 ± 0.1	5.1 ± 2.7
42	13.39	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.2	0.0 ± 0.2	5.9 ± 3.0

43	14.82	0.0 ± 0.0	0.0 ± 0.0	0.8 ± 0.3	0.0 ± 0.0	0.0 ± 0.0
44	13.28	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
45	14.53	0.0 ± 0.7	0.0 ± 0.0	0.2 ± 0.3	0.0 ± 0.0	0.0 ± 1.2
46	13.85	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 1.8	0.0 ± 0.0	0.0 ± 0.0
47	14.85	0.0 ± 0.3	0.0 ± 0.0	0.0 ± 0.2	0.0 ± 2.6	0.0 ± 0.0
48	15.15	3.6 ± 1.4	0.0 ± 0.0	0.5 ± 0.9	0.0 ± 0.0	0.0 ± 0.0
49	15.18	0.0 ± 1.6	0.0 ± 0.0	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
50	13.80	0.0 ± 2.3	0.0 ± 2.4	0.0 ± 1.2	0.0 ± 0.0	5.8 ± 2.3
51	15.55	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 1.5	0.0 ± 0.0	0.0 ± 0.0
52	13.71	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.4	0.0 ± 0.0	0.0 ± 0.0
53	16.60	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
54	16.66	0.0 ± 1.2	0.0 ± 0.0	0.1 ± 1.0	0.0 ± 0.0	0.0 ± 0.0
55	21.87	0.0 ± 0.2	0.0 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
56	19.52	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 1.6	0.0 ± 0.0	5.9 ± 2.4
57	17.82	0.0 ± 0.3	0.0 ± 0.0	4.1 ± 1.8	5.3 ± 3.4	3.4 ± 1.7
58	14.80	0.0 ± 0.1	0.0 ± 0.0	1.8 ± 1.3	0.0 ± 0.0	0.1 ± 0.0
59	26.31	0.0 ± 0.1	0.0 ± 1.7	1.4 ± 1.3	0.0 ± 0.0	0.0 ± 0.0
60	21.03	0.1 ± 1.7	0.8 ± 1.6	4.5 ± 1.8	0.0 ± 0.0	0.0 ± 0.0
61	29.03	3.7 ± 2.0	5.9 ± 2.1	3.3 ± 1.8	0.0 ± 0.0	0.0 ± 0.0
62	23.01	0.0 ± 1.8	0.0 ± 1.9	0.1 ± 1.4	0.0 ± 0.0	0.0 ± 0.0
63	21.54	0.0 ± 0.8	0.0 ± 1.2	2.3 ± 1.3	0.0 ± 0.0	0.0 ± 0.0
64	23.42	2.9 ± 1.3	1.6 ± 2.1	0.0 ± 1.5	0.0 ± 0.0	0.0 ± 0.0
65	35.68	6.4 ± 1.1	6.2 ± 1.0	4.4 ± 0.6	0.0 ± 0.0	0.0 ± 0.0
66	27.89	5.1 ± 2.4	0.0 ± 2.3	2.4 ± 0.9	0.0 ± 0.2	0.0 ± 1.7

Table 8: Contributions of REMD structures to linear fitting of SAXS data. The contribution as a percentage to the linear fitting of the SAXS profiles (and standard deviation from 10 repeat fittings) of the 67 structures taken from the WT REMD simulation are shown for the WT, 2K, 3K, 3KR1 and 4K mutants. The "ID" column refers to the ID given in Figure 33 which shows and image of each structure. Also given is the Rg for each structure calculated using GROMACS. The linear fitting was carried out using GENFIT and repeated 10 times before averaging.

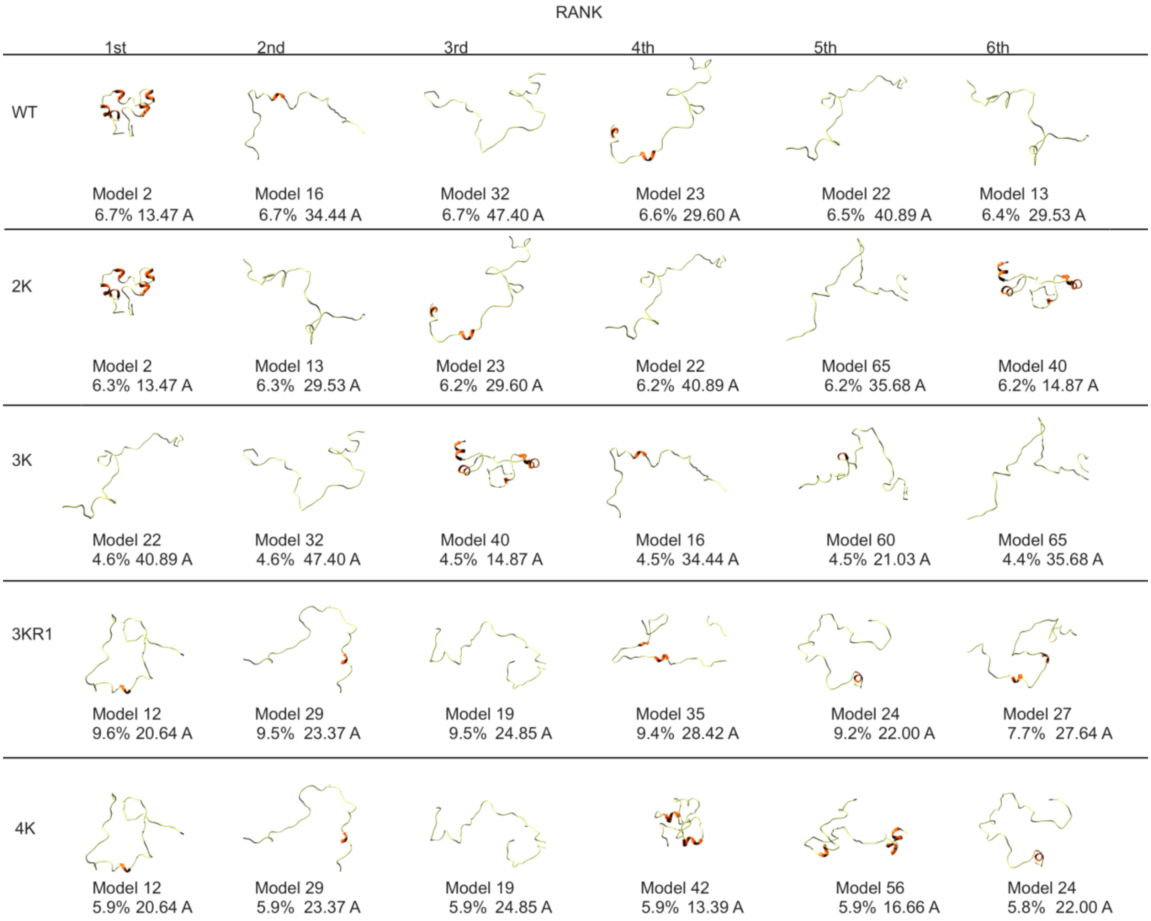


Figure 34: Representative structures from fitting of 67 REMD models to the WT and mutant I(q) profiles in the absence of SDS. The representative structures chosen are the six highest contributing models for each construct. Shown are the model names, their contribution as a percentage and their Rg. The WT's preference towards unfolded, extended structures is evident as five of the models have Rg > 29A. However, the low Rg cluster observed in the REMD is represented by the highest contributing structure (Model 2), which has a "folded" Rg of 13.47 Å. Comparing the wild-type to the mutants, the trend towards more compact ensembles as cationic residues were replaced with alanines can be seen. The comparatively high contribution of "partially unfolded" models to the 3KR1 profile is also represented in the structures above. All structures were visualised using Chimera (Pettersen et al. 2004).

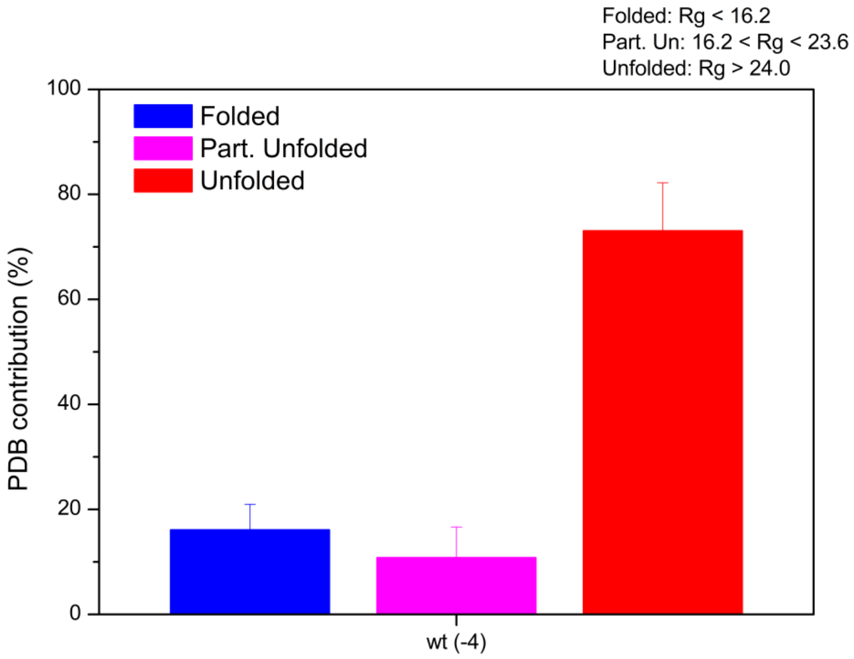


Figure 35: Radius of gyration distribution from PDB fitting of WT SAXS profile. The Rg values of each structure used in the calculation were weighted according to their proportional contribution to the fit, before binning into folded (Rg < 16.2 Å, blue), partially unfolded (16.2 Å < Rg < 23.6 Å, magenta) and unfolded (Rg > 24.0 Å, red). The WT is primarily unfolded in the absence of SDS – however, the smaller compact population observed in the REMD simulation is represented here by the contribution of "folded" structures being larger than that of "partially folded" ones.

As an initial first step to answering these questions, a 200 ns simulation of SHERP with an SDS micelle was performed. Due to the fact that there was only one SHERP molecule present in the simulation, the question of how many SHERP bind to each micelle could not be answered by this simulation. However, it would provide an idea of how SHERP might interact with the micelle, and highlight the role of specific residues in this interaction. The protein, although its structure differed slightly from the NMR solved ordered structure, appeared to be stabilized by the SDS micelle over the course of the simulation (Figure 36). RMSD, Rg and atom 33-819 distance had all appeared to reach equilibrium after approximately 135 ns. The RMSF showed that the majority of the protein fluctuated <0.5 nm during the simulation apart from the disordered C-terminal region and a short N-terminal region. The stability of SHERP in SDS was further supported by the retention of helicity seen in the secondary structure analysis of the protein (Figure 37). The per residue helical propensity showed the protein retained the helix-turn-helix motif found in the NMR structure (Figure 38), with the exception of a loss of helicity around residue 11. Helicity remained between 60-70% during the simulation – the final structure had approximately the same secondary structure content as the initial starting structure, and agreed very well with the CD spectroscopy results obtained previously (Section 3.2.4).

Throughout the entire simulation, the protein was located embedded within the head-groups of the micelle (Figure 39). Radial density function analysis of the simulation showed the clear overlap between the head-group peak and the protein peak (Figure 40). SHERP binding in this manner, rather than inserting into the paraffinic core of the micelle, would allow the greatest number of contacts between cationic side-chains and the anionic head-groups of the micelle, supporting our hypothesis that charge-charge interactions are key to the interaction. This also parallels the “snorkelling” binding mode of apolipoprotein C-II, mentioned in the introduction to this thesis (MacRaild et al. 2004).

One of the main hypotheses of this chapter was the contention that cationic residues played a key role in the interaction of SHERP with SDS. To assess whether that was the case in the MD simulation, the presence of salt bridges between ten cationic side-chains (Figure 41) and the anionic head-groups of the SDS were identified every 50 ps (Figure 42). Seven of the residues – R6, K10, K17, K23, K34, K50 and K52 – were involved in possible salt bridges > 80% of the simulation, the R6 and K17

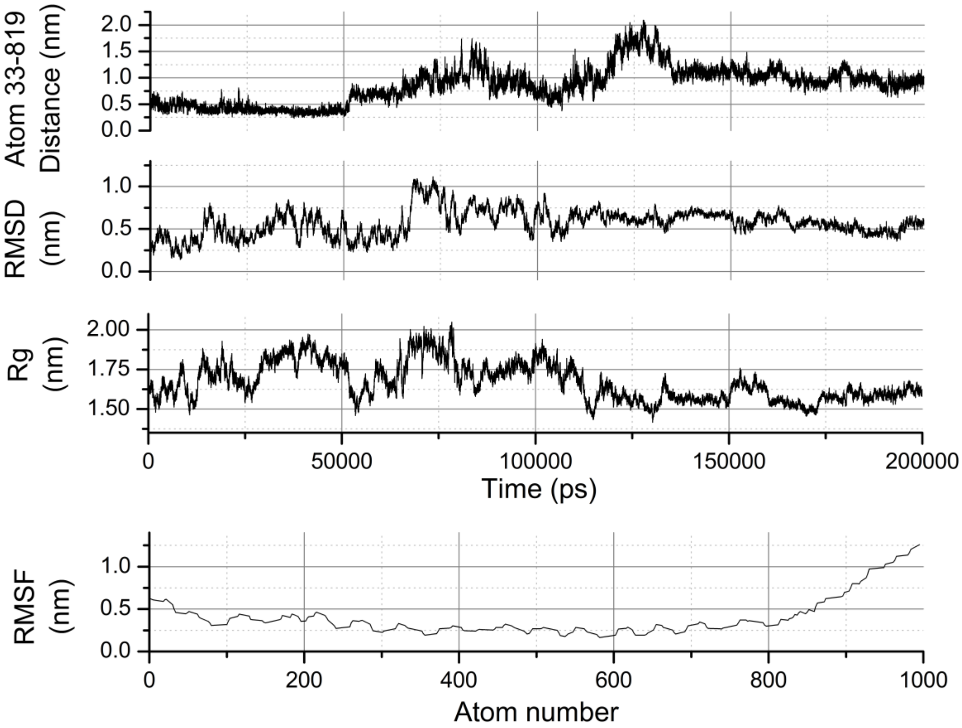


Figure 36: Structural properties of the protein during the MD simulation of the SHERP-SDS complex. Interatomic distance between atoms 33 and 819 (on the N- and C-terminal ends of the first and second helices respectively), RMSD of the backbone from the initial conformation of the protein, radius of gyration and root mean square fluctuation of the backbone atoms are shown. These results indicated that the protein had reached a stable equilibrium after approximately 135ns.

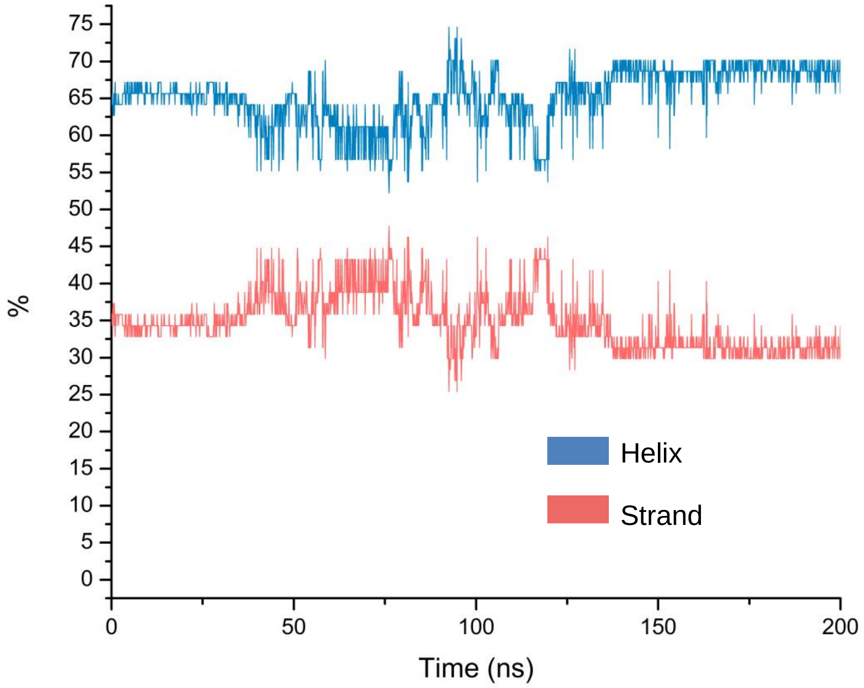


Figure 37: Secondary structure over time of the protein during the WT SHERP-SDS complex MD simulation. Secondary structure calculated using STRIDE (Heinig and Frishman 2004)*. The presence of SDS stabilised the secondary structure content of the protein, with the helicity at the end of the simulation comparable to that of ordered structure (~65%).*

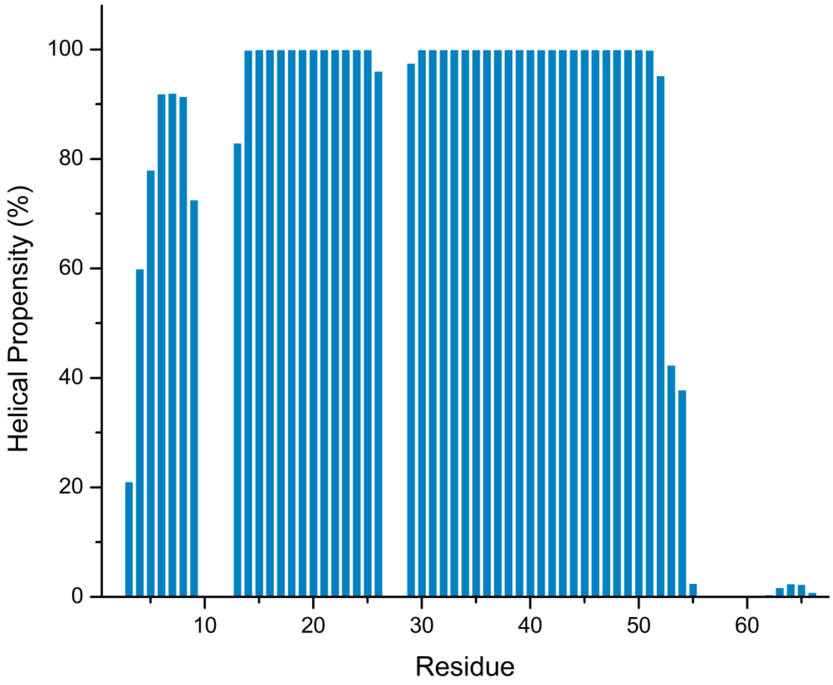


Figure 38: Per residue helical propensity of the WT during the SHERP-SDS complex MD simulation. Helical propensity calculate by STRIDE (Heinig and Frishman, 2004)*. The helix-turn-helix motif observed in the NMR structure of SHERP is retained during the simulation, with the exception of a kink that forms around residue 11.*

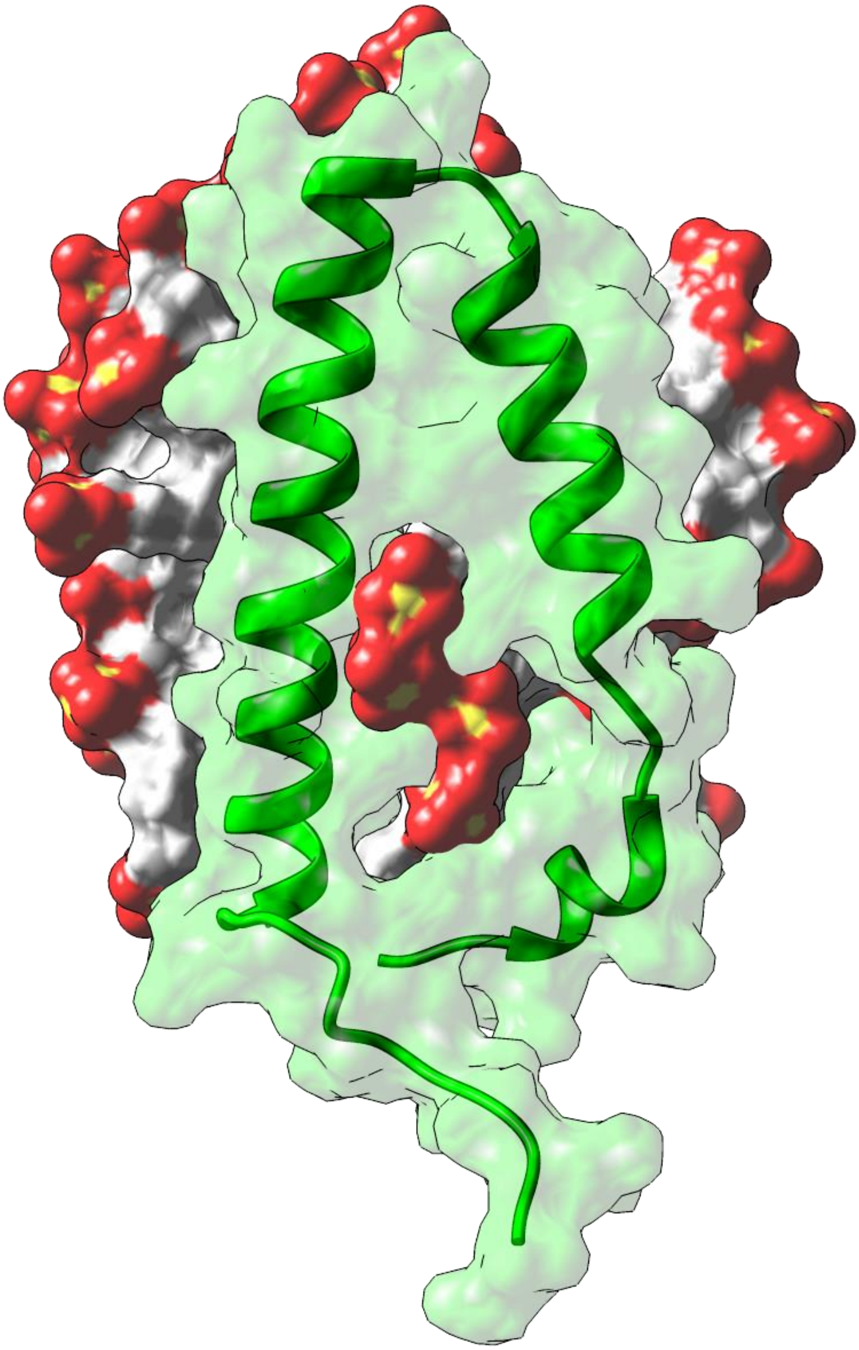


Figure 39: SHERP-SDS complex at 200ns. WT SHERP (green) shown in cartoon and surface representations. SDS molecules (red oxygens, yellow sulphur, white aliphatic chains) shown in surface representation. Structures visualised using Chimera (Pettersen et al. 2004)*. The helix-turn-helix motif is mostly retained from the starting structure. The protein is located amongst the charged sulphate headgroups of the micelle.*

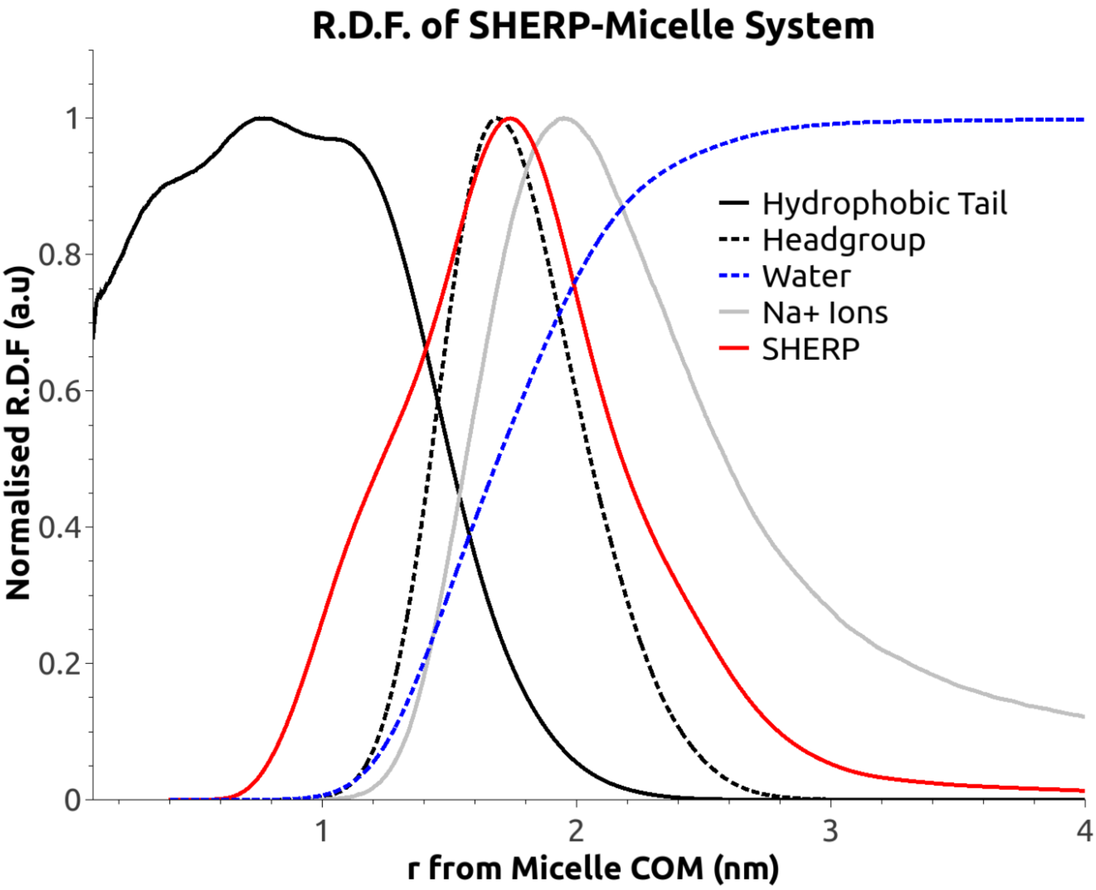


Figure 40: Radial density function (RDF) plot of SHERP-SDS complex simulation. The RDF of SHERP (red), SDS dodecyl tail (solid black), SDS head-group (dashed black), water molecules (dashed blue) and Na+ ions (grey) are shown. The RDF for each were normalised between 0 and 1 for easier visual comparison. It can clearly be seen that the protein (red) occupies the same region as the detergent headgroups (dashed black), with a smaller amount extending both into solution close to the micelle and into the hydrophobic core. No water is found in the center of the micelle.

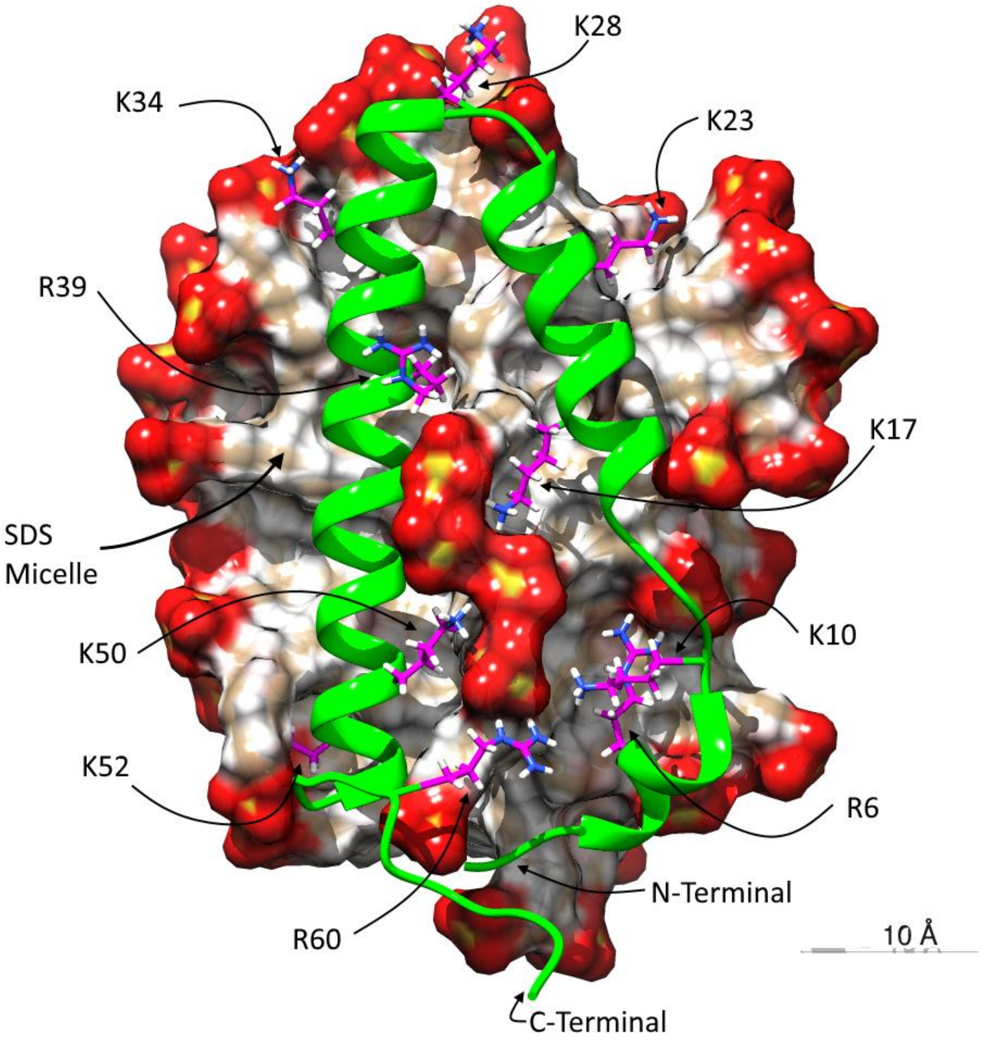


Figure 41: Locations of the ten cationic residues (magenta, stick representation) investigated for their role in salt bridge formation with the sulphate groups of SDS (red and yellow, surface representation) during the SHERP-SDS complex simulation. Structures visualised using Chimera (Pettersen et al. 2004).

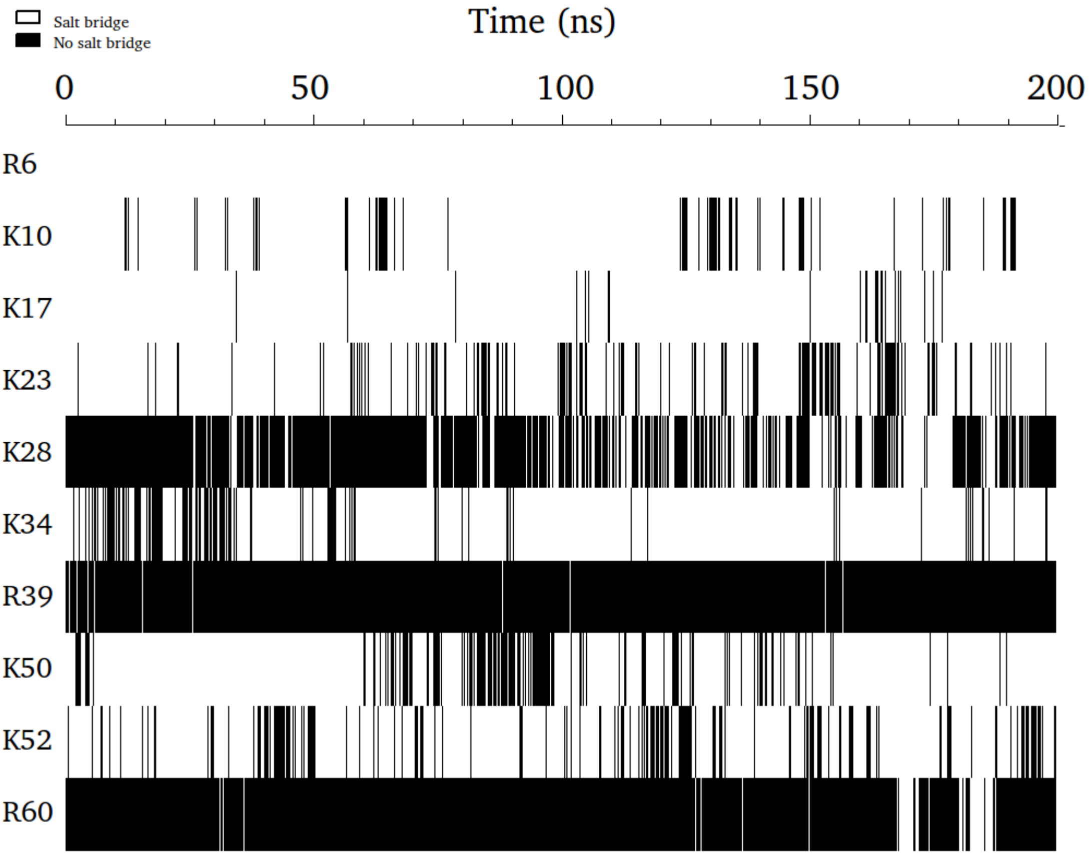


Figure 42: Salt bridge occupancy of the ten cationic residues during the 200 ns SHERP-SDS complex simulation. Presence of a salt bridge is indicated by white colouration, absence by black. Seven of the residues – R6, K10, K17, K23, K34, K50 and K52 – were involved in possible salt bridges > 80% of the simulation, the R6 and K17 residues with the highest salt bridge occupancy at 100% and 97% respectively (Table 9). K28 showed less persistent salt bridge formation at 29%, while R39 and R60 only formed salt bridges for <6% of the simulation.

residues (salt bridge network formed by K17 shown in Figure 43) with the highest salt bridge occupancy at 100% and 97% respectively (Table 9). K28 showed less persistent salt bridge formation at 29%, while R39 and R60 only formed salt bridges for <6% of the simulation. To check whether these salt bridges were resulting in persistent co-ordination with specific SDS molecules, the distance of all SDS molecules from each of the ten residues was found every 50 ps (Figure 44). A cut off of 10 Å was chosen to define a “close” SDS molecule, as the distance measured was between the S atom of the SDS molecule and the terminal C of the side-chains in question rather than the atoms involved in salt bridge formation (for the sake of simplicity). Multiple S atoms within 10 Å with small standard deviations in their distances could be seen, and in greater numbers for the residues with high salt bridge occupancy (points within below red lines, Figure 44). The small standard deviations showed that these SDS molecules did not move much during the simulation, which suggested the salt bridges were having an ordering effect on specific SDS molecules (Figure 45). Those SDS never found within 10 Å of any of the ten residues had much broader distributions of distances, indicating that they were more mobile. The presence of stable, persistent salt bridges between some of the cationic side-chains and specific SDS molecules marked them out as potentially key residues for the interaction of SHERP with the detergent, and it is possible their presence is vital for binding of SHERP to the micelle or for correct folding upon it.

To verify that SHERP bound within and to the head-group region of the micelle, SAXS experiments were carried out. The first step was to characterise SDS micelles in the absence of the protein. A spherocylindrical model (see Figure 9) was fitted to the I(q) plot obtained for the detergent micelles (Figure 46). This allowed us to obtain a number of parameters to describe the dimensions and organisation of the micelle to ascertain the effect of adding the protein (Table 10). The SDS micelle alone was found to have an aggregation number of ~106. The aggregation number is defined as the average number of surfactant monomers making up micelles after the critical micelle concentration is reached. It was 66.6 Å ± 2.3 along its longest axis, with a radius in the cylinder region of ~23 Å. This agrees well with results seen in the literature – SDS forms oblate ellipsoids/spherocylinders at

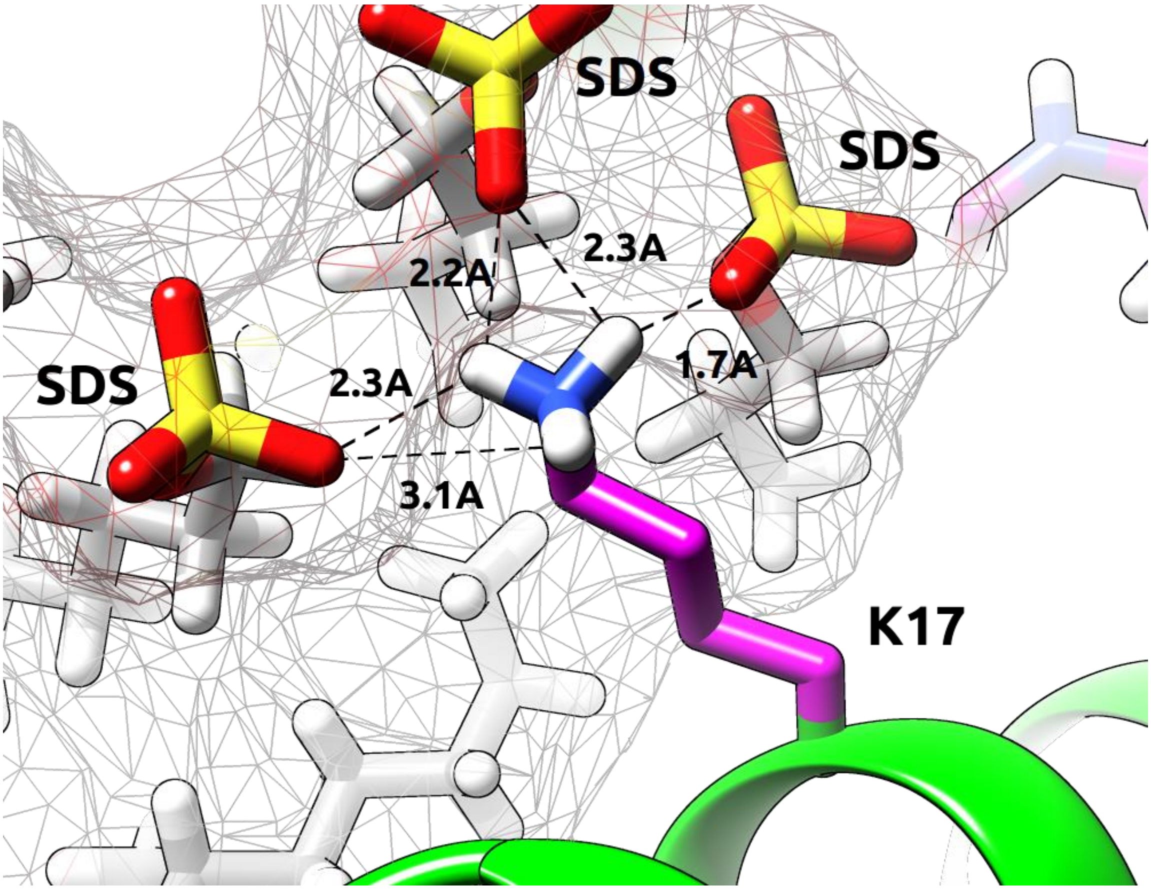


Figure 43: Close up of K17 (magenta) at 200 ns, showing the possible salt bridges it could form with the head-groups of SDS (red and yellow) at the final frame of the SHERP-SDS complex simulation. Distances from –NH groups of K17 to the oxygens of SDS are shown. Structures visualised using Chimera (Pettersen et al. 2004).

Residue	% of simulation involved in a potential salt bridge
R6	100.0
K10	92.1
K17	97.2
K23	83.2
K28	28.9
K34	87.8
R39	1.1
K50	83.7
K52	80.7
R60	5.5

Table 9: The percentage of the simulation each of the ten cationic residues seen in Figure 41 were involved in a potential salt bridge during the SHERP-SDS complex simulation.

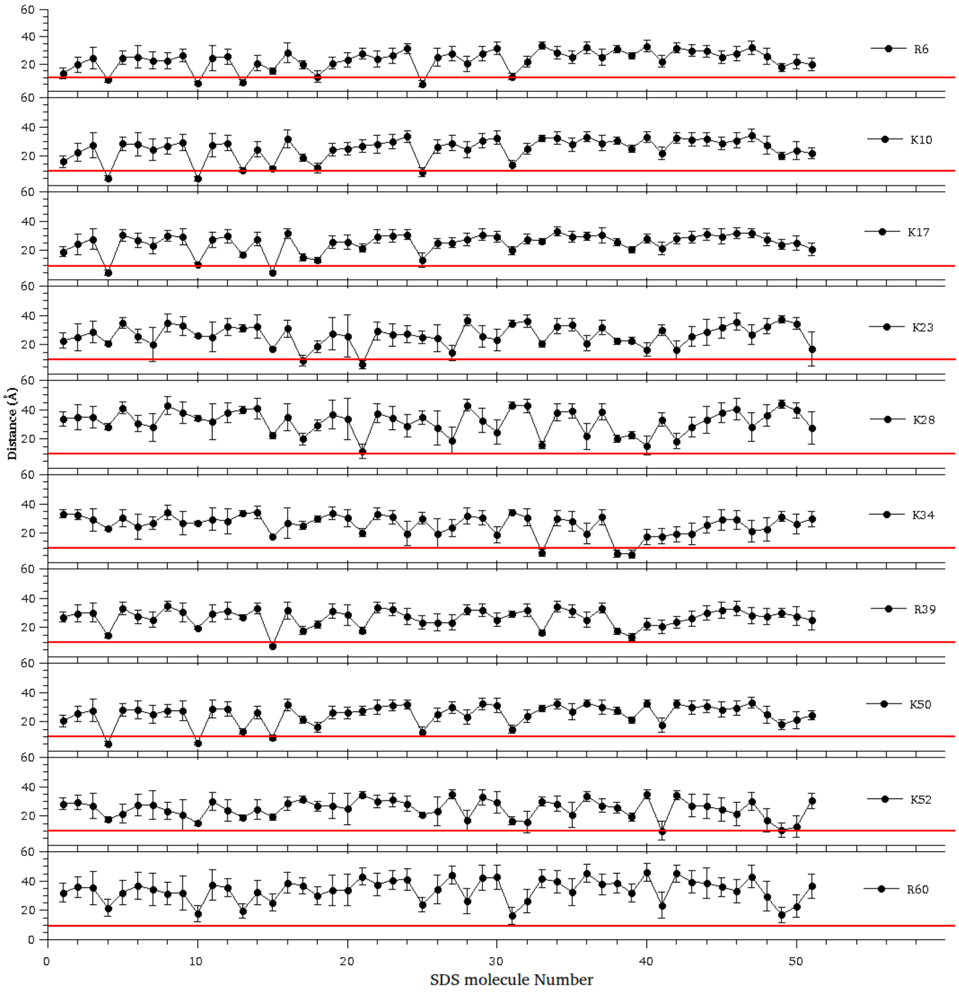


Figure 44: Average distance of each SDS molecule's head-group atoms from each of the ten cationic residues' (R6, K10, K17, K23, K28, K34, R39, K50, K52 and R60) acceptor groups. Standard deviation of the distances observed is shown with the error bars. A lower average distance and smaller standard deviation indicates the SDS molecule may have been stabilised by an interaction with that particular amino acid.

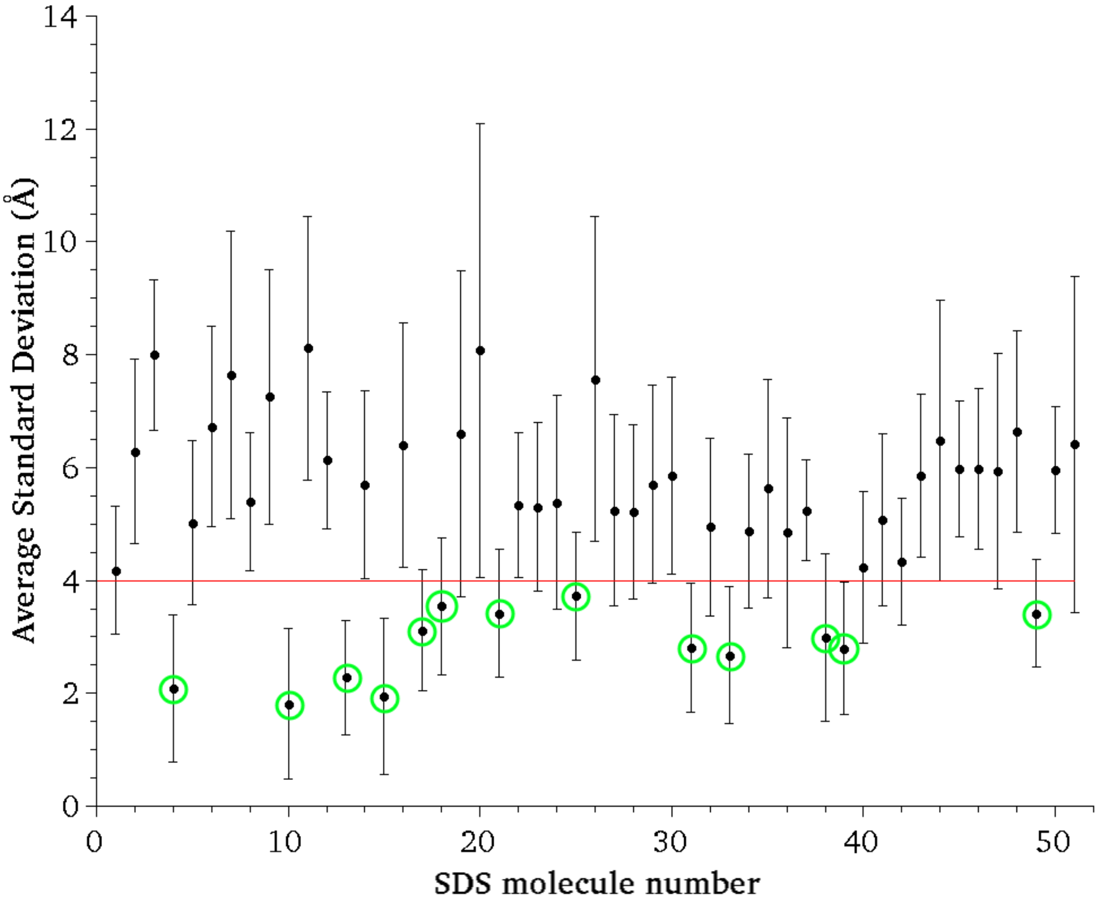


Figure 45: Mean and standard deviations of the standard deviations of SDS molecules from the ten cationic residues (R6, K10, K17, K23, K28, K34, R39, K50, K52 and R60). Circled in green are the SDS molecules with low standard deviations (<±4A) indicating they were stabilised by the formation of salt bridges with the protein.

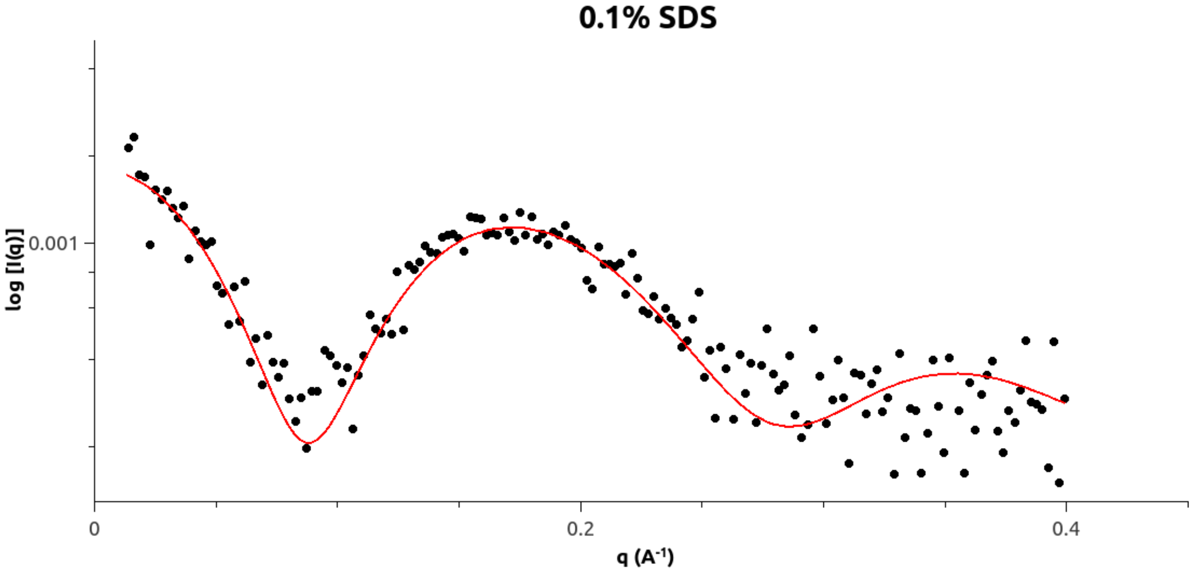


Figure 46: SAXS I(q) profile of 0.1% SDS alone. Experimental data is shown as black dots, with the spherocylinder model fit in red.

Sample	Micelle Length (A)	Agg No.	No. Protein bound	Free protein (%)	Free Micelle (%)
SDS alone	67 ± 2	106 ± 8.5	n/a	n/a	100.0 ± 0
WT + SDS	116 ± 3	88 ± 5	6.1 ± 0.5	35.4 ± 2	0.0 ± 0

Table 10: Parameters obtained for the SDS alone and WT-SDS systems from fitting of the two and three layered spherocylindrical models to the experimental SAXS data. Addition of the protein led to a large increase in the length of the spherocylinders – this occurred in spite of the aggregation number decreasing due to multiple SHERP molecules binding to the micelle headgroups in the cylinder region.

concentrations comparable to that used in this study (Duplâtre et al. 1996; Hammouda 2013; Tang et al. 2014). Examining the electron density profile for the micelle, the interface between the apolar core and polar shell could be seen at ~16 Å – this was roughly equivalent to the length of an SDS molecule (Figure 47).

On addition of protein, large changes were seen. Visual inspection of the I(q) plots for SDS micelles alone and SDS in the presence of the WT showed a clear shift in the characteristic micelle peak at q~=1.7 Å^{-1} to higher q values and an increase in intensity of the scattering (Figure 48). The summed profile of SHERP in the absence of SDS and SDS alone (orange) was significantly different to that of the WT in the presence of SDS – the protein was interacting with the micelle and large structural changes were occurring. The higher intensity of the WT-SDS profile compared to that of the protein alone or SDS alone suggested that this interaction led to an increase in the size of the micelle particle.

For the protein-micelle system, the spherocylinder model was extended to include another shell (representing the exterior solute of the micelle), bound protein, free protein in solution and free micelles. The biggest changes were in the length of the micelle – the addition of protein led to significant elongation along the micelles cylindrical axis to 116 Å ± 4.9 (Table 9). This was similar to the changes seen in pure SDS systems when the concentration of detergent increases – a more rod-like structure forms (Duplâtre et al. 1996; Hammouda 2013; Tang et al. 2014). However, the aggregation number of the micelle decreased to ~89 from 106 (Table 9). Taking the magnitude of the length increase with the decrease in aggregation number, multiple SHERP molecules must have bound to each micelle. That was indeed the case, with our model predicting approximately 6 WT bound per micelle (Table 9). The protein only bound to the cylindrical portion of the micelle – the protein preferred the “flatter” binding surface in this region to the high curvature of the hemisphere region. The protein was also predominantly located in the polar head-group region of the micelle (92%), with insignificant presence in the paraffinic core and a small presence in the exterior of the micelle (8%).

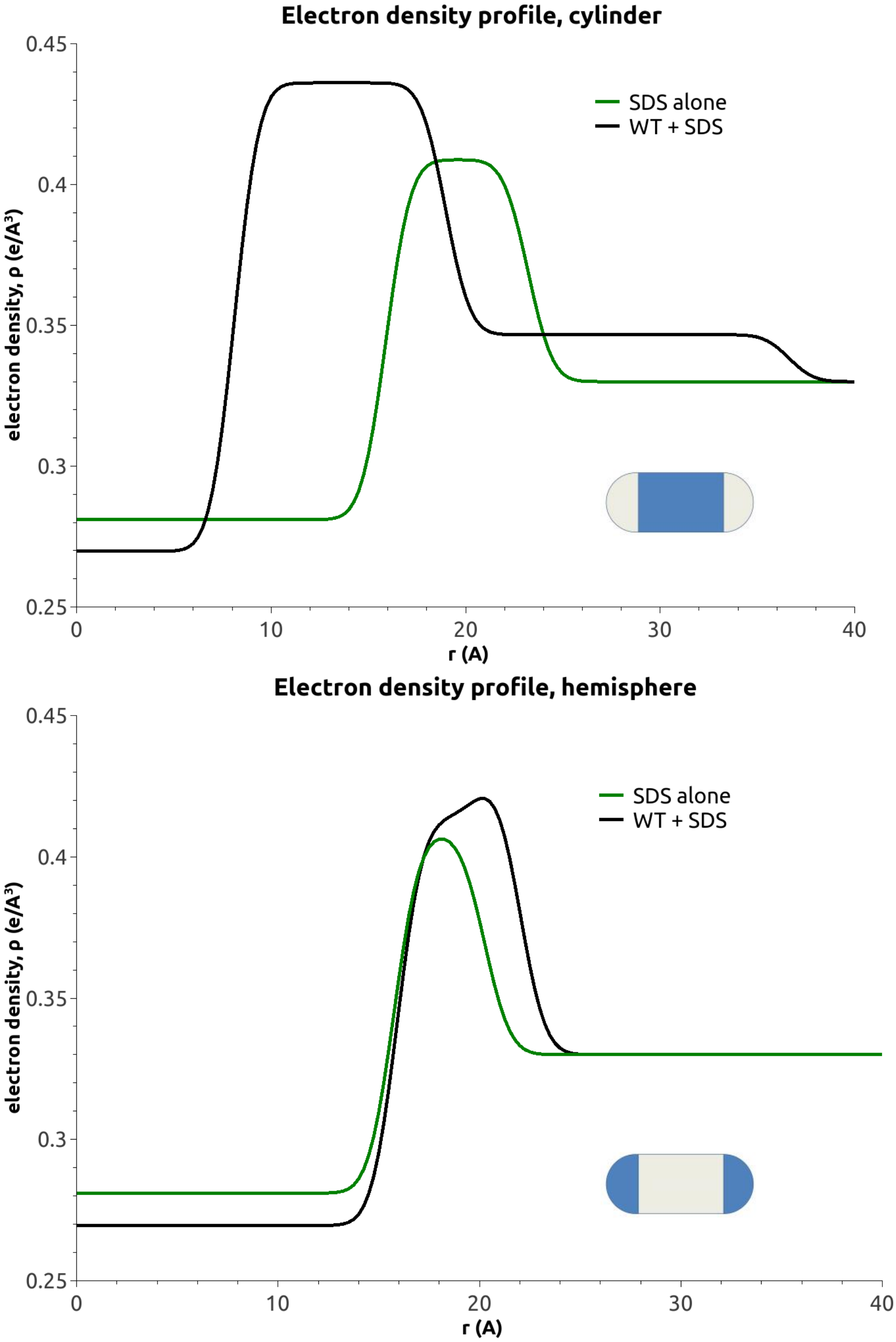


Figure 47: Electron density profiles for the SDS alone (green) and WT-SDS (black) systems. Electron density was obtained from modelling of the SAXS I(q) data with GENFIT. Electron density is shown for the cylindrical region (top, indicated in blue on inset diagram) and hemispherical region (bottom, indicated in blue on inset diagram) of the micelle model.

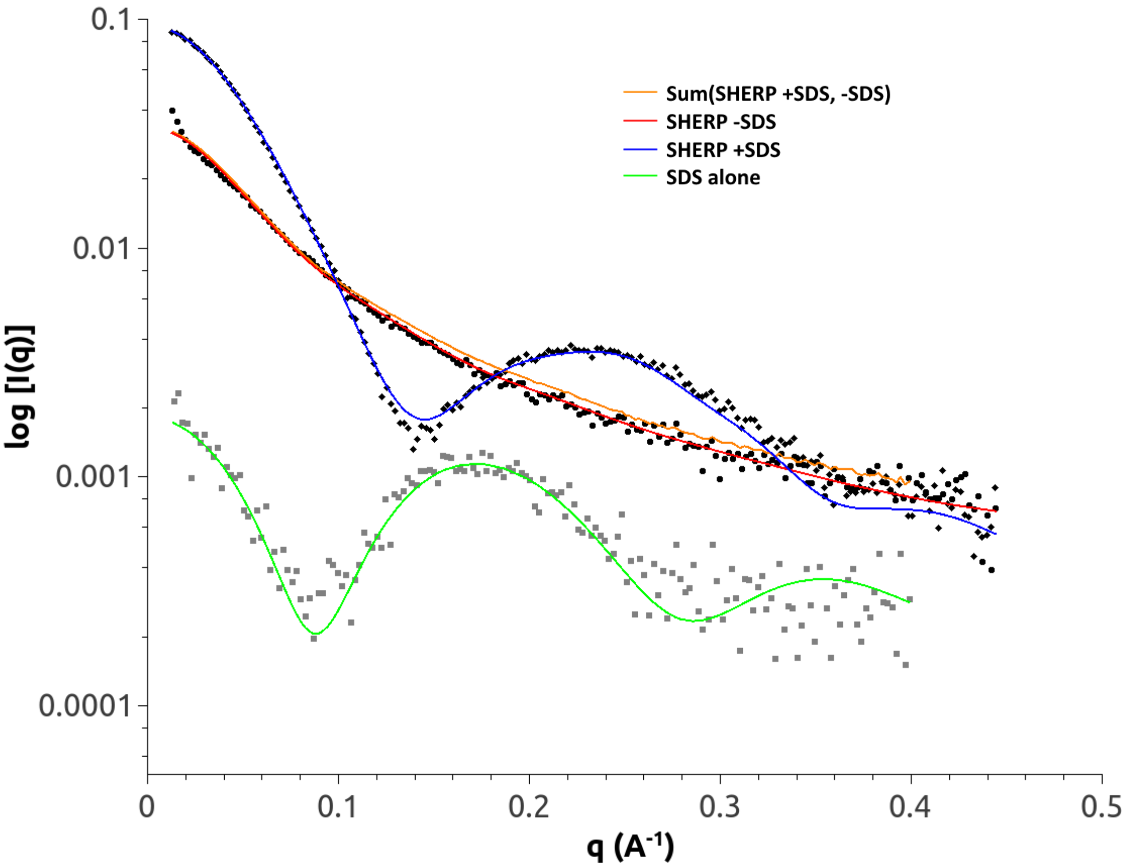


Figure 48: I(q) plots of SDS alone (green), SHERP in absence of SDS (red) and SHERP in the presence of SDS (blue). The sum of intensities for SDS alone and WT alone is also shown (orange). If SHERP and the micelle didn't interact we would expect the orange plot to be the same as the blue plot. As they are not we can confirm that SHERP and SDS micelles interact. The higher intensity of the WT-SDS spectra compared to the SDS alone spectra indicate that the micelles were much larger in the former sample.

Protein binding led to a narrowing of the paraffinic core of the cylindrical portion of the micelle, as the apolar-polar interface of the cylinder region shifted towards the centre of the micelle with respect to the micelle axis (~9-10 Å) (Figure 47). The percentage of micelle with no SHERP bound was predicted to be 0% (Table 9). However, 35% of the SHERP in the sample was predicted to be unbound and free in solution.

In the introduction, the similarities between SHERP and other lipid binding proteins (Davidson, 1998; Boucher et al. 2004; Wang et al. 1997) were discussed. It was hypothesised that SHERP may bind in a similar fashion to these proteins. These results provide clear evidence that SHERP binds among the head-groups of the micelle rather than penetrating the hydrophobic core – similar to the “snorkelling” binding mode of apolipoprotein C-II (MacRaild et al. 2004). The presence of multiple SHERP molecules on a single micelle raises the possibility of inter-protein interactions – this has been observed as a component of the “snorkelling” of apolipoprotein C-II. These protein-protein interactions might contribute to the stability of SHERP secondary structure.

Both alpha synuclein and apolipoprotein C-II have been shown to lead to reorganisation of the membrane – formation of disk-like particles (Varkey et al. 2013), extended tubules (Mizuno et al. 2012) and bending of anionic membranes (Varkey et al. 2010). Due to the similarities between SHERP and these proteins, and the results that indicate elongation of the micelle occurs upon the binding of SHERP to the micelle, it is possible that part of SHERP’s function *in vivo* involves some reorganisation of bilayer structure.

4.2.3 Effect of decreasing charge on SHERP’s interaction with SDS

The presence of stabilising salt bridges in the MD simulations of SHERP and SDS (Figure 42-45) and the fact that SHERP does not fold in neutral or positively charged lipids (Moore et al. 2011) provide evidence that the interaction of cationic side-chains with anionic head-groups play a crucial role in SHERP’s folding and interaction with the micelle. To further test this hypothesis, the series of constructs were expressed and characterised. These were designed so they would have increased negative charge, as follows: 2K = -6, 3K = -7, 3KR1 = -8 and 4K = -8. The alanine substitutions in each mutant were as follows: 2K – [K23A, K28A]; 3K – [K23A, K28A, K34A]; 3KR1 – [K23A, K28A, K34A, R39A]; and 4K – [K17A, K23A, K28A, K34A].

4.2.3.1 In the absence of SDS

In the absence of SDS, no significant differences compared the WT were seen by visual inspection of the mutant CD spectra obtained under the same conditions (Figure 49) apart from the 4K mutant which appeared to have increased secondary structure content compared to the WT due to the smaller peak at 200 nm. DichroWeb analysis corroborated this – the 2K, 3K and 3KR1 mutants were slightly more disordered than the WT (83-85% vs 78%), but the 4K mutant had 8% more strand content than the WT (Table 11). Alanines are helix-promoting residues, so it is surprising that their introduction into the sequence led to no gain in helicity. However, some studies suggest that the helix forming properties of alanine only manifest when the CO and NH groups of the residue are dehydrated by the presence of preferentially hydrated side-chains like charged or polar groups (Vila et al. 2000) – in the case of the mutants, charged groups are being removed from the protein which may provide less favourable conditions for helix formation due to alanine. SAXS also revealed some differences between samples (Figure 50). Fitting of the back-calculated intensities of the 67 PDBs obtained from the REMD simulations (Figure 33, Table 8) was carried out as described earlier for the WT (Figure 51). The 2K mutant had a very similar distribution to the WT, with predominantly unfolded models describing the data. However, the contribution from partially unfolded (16.2 Å < Rg < 23.6 Å) models increased for the 3K, 3KR1 and 4K mutants. The average Rg decreased as cationic residues were substituted with alanines, to a minimum of 22 Å for the 4K mutant (Figure 52, Table 12). The introduction of the more hydrophobic alanines may have resulted in the protein adopting a more compact structure due to the hydrophobic effect.

Comparing the six representative structures for each sample (Figure 34) also supported the trend towards more compact ensembles as cationic residues were replaced. Interestingly, the 3KR1 mutant showed a comparatively higher contribution from its top six models versus the other samples, and these models were all within ~5Å of the unfolded/partially unfolded boundary with no compact structures. It might be that the R39A mutation left the anionic E37 and D40 exposed to one another, and the repulsive nature of this interaction prevented more compact structures forming.

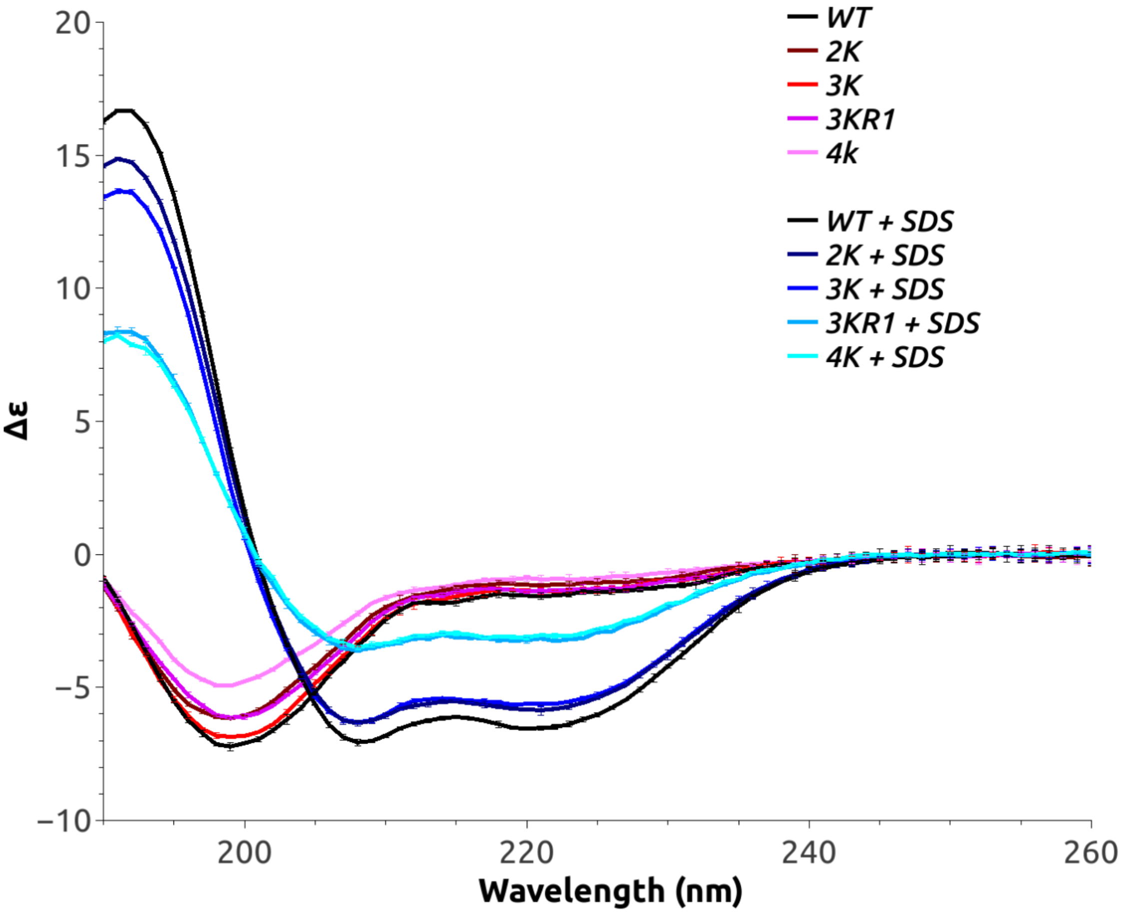


Figure 49: CD spectra of the WT and 2K, 3K, 3KR1 and 4K mutants in the presence (blue hues) and absence (red hues) of 0.1% SDS. With the loss of positive residues, the helical content of the mutants decreased. In the absence of SDS, differences between each protein were observed in the spectra but deconvolution with DichroWeb did not discern any large changes in secondary structure between the samples apart from the 4K mutant, which showed a 8% increase in strand content versus the WT.

Sample	Helix	Sheet	Turn	Other	NMRSD
WT -SDS	8 ± 2	6 ± 3	6 ± 2	80 ± 4	0.09 ± 0.07
WT +SDS	66± 3	4± 2	9± 5	18± 1	0.02± 0.01
2K -SDS	8 ± 1	8 ± 1	7 ± 2	78 ± 3	0.07 ± 0.04
2K +SDS	63 ± 3	6 ± 3	10 ± 5	21 ± 1	0.02± 0.01
3K -SDS	7 ± 2	7 ± 1	7 ± 2	79 ± 3	0.09 ± 0.07
3K +SDS	62 ± 4	6 ± 3	11 ± 4	21 ± 2	0.02 ± 0.02
3KR1 -SDS	9 ± 2	4 ± 3	8 ± 1	80 ± 2	0.12 ± 0.12
3KR1 +SDS	37 ± 2	19 ± 1	18 ± 0	27 ± 1	0.02 ± 0.02
4K -SDS	6 ± 0	8 ± 1	6 ± 1	81 ± 2	0.05 ± 0.02
4K +SDS	35 ± 1	20 ± 1	18 ± 0	27 ± 1	0.02 ± 0.02

Table 11: DichroWeb analysis of WT, 2K, 3K, 3KR1 and 4K mutants in the presence and absence of SDS. A small decrease in helicity is observed for the 2K and 3K mutants vs the WT, then a large drop as net negative charge increased further in the 3KR1 and 4K mutants.

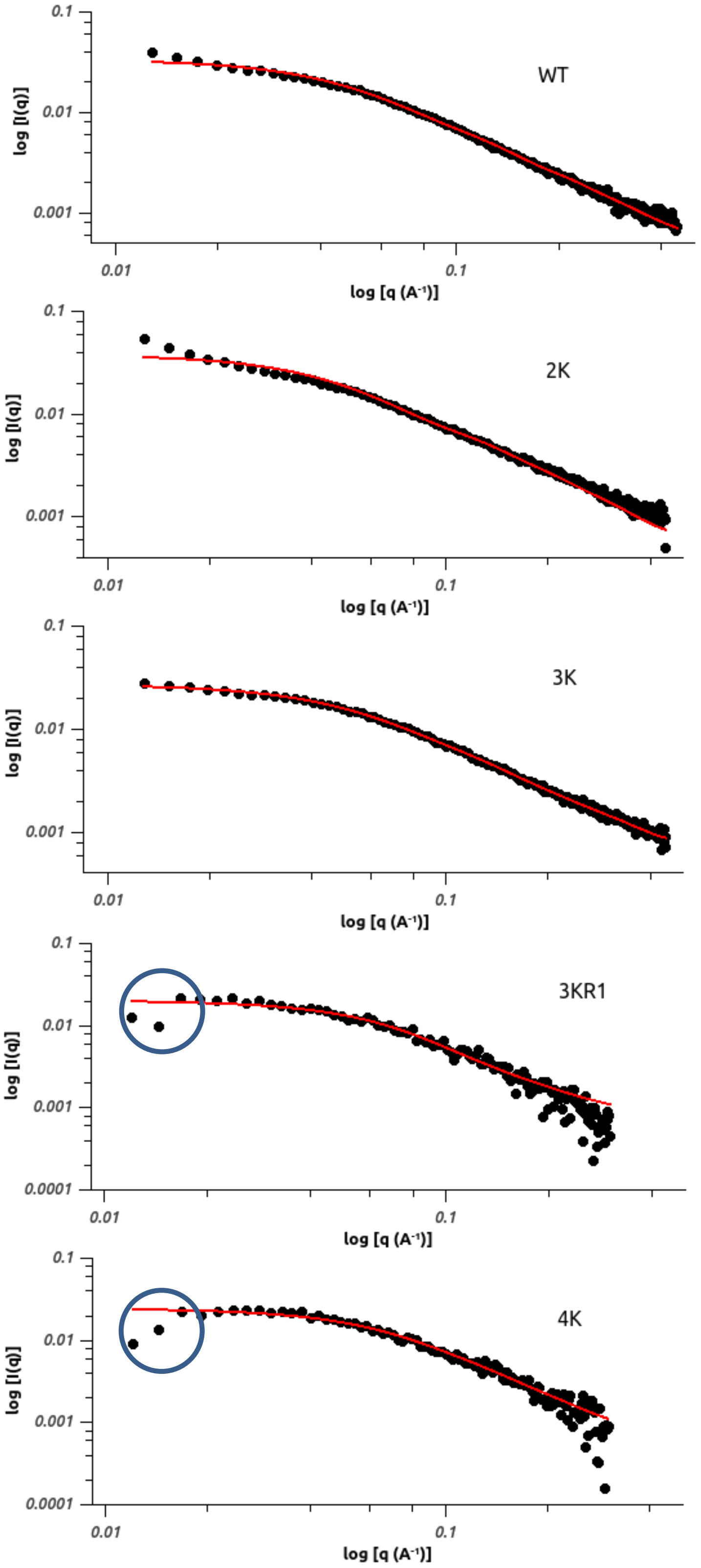


Figure 50: SAXS I(q) plots of the WT and 2K, 3K, 3KR1 and 4K mutants in the absence of SDS. Experimental data is shown as black dots, the fit obtained from fitting with REMD models in red. Evidence of a repulsive interference function in the low q range of the 3KR1 and 4K mutant is circled in blue. This repulsive interference probably stems from the increased net negative charge of the mutants affecting their interaction with one another.

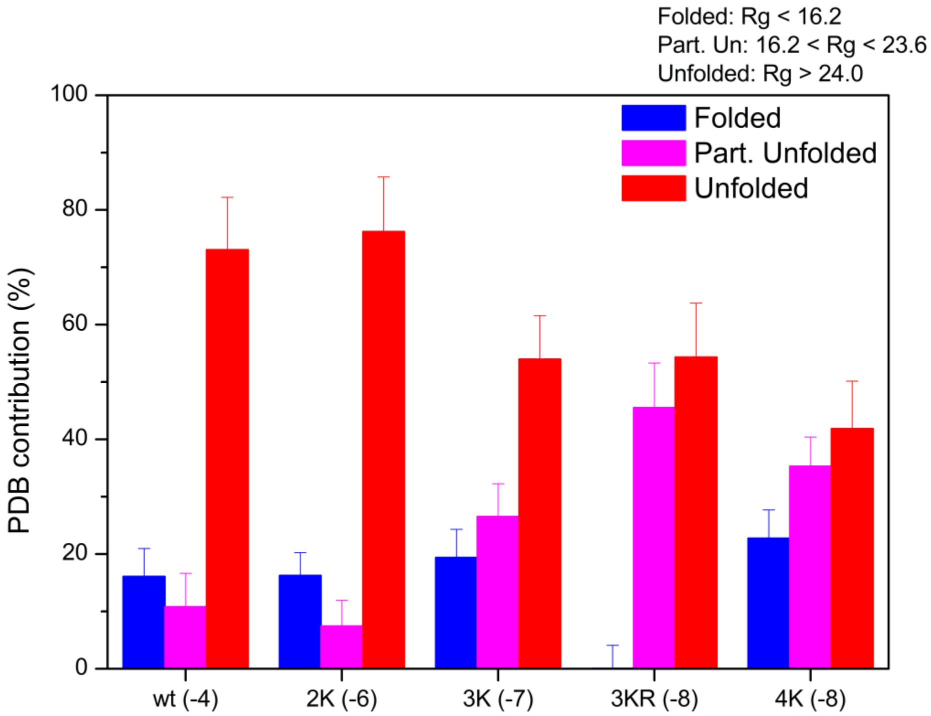


Figure 51: The distribution of Radius of gyrations for the WT and mutants in the absence of SDS. Data obtained by fitting 67 REMD PDBs to the experimental SAXS profiles. The Rg of each model used was calculated then weighted by its relative contribution to the best fit of the data, found using GENFIT. Structures were classified as folded (Rg < 16.2 Å, blue), partially unfolded (16.2 Å < Rg < 23.6 Å, magenta) and unfolded (Rg > 24.0 Å, red). A trend towards a best fit with an ensemble comprising lower Rg structures was observed as the net charge of the protein becomes more negative.

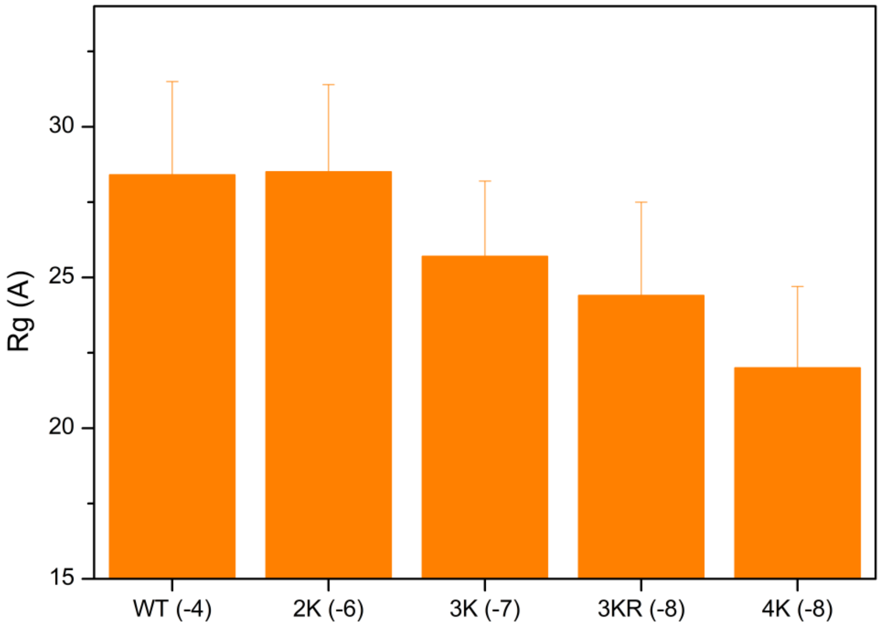


Figure 52: Average radius of gyration of PDB fit ensembles. The average Rg of the WT and each mutant, obtained from the averaged Rg of each of the 67 REMD PDBs used for fitting the experimental SAXS data in the absence of SDS, weighted by their contribution to the fit. The proteins become more compact as positively charged residues are removed – the replacement of charged residues with hydrophobic alanines might have led to this effect due to better hydrophobic packing.

Sample	Rg (A)
WT (-4)	28.4 ± 3.1
2K (-6)	28.5 ± 2.9
3K (-7)	25.7 ± 2.5
3KR (-8)	24.4 ± 3.1
4K (-8)	22 ± 2.7

Table 12: Radius of gyration values of the WT and each mutant in the absence of SDS. Rgs obtained from the weighted contribution of each of the 67 REMD PDBs used to fit the experimental SAXS data in the absence of SDS.

More could be inferred from the low *q* region of the 3KR1 and 4K mutants (Figure 50). The dip in intensity in this low q region suggested the presence of repulsive inter-particle effects for the two most negative mutants. This makes sense, as the net charge would be expected to cause a repulsive effect between individual protein molecules.

4.2.3.2 In the presence of SDS

As our hypothesis stated that charged residues likely played a large role in the interaction of the protein with the micelle, it would be expected that large changes would occur both in the folding of the protein and the organisation of the SHERP-SDS complex when comparing the mutants to the wild-type. From CD spectroscopy in the presence of SDS, it could clearly be seen that the 191 nm, 208 nm and 222 nm peaks decreased in magnitude as the negative charge of the protein increased (blue hued spectra), demonstrating that alpha helicity decreased for all the mutants with respect to the wild-type (Figure 49). DichroWeb analysis of the spectra supported this qualitative assessment of the data (Table 11). All mutants had decreased helicity and increased “other” compared to the WT but a much larger drop was seen when protein net charge decreases to -8, with the 4K and 3KR1 mutants causing the largest change with decreases to 35% and 37% helix respectively. Sheet and turn was also seen to increase compared to the WT. The results confirmed that removal of cationic residues led to decreased helicity in the sample. There could be several explanations for this effect. The protein may have had reduced affinity for the micelle due to increased negative charge but still have folded correctly upon binding (leading to an increase in free SHERP in solution), it may have only partially folded but still have had affinity for the micelle comparable to the wild-type, or it could be a combination of these two factors.

To try and answer this, SAXS experiments under the same conditions as the WT were performed in the presence of SDS. Changes in the position of the first peak of the I(q) plots could be seen – the peak shifted towards smaller *q* values (in the direction of the SDS micelle alone profile) as more positively charged residues were removed (Figure 53). The organisation of the Mutant-SDS systems must therefore differ from that of the WT-SDS complex.

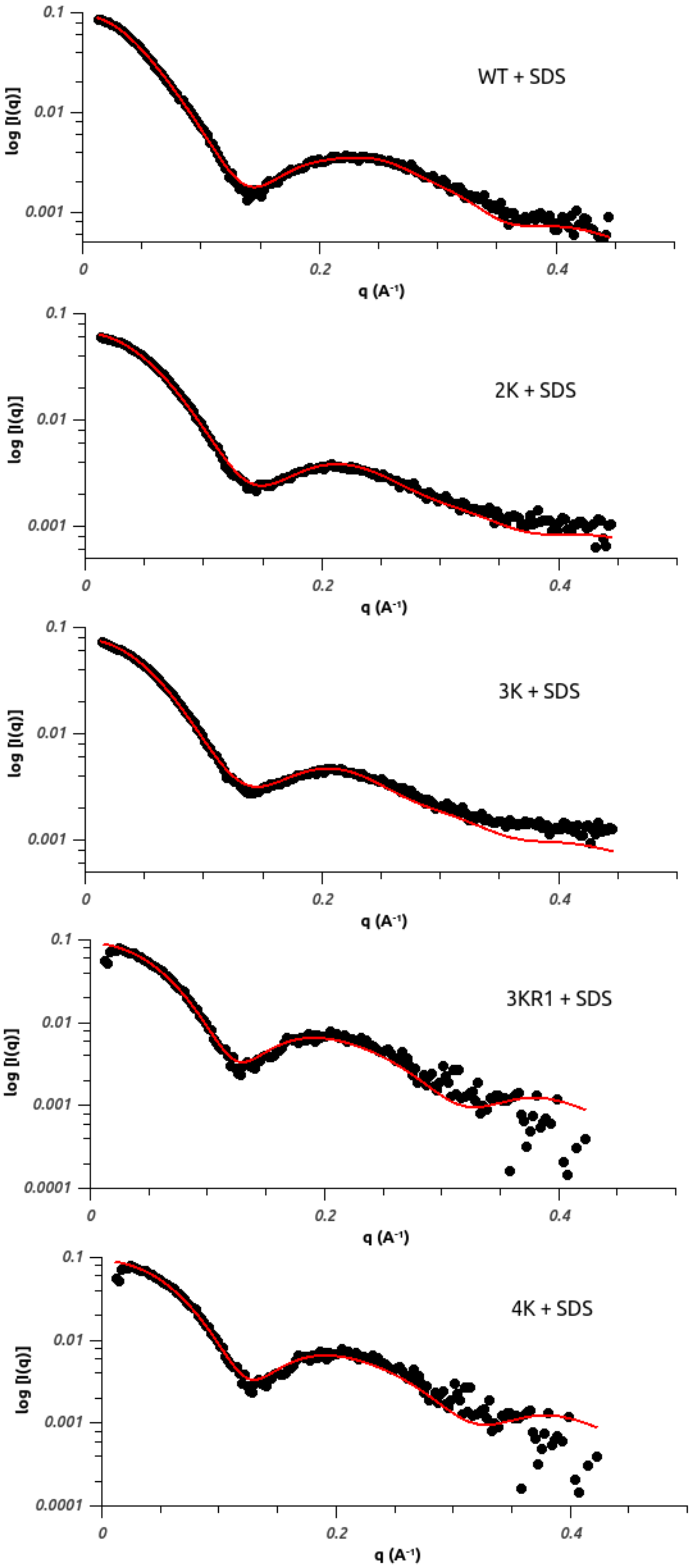


Figure 53: I(q) plots from SAXS experiments of the WT and 2K, 3K, 3KR1 and 4K mutants in the presence of 0.1% SDS. Experimental data shown as black dots, spherocylindrical model fit in red.

Sample	Micelle Length (A)	Agg No.	No. Protein bound	Free protein (%)	Free Micelle (%)
SDS alone	67 ± 2	106 ± 8.5	n/a	n/a	100.0 ± 0
WT + SDS	116 ± 3	88 ± 5	6.1 ± 0.5	35.4 ± 2	0.0 ± 0
2K + SDS	75 ± 1	65 ± 0	2.8 ± 0	60.0 ± 3	0.5 ± 0.1
3K + SDS	79 ± 0	66 ± 0	2.8 ± 0.1	60.0 ± 4	1.0 ± 0.1
3KR1 + SDS	62 ± 0	57 ± 0	2.0 ± 0	49.1 ± 1	1.6 ± 0
4K + SDS	65 ± 0	57 ± 0	2.1 ± 0	53.3 ± 2	2.0 ± 0.1

Table 13: Parameters obtained from fitting of models to SDS alone, WT-SDS and mutant-SDS SAXS data using GENFIT (Spinozzi et al. 2014)*. Micelle length increases considerably upon binding of the WT, while aggregation number decreases. Multiple protein molecules were found to bind per micelle, perhaps driving this elongation. Fewer protein molecules bound per micelle for the mutants, with a greater proportion of free protein and micelle and a shorter micelle length.*

The same three layer spherocylinder model was used for fitting the mutant data (Table 13). The most dramatic change from the WT was seen in the length of the micelle axis (Figure 54). For the mutants, micelle length was greatly reduced – the 2K (75.4 Å ± 1.)] and 3K (79.1 Å ± 0.85) mutants were still elongated with respect to SDS alone (66.6 Å ± 2.3) but the 4K mutant (65.3 Å ± 0.65) was almost identical in length to the micelle in the absence of protein, and the 3KR1 mutant (62.2 Å ± 0.55) was even smaller. Once again, the presence of the protein reduced the aggregation number of the cylinder region of the micelle (Figure 55). The shift in the peak of the SAXS profiles as negative charge increased follows the decrease in micelle length, suggesting the two are linked.

Even with the decreased micelle length observed, a number of protein molecules must have bound to the micelle in order to explain the results. The number of protein molecules bound to each micelle decreased significantly for all the mutants when compared to the WT (Figure 56). The samples with the largest number of mutations, 3KR1 and 4K, had the lowest number of SHERP bound per micelle, 2.0 and 2.1 respectively down from 6 for the WT. This did suggest that there was a significant decrease in affinity for the micelles due to the decrease in positive residues, and demonstrated the importance of cationic residues in stabilising the larger, rod-like micelle seen in the WT. As with the WT, the mutants were only found to be bound to the cylindrical portion of the micelle – again showing favouritism towards the flatter binding surface of this region.

The proportion of free, unbound protein was higher for all the samples with respect to the WT, which also suggested a decreased binding affinity (Figure 57) – especially as the proportion of free micelle increased as the number of Lys/Arg to Ala substitutions increased (Figure 58), even though the protein was in even greater excess due to the smaller number of molecules bound per micelle (Table 13). Increasing the negative charge of the protein led to decreased affinity for the micelle compared to the WT. The proportion of free SHERP was highest for the 2K and 3K mutants (Figure 57). The 3KR1 and 4K mutants, although higher than the WT, had a slightly lower proportion of free SHERP than the 2K and 3K even though fewer SHERP bound per micelle (Figure 56) and the helicity of the sample was lower (Table 11). By comparing the decrease in helicity determined by CD and the increase in the proportion of free SHERP we could see that both decrease relative to the WT for the

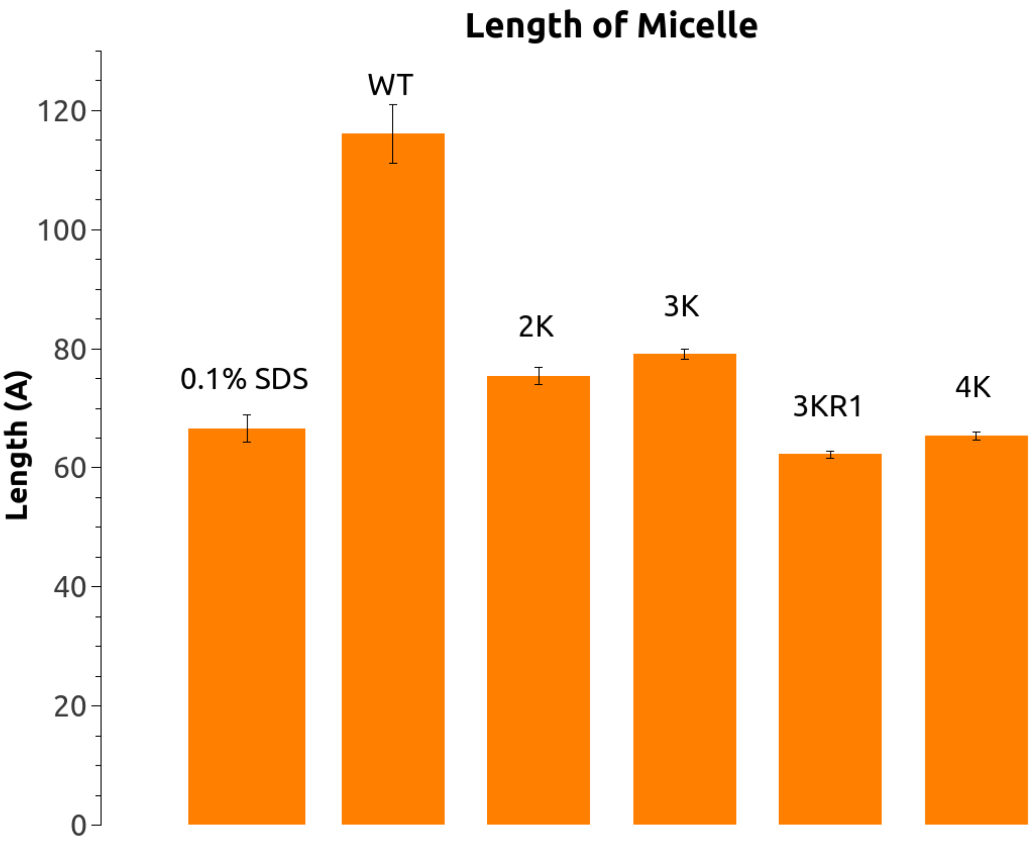


Figure 54: Length of the longest micelle axis for SDS alone, the WT-SDS and mutant-SDS systems from SAXS modelling. Values were obtained from modelling of the SAXS data, with standard deviations arising from 10 repeats of the fitting process. The lengths of the micelle-mutant complexes were greatly reduced compared with the WT, becoming more similar to that of SDS alone.

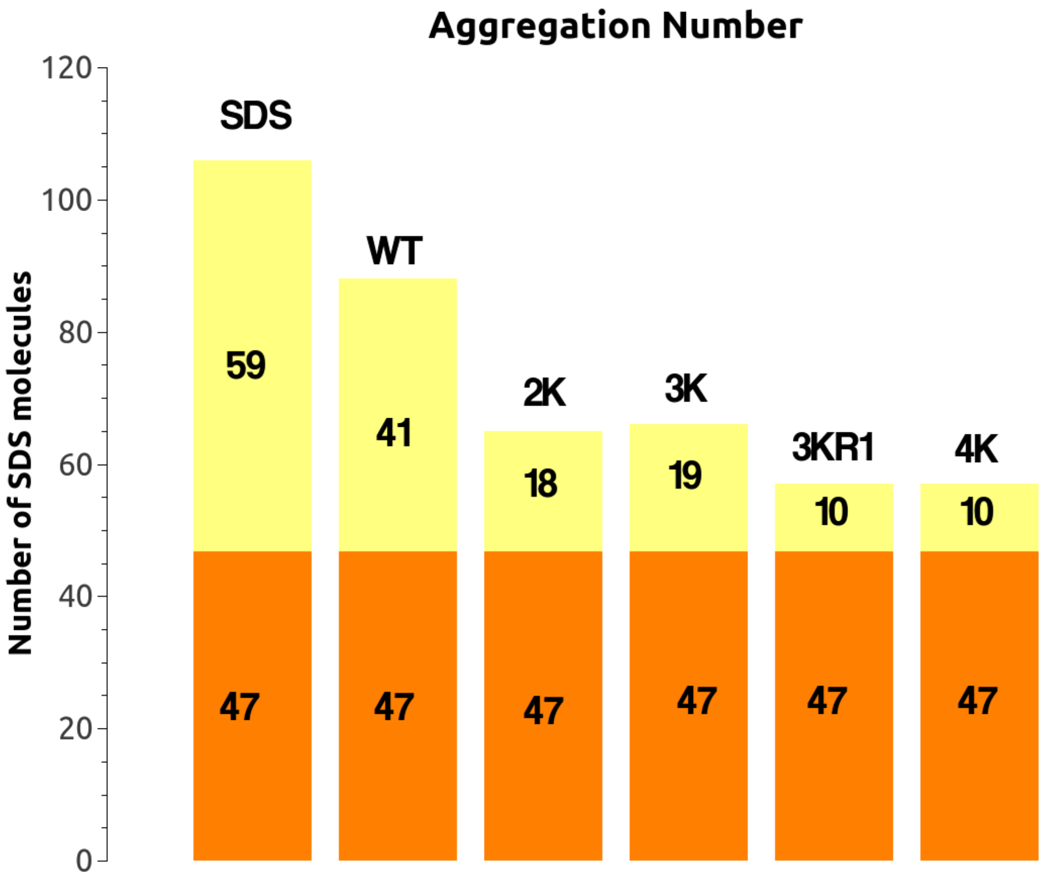


Figure 55: Aggregation number for SDS alone, the WT-SDS and mutant-SDS systems from SAXS modelling. The number of SDS molecules in the hemispherical region (orange) and cylindrical region (yellow) are shown. The aggregation numbers of the protein bound micelles were significantly lower than that of SDS alone, even though the micelle lengths were at least the same magnitude or larger. The decrease comes from the cylindrical portion of the micelle – this is where the protein was found to bind in all samples.

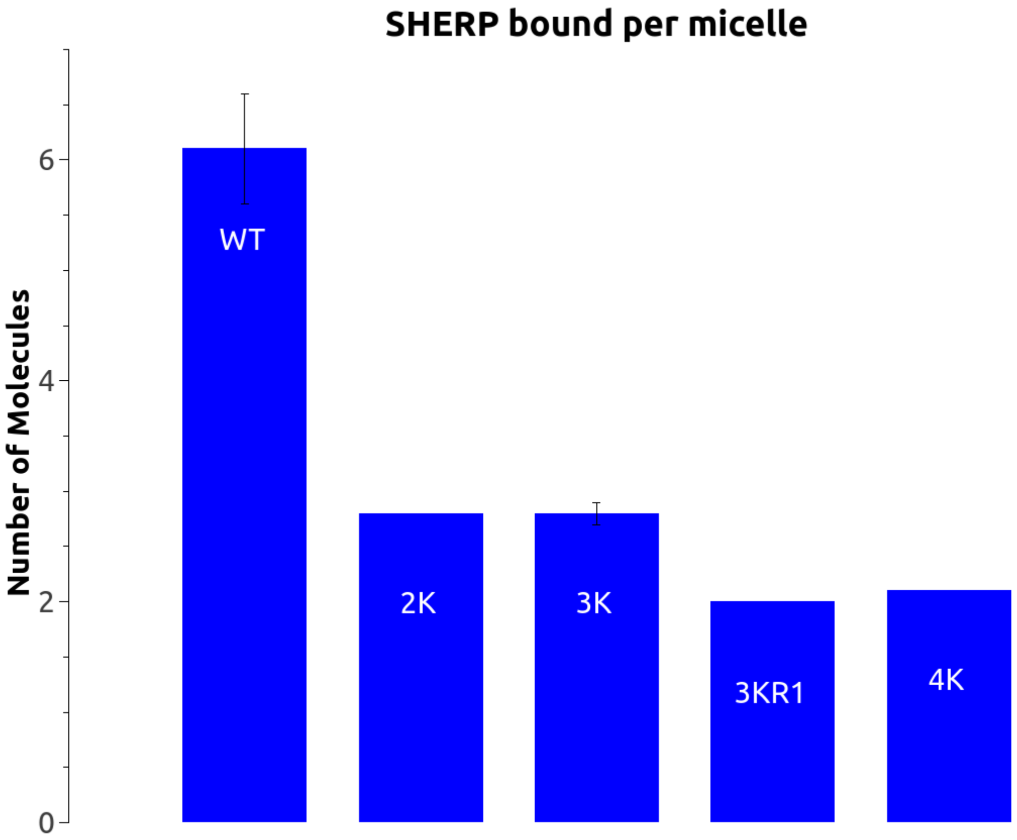


Figure 56: Number of SHERP molecules bound for the WT-SDS and mutant-SDS systems from SAXS modelling. As negative charge increased, the number of SHERP molecules bound to the micelle decreased. The decreases have a very similar pattern to the decreases seen in the micelle length (Figure 54), so it could be that the number of protein molecules bound dictated the length of the micelle.

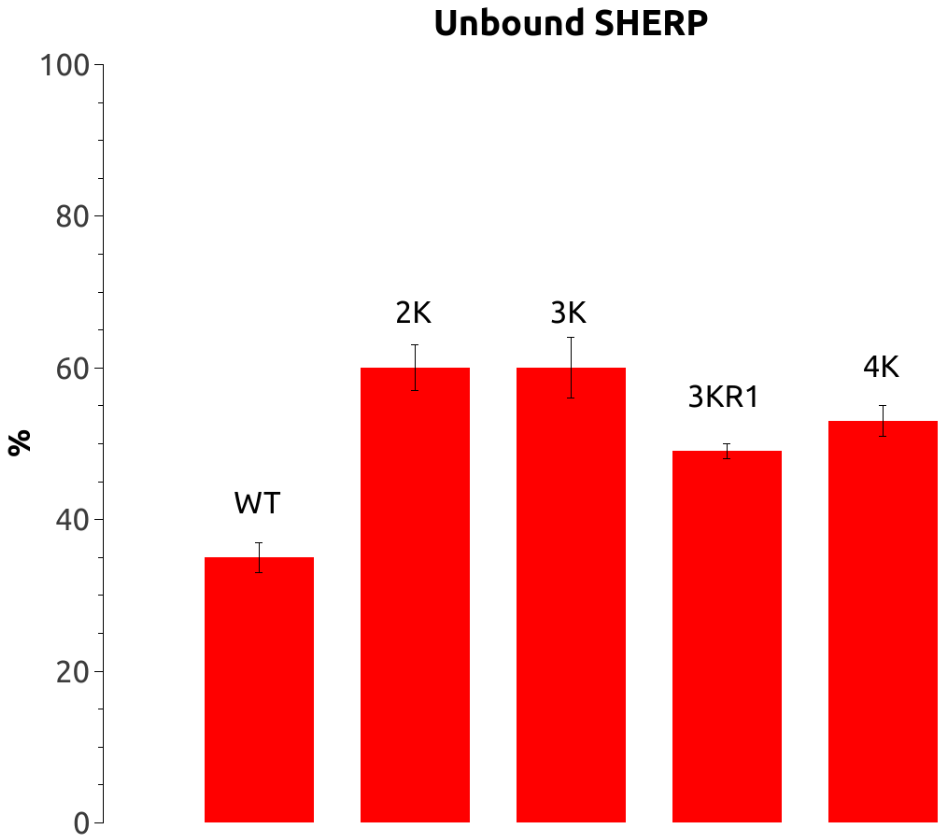


Figure 57: Proportion of unbound SHERP molecules for the WT-SDS and mutant-SDS systems from SAXS modelling. The increase in free SHERP in the mutants compared to the WT suggests the a decreased affinity of the protein for the micelle.

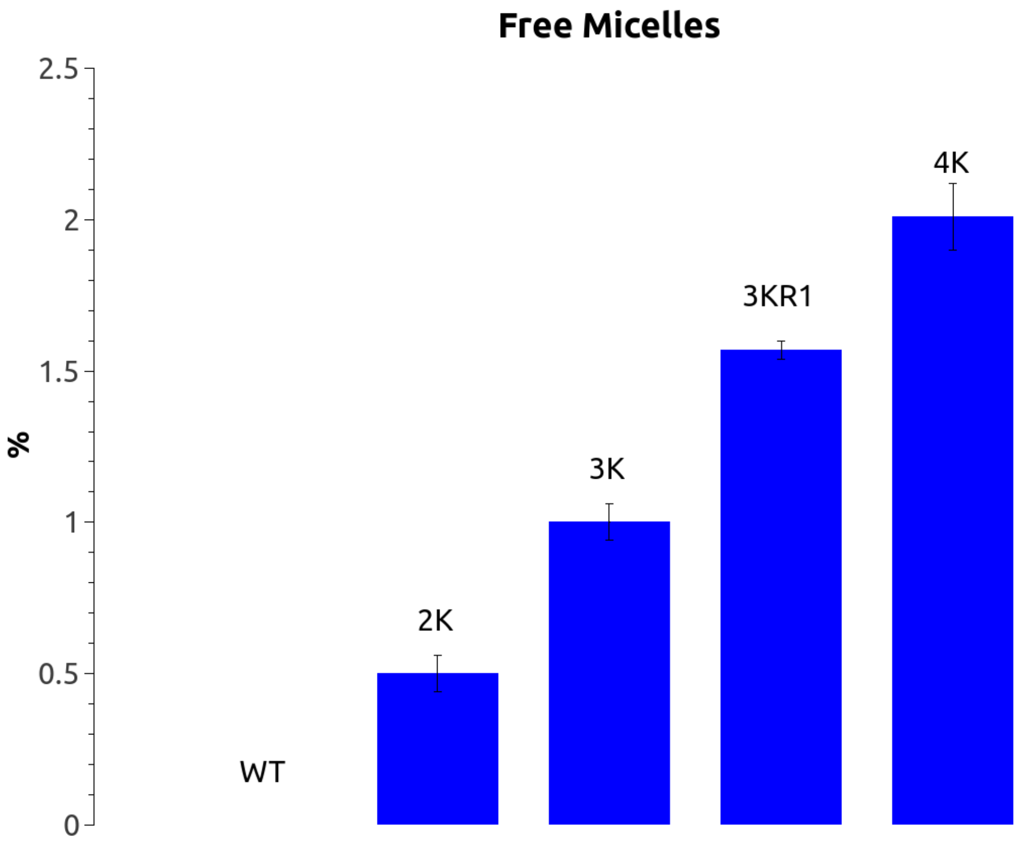


Figure 58: Proportion of free, unbound SDS micelles observed in the WT-SDS and mutant-SDS systems by SAXS modelling. Changes in free micelle were small but consistent as negative charge increased, suggesting that charge affected the proteins affinity for the micelle.

2K and 3K mutants. However, the helicity decreased in the 3KR1 and 4K mutants compared to the 2K and 3K mutants while the proportion of free SHERP decreased (Figure 59). A decrease in free SHERP would be expected to lead to an increase in sample helicity if the protein was fully folded, which is not what we observed. This suggested that the decreases in helicity seen for the 2K and 3K were primarily due to decreased affinity of the mutants for the micelle, while the larger decrease in helicity observed in the 3KR1 and 4K mutants was due to not just decreased affinity for the micelle, but partial, incomplete folding of the protein when it bound to SDS.

The electron density profiles of the mutants also supported this proposal (Figure 60). The 2K and 3K mutants had a profile (for the cylinder region) very similar to the WT, demonstrating that the protein (responsible for the increase in electron density and movement of the polar-apolar interface towards the micelle axis) occupied the same volume for all 3 of these samples – this suggested that the 2K and 3K were probably folding properly on the micelle. The broadening of the profile to higher radius values seen for the 3KR1 and 4K mutants, while the polar-apolar interface remained shifted inward with respect to the SDS alone profile, suggested that these mutants occupied a larger volume than the WT. Therefore, the 3KR1 and 4K mutants might have been partially unfolded when bound to the micelle.

Previously, the hypothesis that protein-protein interactions between SHERP molecules bound to the micelle might have a role in stability of SHERP secondary structure was put forward. It can be seen from the results for the mutants that as the number of protein bound decreased, so did the helicity. Perhaps with fewer molecules bound there were fewer opportunities for inter-protein interactions which decreased the potential stabilising effect it could be having on secondary structure. Therefore, it may be that the loss of cationic residues (increasing protein negative charge) decreases affinity of the protein for the anionic micelle leading to fewer protein molecules binding. This in turn reduced the chance of stabilising interactions occurring between protein molecules, leading to the partial folding described above.

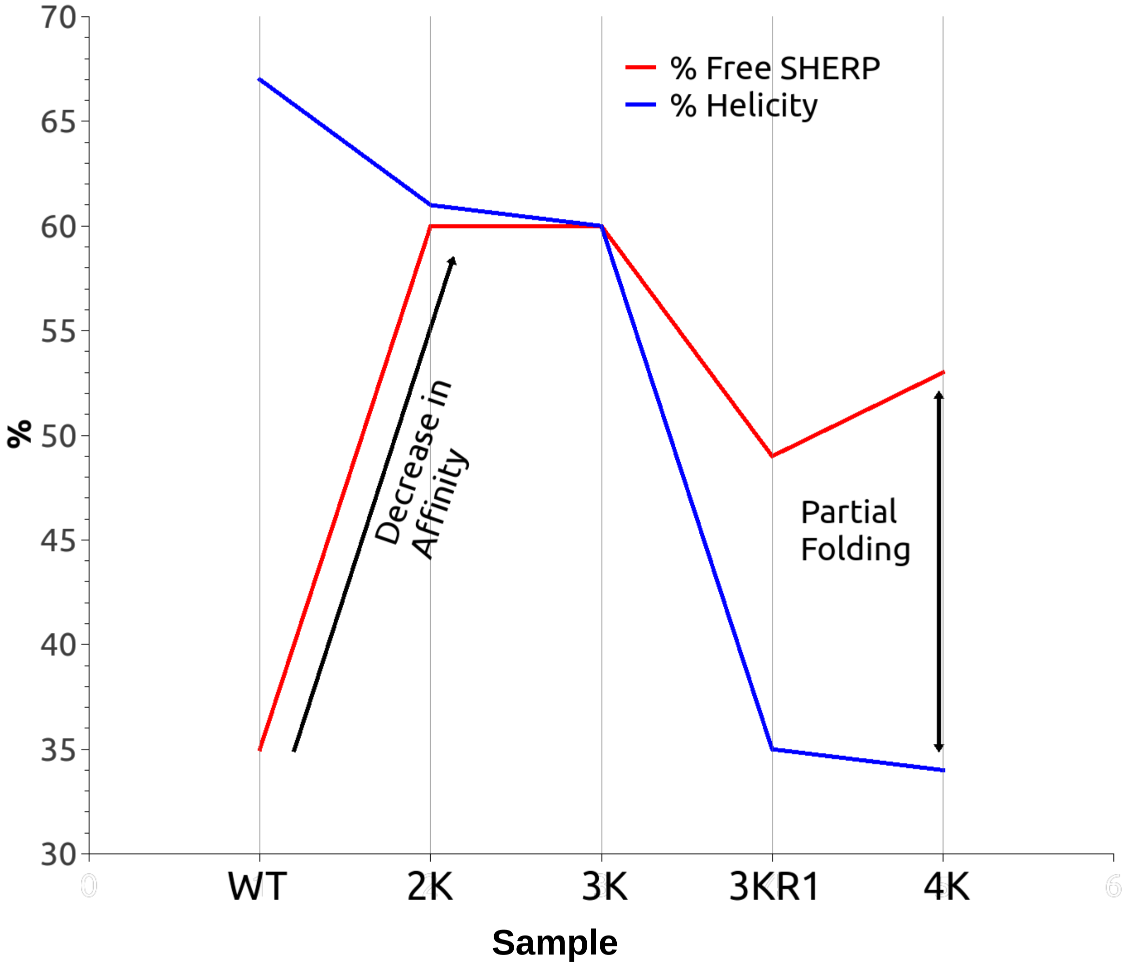


Figure 59: Plot showing the relationship between percentage of free unbound SHERP (red) and percentage helicity (blue) of the WT and each mutant. The large increase in free SHERP with a small decrease in helicity seen from the WT (net charge -4) to the 2K (net charge -6) and 3K (net charge -7) mutants suggests a loss of affinity for the protein to the micelle. The large decrease in helicity seen in the 3KR1 and 4K mutants (net charge of -8) cannot be explained by an increase in free SHERP – it actually decreases compared to the 2K and 3K mutants. This suggests that decreased binding affinity is not the only phenomenon in effect, and that loss of secondary structure due to partial folding of the protein on the micelle surface is also occurring for the mutants with the most negative net charges. This partial folding may be caused by the decreased chance of SHERP-SHERP interactions occurring due to fewer protein molecules binding per micelle (Figure 56).

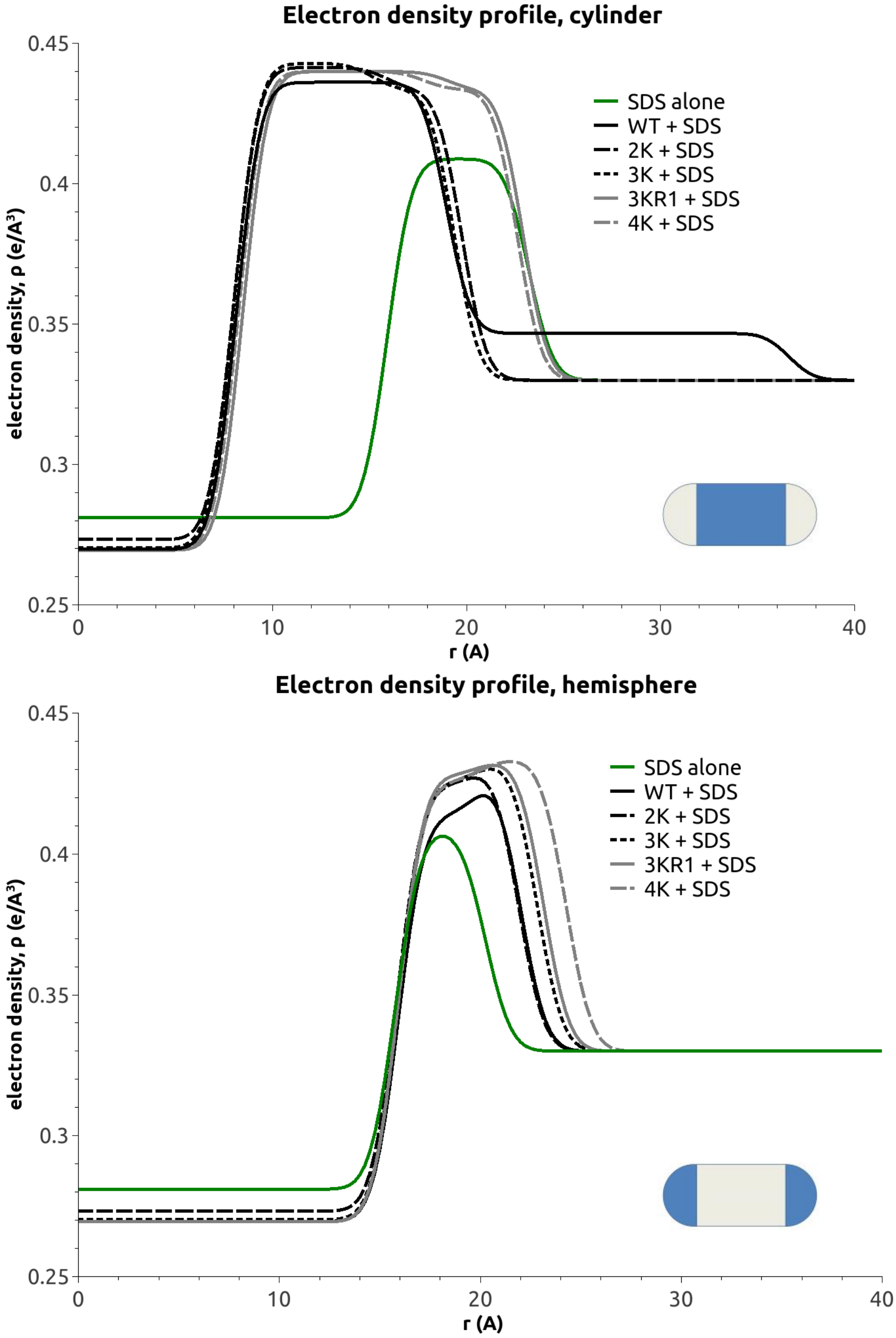


Figure 60: Electron density profiles for the SDS alone (green), WT-SDS (black, solid), 2K-SDS (black dashes), 3K-SDS (black dotted), 3KR1-SDS (grey, solid) and 4K-SDS (grey dashes) systems. Electron density was obtained from modelling of the SAXS I(q) data with GENFIT. Electron density is shown for the cylindrical region (top, indicated in blue on inset diagram) and hemispherical region (bottom, indicated in blue on inset diagram) of the micelle model. The broader density distribution observed for the 3KR1 and 4K mutants are indicative of the protein volume increasing, suggesting the protein was only partially folded compared to the WT.

4.3 Conclusions

In this chapter, the disordered ensemble of SHERP was characterised for the first time by REMD, SAXS and CD spectroscopy. Good agreement was found between the computational the biophysical methods used. The protein was found to be predominantly unfolded in solution, preferentially adopting extended conformations with an ensemble Rg between 22 Å and 28 Å from different methods. Helical content was low at 8%, with the majority of the structure considered unordered.

The organisation of the SHERP-SDS complex and the changes that occur to the structure of the micelle during the disorder-to-order transition of the protein were characterised *in silico* and *in vitro*. Simulation of the SHERP-SDS complex found the helix-turn-helix conformation seen in the NMR structure (Moore et al. 2011) was stabilised with helical content in agreement with our prior CD spectroscopy results, and showed the protein bound in the polar head-group region of the micelle. The presence of persistent, stabilising salt bridges was observed between cationic side-chains and the head-groups of SDS. SAXS data of the WT-SDS system was obtained and modelled using the GENFIT (Spinozzi et al. 2014) software, showing large structural changes occurred upon protein binding. These protein-induced changes to micelle structure draw parallels to the action of similar proteins on lipid bilayers (Varkey et al. 2010; Varkey et al. 2013; Mizuno et al. 2012). This raises the possibility that part of SHERP's function *in vivo* may involve reorganisation of the membrane structure. Multiple SHERP molecules were shown to bind to each micelle, leading to significant lengthening of the micelle compared to SDS alone. The protein bound in the shell of the micelle, amongst the anionic head-groups of SDS. This mirrors the "snorkelling" binding mode observed in other proteins like apolipoprotein C-II (MacRaild et al. 2004).

Several cationic residues identified in the MD simulation as having a range of salt-bridge forming propensities were substituted with alanines in a series of mutants with increasing negative charge, which were expressed and characterised by SRCD spectroscopy and SAXS to examine the importance of charge interactions in the interaction. The increase in negative charge led to dramatic changes, with large decreases in helicity, decreasing micelle length, fewer protein molecules bound, a higher percentage of unbound SHERP and micelles and large changes in the distribution of electron density in the polar head-group region as negative charge increased. These results led to the conclusion that charge-charge interactions are integral to the complex formed between SHERP and SDS, and there is strong evidence that their loss results in decreased affinity of the protein for the micelle leading to fewer protein molecules binding, which may lead to a decreased chance of secondary structure stabilising protein-protein interactions resulting in the partial folding observed via the decreased helicity and broadening of the electron density in the 3KR1 and 4K mutants.

5.0 Chapter 5: Conclusions

5.1 Summary

In Chapter 3 (Section 3.0), the effect of point mutations on the disordered ensemble of SHERP was explored. Predictions were made using DISOPRED2 and mutants were selected for expression and characterisation. In the course of this previous results from the literature showing the disorder-to-order structural transition were successfully reproduced. However, changes in disorder predicted for the mutants were not realised in the *in vitro* results, with no significant changes seen due to any point mutation apart from increased thermal stability and helicity for the D29W mutant in the presence of SDS, and the increase in secondary structure content for the S43F mutant in the absence of SDS as the temperature increased.

Another mutant sequence, the PM, was designed to introduce helix stabilising salt bridges into the structure while making minimal alterations to the proteins amino acid composition. Molecular dynamics simulations showed that the PM retained greater helicity and formed a much larger number of salt bridges than the WT. The helix retention was corroborated by the greater helicity and decreased other of the PM observed by CD spectroscopy in the absence of SDS. However, SAXS data for the WT protein in the absence of SDS showed that it was extended in solution, which contradicted the compact conformations adopted during the MD simulations.

The disordered ensemble of SHERP was characterised by REMD, CD spectroscopy and SAXS scattering in Chapter 4 (Section 4.0). Good agreement was observed between all the methods used. The protein was highly disordered and extended in solution, and after clustering was observed to have a bimodal distribution of conformations – a smaller, compact population centred at an Rg of ~14.9 Å and a broader, extended population around ~28.4 Å. Helical propensity was significantly higher in the low Rg cluster, suggesting that perhaps these population could provide transient secondary structure and allow recognition of the protein for the micelle, driving folding. These two populations were observed over two runs of REMD simulation and during fitting of REMD structures to the experimental SAXS data.

Prior to this work, the organisation of the SHERP/SDS micelle system was unknown. It was hypothesised that cationic side-chains on the protein played a key role by interacting with the anionic SDS head-groups – the protein does not fold without the presence of anionic lipids or detergents. Molecular dynamics simulations indicated that SHERP bound in the head-group region of the micelle, and that this led to retention of the helix-turn-helix motif seen by NMR. Persistent salt bridges were observed between cationic side-chains and sulphate head-groups. SAXS provided greater insights into what occurs when SHERP binds to the micelle. The presence of the protein led to elongation of the micelle, with approximately six SHERP molecules bound in the head-group region of each micelle. This elongation might indicate that SHERP plays a role in reorganisation of the lipid bilayer *in vivo.*

From the SHERP/SDS MD simulation, several cationic residues were identified that had a range of salt-bridge forming propensities with the SDS head-groups. Alanine substitutions of these residues were expressed and characterised by CD spectroscopy and SAXS. The secondary structure content of the samples decreased as the negative charge of the protein increased – suggesting that either the protein was partially folded on the micelle, less protein was binding or a combination of both these phenomena. The mutants still bound to the head-group region of the micelle; however, fewer molecules bound per micelle, and the micelles were much less elongated with respect to the wild-type bound micelle. The proportion of free protein increased, as did (to a smaller degree) the proportion of free micelle. This suggested that a large proportion of the decrease in secondary structure was due to lower affinity of SHERP for the micelle. However, a broadening of the peak in the electron density profile of the most negative mutants away from the central axis of the micelle, and evidence from the modelling that a greater proportion of the protein was found in the exterior solute of the micelle for the mutants compared to the WT suggests that the changes were probably due to both decreased affinity and partial folding potentially due to a decrease in stabilising protein-protein interactions. In summary, the presence of protein induced large changes in micelle organisation, size and structure, and the interactions between cationic side-chains and the anionic head-groups of the micelle were found to be integral to the disorder-to-order structural transition of SHERP in SDS.

This work contributes to the literature in several ways. It provides an example of a well characterised ensemble of a novel parasitic IDP. As a researcher, it can be hard to obtain good examples of IDP ensembles – the protein ensemble database as of the 18th of July 2017 contains 60 ensembles, but these come from just 16 proteins (Varadi and Tompa 2015). When this work is published, the author intends to submit the ensemble described in this work to the database. The impact of the PDB on the study of ordered proteins cannot be overstated; it is probably the most important resource in structural biology, and large swathes of bioinformatics research would never have been published if it did not exist. It is of great importance that more structural ensembles are published and disseminated to allow the same kind of growth in understanding and knowledge of intrinsic disorder.

The binding mode of SHERP provides another example of the "snorkelling" binding mode of IDPs seen in many amphipathic membrane associated proteins (MacRaild et al. 2004). The changes seen in the micelle structure provide further clues to SHERP's function – parallels can be drawn to the membrane reorganisation effected by alpha synuclein and apolipoproteins (Varkey et al. 2010; Varkey et al. 2013; Mizuno et al. 2012), suggesting that SHERP may perform a similar function *in vivo* combined with its known interactions with V-ATPase (Moore et al. 2011).

As was related earlier in the thesis, SHERP is essential to the progression of the *L. major* life cycle (Doehl et al. 2017; Sádlová et al. 2010). No vaccine is in clinical use for humans as of 2017, treatments for leishmaniasis often come with unpleasant side effects, and some strains of *Leishmania* have developed a resistance towards certain classes of drug (Singh and Sivakumar 2004). By characterising the structure of the SHERP-SDS complex and identifying that charge-charge interactions play a key role, this work can provide structural context to functional studies carried out by researchers in the future, perhaps one day resulting in new treatments for the disease.

5.2 Future Work

Chapter 3 showed the failure of sequence based methods for predicting the effect of point mutations on the disorder of SHERP. A method published recently, GADIS, which attempts to design sequences that alter disorder content while maintaining amino acid composition and known functional sites, was discussed (Harmon et al. 2016). This method involves an iterative genetic algorithm, each step involving the simultaneous simulation of hundreds of sequences to optimise secondary structure content. This obviously involves a large amount of computing resources inaccessible to many researchers. However, this method, or elements of this method, would enable much more rigorous and accurate prediction of the effect of mutations on SHERP secondary structure content. As the key residues for the interaction of SDS with SHERP have been identified by experiments in this thesis, the GADIS approach could be employed to identify functional SHERP mutants with varying degrees of disorder.

Probably the most obvious work that could be done to build on this thesis would be characterisation of SHERP in the presence of anionic lipids. SDS was used for several reasons in this project; its ease of use in both CD and SAXS spectroscopy, the simplicity of implementing SDS micelle in molecular dynamics simulation and most importantly the fact that the NMR structure of SHERP was solved in SDS allowing for a baseline, *in vitro* structure to compare MD results to. However, SDS is not a physiological system – in reality the protein would be interacting with anionic lipids in organelle and cell membranes. Previous work showed that there were small differences in secondary structure of the protein in SDS versus anionic lipids (Moore et al. 2011). By repeating the experiments in this thesis with anionic lipids it could be seen whether the phenomena seen using SDS as a model still hold true in a more physiologically accurate environment. The hypothesis that SHERP function may involve some reorganisation of the membrane could be tested using electron microscopy imaging of lipid bilayers in the presence of SHERP.

Due to time constraints, only five out of the ten cationic residues analysed in the SHERP-SDS simulation were used in the alanine substitution mutants. It would be interesting to further alter the charge of SHERP by substituting more of these cationic residues with alanines to see if the trends observed in Chapter 4 continue.

Although we have concluded that increasing the negative charge of the protein leads to large changes in how SHERP interacts with SDS, we have not explored how this might affect its function. Previous studies in the literature have shown that L. *major* parasites with the SHERP containing locus knocked-out are not able to progress in their life cycle (Doehl et al. 2017; Sádlová et al. 2010). Would the introduction of the mutant SHERP genes described in this thesis arrest the life cycle of the parasite in similar fashion? Would it lead to reduced pathogenicity or virulence? On the matter of protein function, cross linking studies have shown that SHERP interacts with a V-ATPase (Moore et al. 2011). If these cross linking studies were repeated with the mutant proteins, would this interaction still occur?

The disordered ensemble of SHERP was calculated by REMD and validated by SAXS and CD. A method like NMR could provide more detailed information and further improve the ensemble obtained. It is a common practice to use NMR data of small, disordered proteins like SHERP as constraints in MD (Fisher and Stultz 2011; Rosenman et al. 2013), or even through statistical coil generation using programs like Flexible-Meccano (Ozenne et al. 2012). The protein ensemble database mentioned previously has been developed to store and validate ensembles calculated through methods like the above (as well as SAXS) – as of June 2017, 60 ensembles of 16 proteins encompassing 25473 structures have been deposited, many of which rely on NMR data (Varadi and Tompa 2015).

NMR spectroscopy would also be of great utility in characterising the ordered state of the mutants in the presence of SDS. It could answer the question of whether decreased protein-protein interactions between SHERP molecules, caused by a decrease in affinity of the protein for the micelle, are responsible for the changes in secondary structure seen by CD spectroscopy. Is there some level of oligomerisation/association between protein molecules? Could changes in oligomerisation state due to fewer SHERP molecules per micelle as protein net charge becomes more negative explain the changes in secondary structure and micelle structure? Atomistic structures of each mutant may be able to answer these questions.

Bibliography

Abdul-Gader, Ali, Andrew John Miles, and B. A. Wallace. 2011. “A Reference Dataset for the Analyses of Membrane Protein Secondary Structures and Transmembrane Residues Using Circular Dichroism Spectroscopy.” *Bioinformatics* 27(12):1630–1636.

Adkins, Joshua N. and Kevin J. Lumb. 2002. “Intrinsic Structural Disorder and Sequence Features of the Cell Cycle Inhibitor p57Kip2.” *Proteins* 46(1):1–7.

Alder, B. J. and T. E. Wainwright. 1959. “Studies in Molecular Dynamics. I. General Method.” *The Journal of Chemical Physics* 31(2):459–466.

Altschul, Stephen F. et al. 1997. “Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs.” *Nucleic Acids Research* 25(17):3389–3402.

Anfinsen, Christian B. 1973. “Principles That Govern the Folding of Protein Chains.” *Science* 181(4096):223-230.

Apweiler, Rolf et al. 2004. “UniProt: The Universal Protein Knowledgebase.” *Nucleic Acids Research* 32(Database issue):D115-119.

Arents, G., R. W. Burlingame, B. C. Wang, W. E. Love, and E. N. Moudrianakis. 1991. “The Nucleosomal Core Histone Octamer at 3.1 A Resolution: A Tripartite Protein Assembly and a Left-Handed Superhelix.” *Proceedings of the National Academy of Sciences of the United States of America* 88(22):10148–10152.

Ashbaugh, Henry S. and Harold W. Hatch. 2008. “Natively Unfolded Protein Stability as a Coil-to-Globule Transition in Charge/hydropathy Space.” *Journal of the American Chemical Society* 130(29):9536–9542.

Barrett, Michael P. and Simon L. Croft. 2012. “Management of Trypanosomiasis and Leishmaniasis.” *British Medical Bulletin* 104(1):175–196.

Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. “GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation.” *Computer Physics Communications* 91(1–3):43–56.

Best, Robert B. et al. 2012. “Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Φ, ψ and Side-Chain χ(1) and χ(2) Dihedral Angles.” *Journal of Chemical Theory and Computation* 8(9):3257–3273.

Bienkiewicz, Ewa A., Joshua N. Adkins, and Kevin J. Lumb. 2002. “Functional Consequences of Preorganized Helical Structure in the Intrinsically Disordered Cell-Cycle Inhibitor p27 Kip1.” *Biochemistry* 41(3):752–759.

Bodner, Christina R., Christopher M. Dobson, and Ad Bax. 2009. “Multiple Tight Phospholipid-Binding Modes of Alpha-Synuclein Revealed by Solution NMR Spectroscopy.” *Journal of Molecular Biology* 390(4):775–790.

Boucher, Jonathan et al. 2004. “Apolipoprotein A-II Regulates HDL Stability and Affects Hepatic Lipase Association and Activity.” *Journal of Lipid Research* 45(5):849–858.

Brooks, B. R. et al. 1983. “CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations.” *Journal of Computational Chemistry* 4:187–217.

Brown, Celeste J. et al. 2002. “Evolutionary Rate Heterogeneity in Proteins with Long Disordered Regions.” *Journal of Molecular Evolution* 55(1):104–110.

Brown, Celeste J., Audra K. Johnson, and Gary W. Daughdrill. 2010. “Comparing Models of Evolution for Ordered and Disordered Proteins.” *Molecular Biology and Evolution* 27(3):609–621.

Campen, Andrew et al. 2008. “TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder.” *Protein and Peptide Letters* 15(9):956–963.

Casu, Fabio, Brendan M. Duggan, and Mirko Hennig. 2013. “The Arginine-Rich RNA-Binding Motif of HIV-1 Rev Is Intrinsically Disordered and Folds upon RRE Binding.” *Biophysical Journal* 105(4):1004–1017.

Chakrabarti, R. and C. E. Schutt. 2001. “The Enhancement of PCR Amplification by Low Molecular-Weight Sulfones.” *Gene* 274(1–2):293–298.

Chen, Jessica Walton, Pedro Romero, Vladimir N. Uversky, and A. Keith Dunker. 2006. “Conservation of Intrinsic Disorder in Protein Domains and Families: I. A Database of Conserved Predicted Disordered Regions.” *Journal of Proteome Research* 5(4):879–887.

Chen, Sean Chun-Chang, Feng-Chi Chen, and Wen-Hsiung Li. 2010. “Phosphorylated and Nonphosphorylated Serine and Threonine Residues Evolve at Different Rates in Mammals.” *Molecular Biology and Evolution* 27(11):2548–2454.

Cheng, Yugong, Tanguy LeGall, Christopher J. Oldfield, A. Keith Dunker, and Vladimir N. Uversky. 2006. “Abundance of Intrinsic Disorder in Protein Associated with Cardiovascular Disease.” *Biochemistry* 45(35):10448–10460.

Cornell, Wendy D. et al. 1995. “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.” *Journal of the American Chemical Society* 117(19):5179–5197.

Cumberworth, Alexander, Guillaume Lamour, M.Madan Babu, and Jörg Gsponer. 2013. “Promiscuity as a Functional Trait: Intrinsically Disordered Regions as Central Players of Interactomes.” *The Biochemical Journal* 454(3):361–369.

Das, R. K. and R. V. Pappu. 2013. “Conformations of Intrinsically Disordered Proteins Are Influenced by Linear Sequence Distributions of Oppositely Charged Residues.” *Proceedings of the National Academy of Sciences of the United States of America* 110(33):13392–13397.

Das, Rahul K., Scott L. Crick, and Rohit V. Pappu. 2011. “Intrinsic Disorder in the Basic Regions of bZIP Transcription Factors: What It Means to Be Disordered and Why It Might Matter!" *Biophysical Journal* 100(3):519a.

Davidson, W. S. 1998. "Stabilization of Alpha -Synuclein Secondary Structure upon Binding to Synthetic Membranes." *Journal of Biological Chemistry* 273(16):9443–9449.

Dayhoff, M.O., Schwartz, R.M. & Orcutt, B. C. 1978. "A Model of Evolutionary Change in Proteins." *Atlas of Protein Sequence and Structure* 5(3):345–352.

Debye, P. 1947. "Molecular-Weight Determination by Light Scattering." *The Journal of Physical and Colloid Chemistry* 51(1):18–32.

Desjeux, P. 2004. "(2004) Nat. Rev. Microbiol. 2, 692–693." *Nat. Rev. Microbiol.* 2:692–693.

Dikiy, Igor and David Eliezer. 2012. "Folding and Misfolding of Alpha-Synuclein on Membranes." *Biochimica et Biophysica Acta* 1818(4):1013–1018.

Disfani, Fatemeh M. et al. 2012. "MoRFpred, a Computational Tool for Sequence-Based Prediction and Characterization of Short Disorder-to-Order Transitioning Binding Regions in Proteins." *Bioinformatics* 28(12):75-83.

Doehl, Johannes S. P. et al. 2017. "*Leishmania* HASP and SHERP Genes Are Required for In Vivo Differentiation, Parasite Transmission and Virulence Attenuation in the Host" edited by S. M. Beverley. *PLOS Pathogens* 13(1):e1006130.

Dosztányi, Zsuzsanna, Veronika Csizmok, Peter Tompa, and István Simon. 2005. "IUPred: Web Server for the Prediction of Intrinsically Unstructured Regions of Proteins Based on Estimated Energy Content." *Bioinformatics* 21(16):3433–3434.

Dunker, A. K., Z. Obradovic, P. Romero, E. C. Garner, and C. J. Brown. 2000. "Intrinsic Protein Disorder in Complete Genomes." *Genome Informatics. Workshop on Genome Informatics* 11:161–171.

Dunker, A. Keith. et al. 2001. “Intrinsically Disordered Protein.” *Journal of Molecular Graphics and Modelling* 19(1):26–59.

Dunker, A. Keith et al. 2008. “The Unfoldomics Decade: An Update on Intrinsically Disordered Proteins.” *BMC Genomics* 9 Suppl 2(Suppl 2):S1.

Duplâtre, G., M. F.Ferreira Marques, and M.da Graça Miguel. 1996. “Size of Sodium Dodecyl Sulfate Micelles in Aqueous Solutions as Studied by Positron Annihilation Lifetime Spectroscopy.” *The Journal of Physical Chemistry* 100 (41), 16608-16612.

Dyson, H.Jan. and Peter E. Wright. 2002. “Coupling of Folding and Binding for Unstructured Proteins.” *Current Opinion in Structural Biology* 12(1):54–60.

Eiríksdóttir, Emelía, Karidia Konate, Ülo Langel, Gilles Divita, and Sébastien Deshayes. 2010. “Secondary Structure of Cell-Penetrating Peptides Controls Membrane Interaction and Insertion.” *Biochimica et Biophysica Acta - Biomembranes* 1798(6):1119–1128.

Fisher, Charles K. and Collin M. Stultz. 2011. “Constructing Ensembles for Intrinsically Disordered Proteins.” *Current Opinion in Structural Biology* 21(3):426–431.

Flagmeier P et al. (2017). "Ultrasensitive Measurement of Ca2+ Influx into Lipid Vesicles Induced by Protein Aggregates." *Angewandte Chemie* 56 (27): 7750–7754.

Förster, S., L. Apostol, and W. Bras. 2010. “Scatter : Software for the Analysis of Nano- and Mesoscale Small-Angle Scattering.” *Journal of Applied Crystallography* 43(3):639–646.

Fuxreiter, Monika, István Simon, Peter Friedrich, and Peter Tompa. 2004. “Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins.” *Journal of Molecular Biology* 338(5):1015–1026.

Fuxreiter, Monika, Peter Tompa, and István Simon. 2007. “Local Structural Disorder Imparts Plasticity on Linear Motifs.” *Bioinformatics* 23(8):950–956.

Gast, Klaus, Dietrich Zirwer, and Gregor Damaschun. 2003. “Are There Temperature-Dependent Structural Transitions in the ’Intrinsically Unstructured’ Protein Prothymosin Alpha?” *European Biophysics Journal* 31(8):586–594.

Gasteiger E. et al. 2005. “Protein Identification and Analysis Tools on the ExPASy Server”**.** *The Proteomics Protocols Handbook*, Humana Press pp. 571-607

Gazi, Anastasia D. et al. 2008. “Evidence for a Coiled-Coil Interaction Mode of Disordered Proteins from Bacterial Type III Secretion Systems.” *The Journal of Biological Chemistry* 283(49):34062–34068.

Gibbs, J. W. 1902. *Elementary Principles of Statistical Mechanics*. New York: Charles Scribner’s Sons.

Granata, Daniele et al. 2015. “The Inverted Free Energy Landscape of an Intrinsically Disordered Peptide by Simulations and Experiments.” *Scientific Reports* 5:15449.

Greenfield, Norma J. 2006. “Using Circular Dichroism Spectra to Estimate Protein Secondary Structure.” *Nature Protocols* 1(6):2876–2890.

Gsponer, Jörg, Matthias E. Futschik, Sarah A. Teichmann, and M.Madan Babu. 2008. “Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation.” *Science* 322(5906):1365–1368.

Guinier, A. and G. Fournet. 1955. “Small Angle Scattering of X-Rays.” John Wiley & Sons, New York.

Guy, Andrew J. et al. 2015. “Insights into the Immunological Properties of Intrinsically Disordered Malaria Proteins Using Proteome Scale Predictions.” *PLOS ONE* 10(10):e0141729.

Hammouda, Boualem. 2013. “Temperature Effect on the Nanostructure of SDS Micelles in Water.” *Journal of Research of the National Institute of Standards and Technology* 118:151–167.

Hansen, Jeffrey C., Xu Lu, Eric D. Ross, and Robert W. Woody. 2006. “Intrinsic Protein Disorder, Amino Acid Composition, and Histone Terminal Domains.” *Journal of Biological Chemistry* 281(4):1853–1856.

Harmon, Tyler S. et al. 2016. “GADIS: Algorithm for Designing Sequences to Achieve Target Secondary Structure Profiles of Intrinsically Disordered Proteins.” *Protein Engineering Design and Selection* 29(9):339–346.

Harvey, Anne C. and Jessica A. Downs. 2004. “What Functions Do Linker Histones Provide?” *Molecular Microbiology* 53(3):771–775.

Haynes, Chad et al. 2006. “Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes.” *PLOS Computational Biology* 2(8):e100.

Heinig, Matthias and Dmitrij Frishman. 2004. “STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins.” *Nucleic Acids Research* 32(Web server issue):W500-W502.

Henikoff, S. and J. G. Henikoff. 1992. “Amino Acid Substitution Matrices from Protein Blocks.” *Proceedings of the National Academy of Sciences of the United States of America* 89(22):10915–10919.

Holt, Carl and Lindsay Sawyer. 1993. “Caseins as Rheomorphic Proteins: Interpretation of Primary and Secondary Structures of the α_{S1}-, β- and κ-Caseins.” *Journal of the Chemical Society, Faraday Transactions* 89(15):2683-2692.

Huang, Yongqi and Zhirong Liu. 2009. “Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding–Binding Process: A Critical Assessment of the ‘Fly-Casting’ Mechanism.” *Journal of Molecular Biology* 393(5):1143–1159.

Humphrey, W., A. Dalke, and K. Schulten. 1996. “VMD: Visual Molecular Dynamics.” *Journal of Molecular Graphics* 14(1):33–38.

Iakoucheva, Lilia M., Celeste J. Brown, J.Davi. Lawson, Zoran Obradović, and A.Keith. Dunker. 2002. “Intrinsic Disorder in Cell-Signaling and Cancer-Associated Proteins.” *Journal of Molecular Biology* 323(3):573–584.

Jones, J. E. 1924. “On the Determination of Molecular Fields. II. From the Equation of State of a Gas.” *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 106(738):463-477.

Schymkowitz, J. et al. 2005. “The FoldX Web Server: An Online Force Field.” *Nucleic Acids Research* 33(Web server issue):W382–W388.

Jorgensen, William L., David S. Maxwell, and Julian Tirado-Rives. 1996. “Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids.” *Journal of the American Chemical Society* 118(45):11225–11236.

Jurrus, Elizabeth et al. 2018. “Improvements to the APBS Biomolecular Solvation Software Suite.” *Protein Science* 27(1):112–128.

Kalé, Laxmikant et al. 1999. “NAMD2: Greater Scalability for Parallel Molecular Dynamics.” *Journal of Computational Physics* 151(1):283–312.

Karush, Fred. 1950. “Heterogeneity of the Binding Sites of Bovine Serum Albumin.” *Journal of the American Chemical Society* 72(6):2705–2713.

Kelly, Sharon and Nicholas Price. 2000. “The Use of Circular Dichroism in the Investigation of Protein Structure and Function.” *Current Protein & Peptide Science* 1(4):349–384.

Killick-Kendrick, R. 1990. “The Life-Cycle of *Leishmania* in the Sandfly with Special Reference to the Form Infective to the Vertebrate Host.” *Annales de Parasitologie Humaine et Comparée* 65 Suppl 1:37–42.

Knuepfer, E., Y. D. Stierhof, P. G. McKean, and D. F. Smith. 2001. “Characterization of a

Differentially Expressed Protein That Shows an Unusual Localization to Intracellular Membranes in *Leishmania major*." *The Biochemical Journal.* 356(2):335–44.

Krishnarmurti, P. 1930. "Studies in X-Ray Diffraction. I. The Structure of Amorphous Carbon." *Indian Journal of Physics* 5(4):73–86.

Kriwacki, R. W., L. Hengst, L. Tennant, S. I. Reed, and P. E. Wright. 1996. "Structural Studies of p21Waf1/Cip1/Sdi1 in the Free and Cdk2-Bound State: Conformational Disorder Mediates Binding Diversity." *Proceedings of the National Academy of Sciences of the United States of America* 93(21):11504–11509.

Kussie, P. H. et al. 1996. "Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain." *Science* 274(5289):948–953.

Lacy, Eilyn R. et al. 2004. "p27 Binds Cyclin-CDK Complexes through a Sequential Mechanism Involving Binding-Induced Protein Folding." *Nature Structural & Molecular Biology* 11(4):358–364.

Lee, H. et al. 2000. "Local Structural Elements in the Mostly Unstructured Transcriptional Activation Domain of Human p53." *Journal of Biological Chemistry* 275(38):29426–29432.

Lee, John, Dennis J. O'Kane, and Antonie J. W. G. Visser. 1985. "Spectral Properties and Function of Two Lumazine Proteins from Photobacterium." *Biochemistry* 24(6):1476–1483.

Lees, J. G., B. R. Smith, F. Wien, A. J. Miles, and B. A. Wallace. 2004. "CDtool-an Integrated Software Package for Circular Dichroism Spectroscopic Data Processing, Analysis, and Archiving." *Analytical Biochemistry* 332(2):285–289.

Leopold, P., M. Montal, and J. Onuchic. 1992. "Protein Folding Funnels: A Kinetic Approach to the Sequence-Structure Relationship." *Proceedings of the National Academy of Sciences of the United States of America* 89(18):8721–8725.

Li, Xiaohong, Zoran Obradovic, Celeste J. Brown, Ethan C. Garner, and A. Keith Dunker. 2000. “Comparing Predictors of Disordered Protein.” *Genome Informatics Series* 172–184.

Lin, Guorong and John F. Oram. 2000. “Apolipoprotein Binding to Protruding Membrane Domains during Removal of Excess Cellular Cholesterol.” *Atherosclerosis* 149(2):359–370.

Livesay, Dennis R. 2010. “Protein Dynamics: Dancing on an Ever-Changing Free Energy Stage.” *Current Opinion in Pharmacology* 10(6):706–708.

“CD Sample Preparation.” (2014). Website retrieved August 15, 2017 (http://structbio.vanderbilt.edu/wetlab/cd.sample.prep.php).

Lopes, Jose L., Debora Orcia, Ana P. Araujo, Ricardo DeMarco, and B. A. Wallace. 2013. “Folding Factors and Partners for the Intrinsically Disordered Protein Micro-Exon Gene 14 (MEG-14).” *Biophysical Journal* 104(11):2512–2520.

Lozano, Rafael et al. 2012. “Global and Regional Mortality from 235 Causes of Death for 20 Age Groups in 1990 and 2010: A Systematic Analysis for the Global Burden of Disease Study 2010.” *Lancet* 380(9859):2095–2128.

Luger, K., A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. 1997. “Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution.” *Nature* 389(6648):251–260.

MacKerell, A. D. et al. 1998. “All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins.” *The Journal of Physical Chemistry. B* 102(18):3586–3616.

MacRaild, Christopher A., Geoffrey J. Howlett, and Paul R. Gooley. 2004. “The Structure and Interactions of Human Apolipoprotein C-II in Dodecyl Phosphocholine.” *Biochemistry* 43(25):8084–8093.

Malhis, Nawar, Matthew Jacobson, and Jörg Gsponer. 2016. “MoRFchibi SYSTEM: Software Tools for the Identification of MoRFs in Protein Sequences.” *Nucleic Acids Research* 44(Webserver issue):W488–493.

Mao, D., E. Wachter, and B. A. Wallace. 1982. “Folding of the Mitochondrial Proton Adenosinetriphosphatase Proteolipid Channel in Phospholipid Vesicles.” *Biochemistry* 21(20):4960–4968.

Marrink, Siewert J., H.Jelger Risselada, Serge Yefimov, D.Peter Tieleman, and Alex H. De Vries. 2007. “The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations.” *Journal of Physical Chemistry B* 111(27):7812–7824.

McLaughlin, Stuart and Diana Murray. 2005. “Plasma Membrane Phosphoinositide Organization by Protein Electrostatics.” *Nature* 438(7068):605–611.

Mészáros, Bálint, István Simon, Zsuzsanna Dosztányi, A. Verloop, and V. Pliska. 2009. “Prediction of Protein Binding Regions in Disordered Proteins.” *PLOS Computational Biology* 5(5):e1000376.

Miles, Andrew J. et al. 2003. “Calibration and Standardisation of Synchrotron Radiation Circular Dichroism and Conventional Circular Dichroism Spectrophotometers.” *Spectroscopy* 17(4):653–661.

Miles, Andrew J. and B. A. Wallace. 2006. “Synchrotron Radiation Circular Dichroism Spectroscopy of Proteins and Applications in Structural and Functional Genomics.” *Chemical Society Reviews* 35(1):39–51.

Minde, David P., Zeinab Anvarian, Stefan Gd Rüdiger, and Madelon M. Maurice. 2011. “Messing up Disorder: How Do Missense Mutations in the Tumor Suppressor Protein APC Lead to Cancer?” *Molecular Cancer* 10(1):101.

Minde, David P., Madelon M. Maurice, and Stefan G. D. Rüdiger. 2012. “Determining Biophysical Protein Stability in Lysates by a Fast Proteolysis Assay, FASTpp.” *PLOS ONE* 7(10):e46147.

Minor, D. L. and P. S. Kim. 1996. “Context-Dependent Secondary Structure Formation of a Designed Protein Sequence.” *Nature* 380(6576):730–734.

Mizuno, N et al. 2012. "Remodeling of lipid vesicles into cylindrical micelles by α-synuclein in an extended α-helical conformation". *Journal of Biological Chemistry* 287 (35): 29301–29311.

Moore, Benjamin et al. 2011. “Structural Basis of Molecular Recognition of the *Leishmania* Small Hydrophilic Endoplasmic Reticulum-Associated Protein (SHERP) at Membrane Surfaces.” *Journal of Biological Chemistry* 286(11):9246–9256.

Morgan, D. O. 1995. “Principles of CDK Regulation.” *Nature* 374(6518):131–134.

Müller-Späth, Sonja et al. 2010. “From the Cover: Charge Interactions Can Dominate the Dimensions of Intrinsically Disordered Proteins.” *Proceedings of the National Academy of Sciences of the United States of America* 107(33):14609–14614.

Munishkina, Larissa A., Anthony L. Fink, and Vladimir N. Uversky. 2004. “Conformational Prerequisites for Formation of Amyloid Fibrils from Histones.” *Journal of Molecular Biology* 342(4):1305–1324.

Oldfield, Christopher J. et al. 2008. “Flexible Nets: Disorder and Induced Fit in the Associations of p53 and 14-3-3 with Their Partners.” *BMC Genomics* 9 Suppl 1(Suppl 1):S1.

Onuchic, J. N., N. D. Socci, Z. Luthey-Schulten, and P. G. Wolynes. 1996. “Protein Folding Funnels: The Nature of the Transition State Ensemble.” *Folding & Design* 1(6):441–450.

Oroguchi, Tomotaka, Mitsunori Ikeguchi, and Mamoru Sato. 2011. “Towards the Structural Characterization of Intrinsically Disordered Proteins by SAXS and MD Simulation.” *Journal of Physics: Conference Series* 272(1):12005.

Ozenne, Valéry et al. 2012. “Flexible-Meccano: A Tool for the Generation of Explicit Ensemble Descriptions of Intrinsically Disordered Proteins and Their Associated Experimental

Observables.” *Bioinformatics* 28(11):1463–1470.

Pancsa, Rita, Peter Tompa, D. Dash, N. Klitgord, and ME Cusick. 2012. “Structural Disorder in Eukaryotes” edited by L. Buday. *PLOS ONE* 7(4):e34687.

Parrinello, M. and A. Rahman. 1981. “Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method.” *Journal of Applied Physics* 52(12):7182–7190.

Patriksson, Alexandra and David van der Spoel. 2008. “A Temperature Predictor for Parallel Tempering Simulations.” *Physical Chemistry Chemical Physics* 10(15):2073.

Pattanayek, Rekha et al. 2011. “Combined SAXS/EM Based Models of the S. Elongatus Post-Translational Circadian Oscillator and Its Interactions with the Output His-Kinase SasA.” *PLOS ONE* 6(8):e23697.

Permyakov, Sergei E., Ian S. Millett, Sebastian Doniach, Eugene A. Permyakov, and Vladimir N. Uversky. 2003. “Natively Unfolded C-Terminal Domain of Caldesmon Remains Substantially Unstructured after the Effective Binding to Calmodulin.” *Proteins* 53(4):855–862.

Petoukhov, Maxim V. et al. 2012. “New Developments in the ATSAS Program Package for Small-Angle Scattering Data Analysis.” *Journal of Applied Crystallography* 45(2):342–350.

Pettersen, Eric F. et al. 2004. “UCSF Chimera—A Visualization System for Exploratory Research and Analysis.” *Journal of Computational Chemistry* 25:1605–1612.

Plimpton, Steve. 1995. “Fast Parallel Algorithms for Short-Range Molecular Dynamics.” *Journal of Computational Physics* 117(1):1–19.

Provencher, S. W. and J. Glöckner. 1981. “Estimation of Globular Protein Secondary Structure from Circular Dichroism.” *Biochemistry* 20(1):33–37.

Qi, Jie, Yanru Wang, and Michael Forgac. 2007. “The Vacuolar (H+)-ATPase: Subunit Arrangement and in Vivo Regulation.” *Journal of Bioenergetics and Biomembranes* 39(5–6):423–426.

Rosenman, David J., Christopher R. Connors, Wen Chen, Chunyu Wang, and Angel E. García. 2013. “Aβ Monomers Transiently Sample Oligomer and Fibril-like Configurations: Ensemble Characterization Using a Combined MD/NMR Approach.” *Journal of Molecular Biology* 425(18):3338–3359.

Russo, A. A., P. D. Jeffrey, A. K. Patten, J. Massagué, and N. P. Pavletich. 1996. “Crystal Structure of the p27Kip1 Cyclin-Dependent-Kinase Inhibitor Bound to the Cyclin A-Cdk2 Complex.” *Nature* 382(6589):325–331.

Ruy, Patrícia et al. 2014. “Intrinsically Disordered Proteins (IDPs) in Trypanosomatids.” *BMC Genomics* 15(1):1100.

Sádlová, Jovana et al. 2010. “The Stage-Regulated HASPB and SHERP Proteins Are Essential for Differentiation of the Protozoan Parasite *Leishmania major* in Its Sand Fly Vector, Phlebotomus Papatasi.” *Cellular Microbiology* 12(12):1765–1779.

Schrödinger, LLC. 2010. *The {PyMOL} Molecular Graphics System, Version~1.3r1*.

Serpell, L. C., J. Berriman, R. Jakes, M. Goedert, and R. A. Crowther. 2000. “Fiber Diffraction of Synthetic Alpha-Synuclein Filaments Shows Amyloid-like Cross-Beta Conformation.” *Proceedings of the National Academy of Sciences of the United States of America* 97(9):4897–4902.

Shoemaker, B. A., J. J. Portman, and P. G. Wolynes. 2000. “Speeding Molecular Recognition by Using the Folding Funnel: The Fly-Casting Mechanism.” *Proceedings of the National Academy of Sciences of the United States of America* 97(16):8868–8873.

Sigalov, Alexander B., Dikran A. Aivazian, Vladimir N. Uversky, and Lawrence J. Stern. 2006. “Lipid-Binding Activity of Intrinsically Unstructured Cytoplasmic Domains of Multichain Immune Recognition Receptor Signaling Subunits.” *Biochemistry* 45(51):15731–15739.

Sigalov, Alexander B., Anastasia V. Zhuravleva, and Vladislav Yu. Orekhov. 2007. “Binding of

Intrinsically Disordered Proteins Is Not Necessarily Accompanied by a Structural Transition to a Folded Form." *Biochimie* 89(3):419–421.

Sinclair, Alison J. 2006. "Unexpected Structure of Epstein-Barr Virus Lytic Cycle Activator Zta." *Trends in Microbiology* 14(7):289–291.

Singh, Sarman and Ramu Sivakumar. 2004. "Challenges and New Discoveries in the Treatment of Leishmaniasis." *Journal of Infection and Chemotherapy : Official Journal of the Japan Society of Chemotherapy* 10(6):307–315.

Slupsky, Joseph R., Mamoru Ohnishi, Michael R. Carpenter, and Reinhart A. F. Reithmeier. 1987. "Characterization of Cardiac Calsequestrin." *Biochemistry* 26(20):6539–6544.

Spillantini, M. G. et al. 1997. "Alpha-Synuclein in Lewy Bodies." *Nature* 388(6645):839–840.

Spinozzi, Francesco, Claudio Ferrero, Maria Grazia Ortore, Alejandro De Maria Antolinos, and Paolo Mariani. 2014. " GENFIT: Software for the Analysis of Small-Angle X-Ray and Neutron Scattering Data of Macro-molecules in Solution." *Journal of Applied Crystallography* 47(Pt 3):1132–1139.

Sreerama, N. and R. W. Woody. 2000. "Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set." *Analytical Biochemistry* 287(2):252–260.

Stark, Chris et al. 2006. "BioGRID: A General Repository for Interaction Datasets." *Nucleic Acids Research* 34(Database issue):D535-D539.

Still, W.Clark, Anna Tempczyk, Ronald C. Hawley, and Thomas Hendrickson. 1990. "Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics." *Journal of the American Chemical Society* 112(16):6127–6129.

Strutt, J. W. (1899) "XXXIV. On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky" *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 47:287, 375-384.

Sugita, Yuji and Yuko Okamoto. 1999. "Replica-Exchange Molecular Dynamics Method for Protein Folding." *Chemical Physics Letters* 314(1–2):141–151.

Svergun, D. I. 1999. "Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing." *Biophysical Journal* 76(6):2879–2886.

Takeda, A. et al. 1998. "Abnormal Accumulation of NACP/alpha-Synuclein in Neurodegenerative Disorders." *The American Journal of Pathology* 152(2):367–372.

Tang, Xueming, Peter H. Koenig, and Ronald G. Larson. 2014. "Molecular Dynamics Simulations of Sodium Dodecyl Sulfate Micelles in Water-the Effect of the Force Field." *The Journal of Physical Chemistry. B* 118(14):3864–3880.

Tria, Giancarlo, Haydyn D. T. Mertens, Michael Kachala, and Dmitri I. Svergun. 2015. "Advanced Ensemble Modelling of Flexible Macromolecules Using X-Ray Solution Scattering." *International Union of Crystallography Journal* 2(2):207–217.

Trojanowski, J. Q., M. Goedert, T. Iwatsubo, and V. M. Lee. 1998. "Fatal Attractions: Abnormal Protein Aggregation and Neuron Death in Parkinson's Disease and Lewy Body Dementia." *Cell Death and Differentiation* 5(10):832–837.

Ulmschneider, Jakob P. and Martin B. Ulmschneider. 2009. "Sampling Efficiency in Explicit and Implicit Membrane Environments Studied by Peptide Folding Simulations." *Proteins: Structure, Function, and Bioinformatics* 75(3):586–597.

Uversky, V. N., J. R. Gillespie, and A. L. Fink. 2000. "Why Are 'Natively Unfolded' proteins Unstructured under Physiologic Conditions?" *Proteins* 41(3):415–427.

Uversky, Vladimir N. 2003. "A Protein-Chameleon: Conformational Plasticity of Alpha-Synuclein, a

Disordered Protein Involved in Neurodegenerative Disorders." *Journal of Biomolecular Structure & Dynamics* 21(2):211–234.

Uversky, Vladimir N. 2011. "Multitude of Binding Modes Attainable by Intrinsically Disordered Proteins: A Portrait Gallery of Disorder-Based Complexes." *Chemical Society Reviews* 40(3):1623–1634.

Uversky, Vladimir N., Jie Li, and Anthony L. Fink. 2001. "Evidence for a Partially Folded Intermediate in α-Synuclein Fibril Formation." *Journal of Biological Chemistry* 276(14):10737–10744.

Uversky, Vladimir N., Christopher J. Oldfield, and A. Keith Dunker. 2008. "Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept." *Annual Review of Biophysics* 37:215-246

van der Lee, Robin et al. 2014. "Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution." *Cell Reports* 8(6):1832–1844.

van der Spoel, David et al. 2005. "GROMACS: Fast, Flexible, and Free." *Journal of Computational Chemistry* 26(16):1701–1718.

van Stokkum, I. H., H. J. Spoelder, M. Bloemendal, R. van Grondelle, and F. C. Groen. 1990. "Estimation of Protein Secondary Structure and Error Analysis from Circular Dichroism Spectra." *Analytical Biochemistry* 191(1):110–118.

Varadi, M. and P. Tompa. 2015 "The Protein Ensemble Database." *Advances in Experimental Medicine and Biology* 870:335-349.

Vargas, Dania Alarcon, Shoichi Takahashi, and Ze'ev Ronai. 2003. "Mdm2: A Regulator of Cell Growth and Death." *Advances in Cancer Research* 89:1–34.

Varkey, J et al. 2010. "Membrane curvature induction and tubulation are common features of synucleins and apolipoproteins". *Journal of Biological Chemistry* 285 (42): 32486–32493.

Varkey, J et al. 2013. "α-Synuclein oligomers with broken helical conformation form lipoprotein nanoparticles". *Journal of Biological Chemistry* 288 (24): 17620–17630.

Verlet, Loup. 1967. "Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules." *Physical Review* 159(1):98–103.

Vila, J. A., Daniel R. Ripoll and H. A. Scheraga 2000 "Physical reasons for the unusual α-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides." *Proceedings of the National Academy of Sciences of the United States of America* 97(24):13075-13079

Vitalis, Andreas and Rohit V. Pappu. 2009. "ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions." *Journal of Computational Chemistry* 30(5):673–699.

Vucetic, S. et al. 2005. "DisProt: A Database of Protein Disorder." *Bioinformatics* 21(1):137–140.

Wallace, B. A. and Robert W. Janes. 2009. *Modern Techniques for Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy*. IOS Press.

Wang, G., J. T. Sparrow, and R. J. Cushley. 1997. "The Helix-Hinge-Helix Structural Motif in Human Apolipoprotein A-I Determined by NMR Spectroscopy." *Biochemistry* 36(44):13657–13666.

Ward, J. J., J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones. 2004a. "Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life." *Journal of Molecular Biology* 337(3):635–645

Ward, J. J., Liam J. McGuffin, Kevin Bryson, Bernard F. Buxton, and David T. Jones. 2004b. "The DISOPRED Server for the Prediction of Protein Disorder." *Bioinformatics* 20(13):2138–2139.

Whitmore L., A. Miles, L. Mavridis, R. W. Janes and B. A. Wallace. 2017. "PCDDB: new developments at the Protein Circular Dichroism Data Bank." *Nucleic Acids Research*

4;45(Database issue):D303-D307

. Whitmore, L. and B. A. Wallace. 2004. “DICHROWEB, an Online Server for Protein Secondary Structure Analyses from Circular Dichroism Spectroscopic Data.” *Nucleic Acids Research* 32(Web server issue):W668-W673.

Whitmore, L. and B. A. Wallace. 2008. “Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases.” *Biopolymers* 89(5):392–400.

Wilke, Claus O. and D.Allan Drummond. 2010. “Signatures of Protein Biophysics in Coding Sequence Evolution.” *Current Opinion in Structural Biology* 20(3):385–389.

Xia, Yu, Eric A. Franzosa, and Mark B. Gerstein. 2009. “Integrated Assessment of Genomic Correlates of Protein Evolutionary Rate” edited by M. Levitt. *PLOS Computational Biology* 5(6):e1000413.

Xue, Bin, Roland L. Dunbrack, Robert W. Williams, A. Keith Dunker, and Vladimir N. Uversky. 2010. “PONDR-FIT: A Meta-Predictor of Intrinsically Disordered Amino Acids.” *Biochimica et Biophysica Acta* 1804(4):996–1010.

Yamada, Justin et al. 2010. “A Bimodal Distribution of Two Distinct Categories of Intrinsically Disordered Structures with Separate Functions in FG Nucleoporins.” *Molecular & Cellular Proteomics : MCP* 9(10):2205–2224.

Yeung, Tony et al. 2008. “Membrane Phosphatidylserine Regulates Surface Charge and Protein Localization.” *Science* 319(5860):210–213.