UNDERSTANDING THE MECHANISM OF PROTEIN AGGREGATION IN THERMAL AND REFOLDING STUDIES

A thesis submitted to the University of Huddersfield in partial fulfilment of

the requirements for the degree of Doctor of Philosophy

JAMES IAN AUSTERBERRY

The University of Huddersfield

April 2013

Copyright

Acknowledgements

Abstract

Biopharmaceutical production of protein for therapeutic use is an expanding practice for treatment of numerous diseases in medicine. However the full benefits of this technique have not yet been fully realised due to a number of barriers. The largest of these barriers is that of protein aggregation, where mis-folded protein monomers self associate to form non functional macromolecules; aggregates. Further understanding of protein aggregation may lead to an improvement in the effectiveness and availability of these therapeutic treatments. Here is presented work which utilises novel or under-used techniques to elucidate information on the structure of protein aggregates, their formation mechanisms and the kinetics and thermodynamics of their growth.

Results presented in the chapter on aggregate nucleation indicate that the nucleation stage of aggregation in bovine serum albumin has a temperature dependant mechanism, which in the middle of the temperature range follow mechanisms for stable nuclei population postulated in the literature. However at the extremes of this range, it appears that the nucleation mechanisms deviate from this and that there may be clustering of highly reactive nuclei at high temperatures, and continual formation of aggregate nuclei at low temperatures. Possible explanations for this behaviour are discussed.

Analysis presented on the growth of particulate aggregates show that the model of monomer addition to the aggregate nuclei appears to be a fitting description of the growth process, which is generic across proteins. Furthermore the detailed analysis from an ultra violet light scattering spectroscopy technique provides a numerical method for examining the efficiency of aggregate preventing additives, and also illustrates the mechanism by which the additives prevent aggregation through stabilising the native state.

Finally; results presented in the chapter on aggregation during refolding indicate the use of fluorescence anisotropy to monitor the molten globule state during refolding of proteins. Most strikingly, it is shown there is an obvious relationship between the mobility period of the protein and its propensity to aggregate. It is also shown that the presence of salt and urea can be utilised to moderate the presence of the molten globule state, and therefore the resultant aggregation.

Abbreviations

ANS	1-anilinonaphthalene-8-sulfonic acid
AUC	Analytical ultracentrifugation
BSA	Bovine serum albumin
DLS	Dynamic light scattering
FFF	Field flow fractionation
FRET	Förster resonance energy transfer
FTIR	Fourier transform infrared spectroscopy
GFP	Green fluorescent protein
SANS	Small angle neutron scattering
SAXS	Small angle x-ray scattering
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SEM	Scanning electron microscopy
SLD	Scattering length density
SLS	Static light scattering
UVLSS	Ultra violet light scattering spectroscopy
VADAR	Volume, Area, Dihedral Angle Reporter

Contents

List of Figures

17

List of Tables

Chapter 1 : **Introduction**

Biopharmaceutical production of protein for therapeutic use is an expanding practice for treatment of numerous diseases in medicine. Such methods are common in the treatment of diabetes; where recombinant insulin is used for treatment. However the full benefits of this technique have not yet been fully realised in areas such as monoclonal antibody, and vaccine antigen production, as there are a number of barriers present to commercial production. The largest of these barriers is that of protein aggregation, where mis-folded protein monomers self associate to form non functional macromolecules; aggregates. The presence of these aggregates not only represents a loss of yield (and the cost and further processing in removing them), but their presence in therapeutics can trigger an immunogenetic response; producing anti-drug antibodies which break down the drug and render the treatment ineffective. Further understanding of protein aggregation may lead to an improvement in the effectiveness and availability of these therapeutic treatments.

The current literature regarding protein aggregation describes a number of different aggregate morphologies and their structures; it also postulates multiple mechanisms for their creation and propagation. However, the literature on protein aggregation is not complete, and a number of areas merit further investigation including some structural aspects of proteins including kinetics and thermodynamics of the aggregation systems. This body of work utilises novel or under-used techniques to elucidate information on the structure of protein aggregates, their formation mechanisms and the kinetics and thermodynamics of their growth. The information provided by these techniques are compared and contrasted with more frequently used techniques in order to validate the conclusions made.

Separate chapters focus on the class of aggregate with a spherical morphology; the particulate aggregate, where two aspects of its formation are examined. The nucleation of particulate aggregates is examined using small angle neutron scattering to probe the system on a nanometer scale and to understand the mechanisms of its formation and its temperature dependant nature. Complementing this, a second chapter details the growth of these nuclei using ultra violet light scattering spectroscopy, which can be used to give numerical values regarding the kinetics and thermodynamics of the aggregate growth. Providing numerical values for aggregation, this technique can then be used to test the effectiveness of aggregate preventing additive and elucidate the mechanism through which prevention occurs.

Work here also focuses on the mobility of proteins during refolding, using fluorescence anisotropy; a rarely utilised technique in relation to aggregation and attempts to understand what relationship may exist, if any between a proteins mobility and its propensity to aggregate.

The following chapters detail the body of work which uses these techniques to contribute to the literature of protein aggregation and examines methods of reducing its occurrence.

Chapter 2 **: Literature Review**

2.1 Protein Structure

There is a large amount of literature that covers the topic of protein structure [1, 2]. Proteins are responsible for the vast array of biological functions seen in all organisms. Functions range from large scale enzymatic action and breakdown of substrates, to delicate and complex environmental regulation. This range of actions is performed by molecules which consist of a very similar basic structure, but slight variances in this structure confer large differences in their function.

Proteins are long polypeptide molecules that consist of a chain of multiple amino acids. This number of amino acids can vary between tens; in the case of peptides, to tens of thousands; the largest being the muscle protein Titin at over 34000 residues [3]. The amino acids themselves contain an identical backbone structure consisting of one amino and one carboxy terminal, linked by a carbon (**Error! Reference source not found.**). **Error! Reference source not found.**Attached to this carbon group is the side (R) group, which varies between different amino acids, to provide each with an individual structure. It is these amino acids, and the sequence in which they are combined, which determines the structure, and thus the function of the protein.
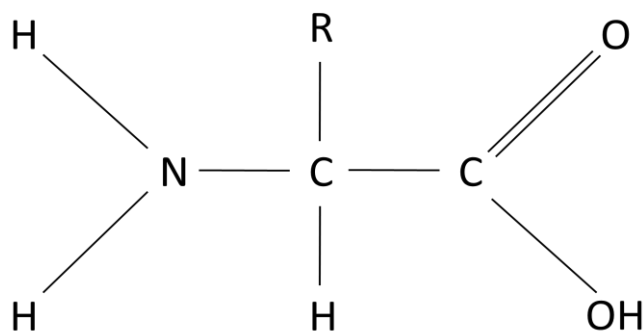
Figure 2.1 A typical amino acid structure. R represents the functional group of the amino acid.
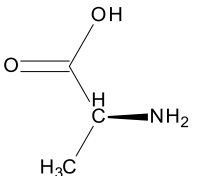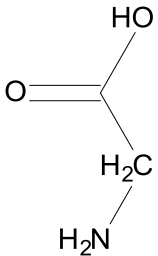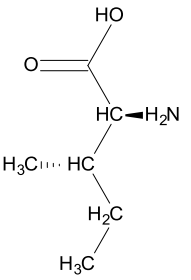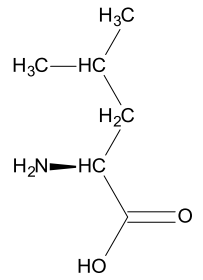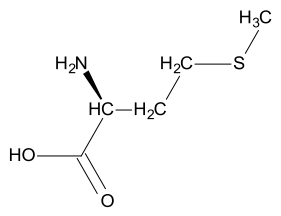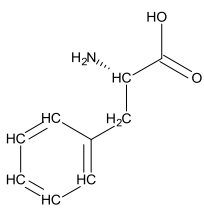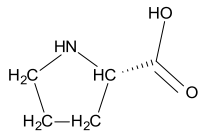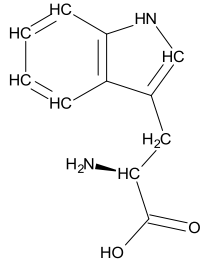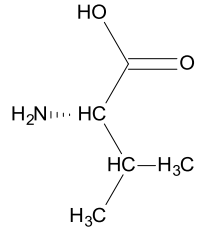
2.1.1 Primary Structure

There are 21 amino acids present in eukaryotes. Each of these amino acids consists of a backbone structure which is identical to all amino acids, and the variable R group. The different R groups confer different properties to the amino acid. These properties include; hydrophobicity, hydrophilicity, the ability to pack tightly and disulphide bridge forming capability. The full list of amino acids are pictured (Table 2.1) are arranged into groups according to some of these properties. The primary structure of a protein corresponds to the order in which these amino acids are present.

The amino acid chain is created by linkage of the individual amino acids through formation of a peptide bond between the amino and carboxy groups on neighbouring amino acids (Figure 2.2). This is a planar bond; due to the partial double bond properties, which allows for the variable R groups to protrude from the side of the amino acid, allowing them opportunity to interact with similar R groups at distant locales on the peptide chain.

1. Berg, J.M., J.L. Tymoczko, and L. Stryer, *3.6 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure*, in *Biochemistry*. 2002.
2. Alberts, B., et al., *Molecular Biology of the Cell*. 4th ed. 2002: Garland Science.
3. Berman, H., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.

Table 2.1:Amino acid structures grouped by properties.

Hydrophobic:				
Alanine	Glycine	Isoleucine	Leucine	Methionine
Phenylalanine	Proline	Tryptophan	Valine	

Hydrophilic				
Asparagine	Cysteine	Glutamine	Serine	Threonine
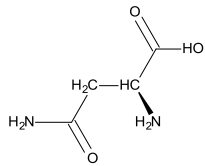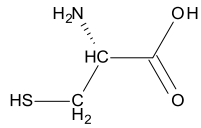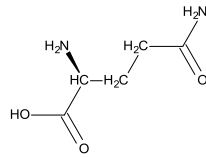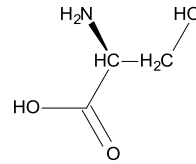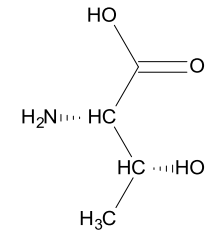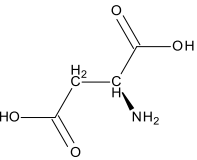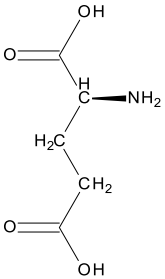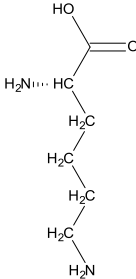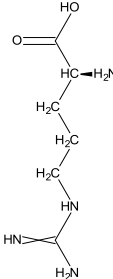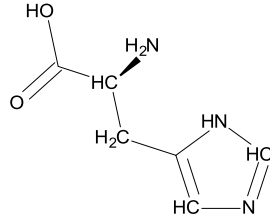
Selenocysteine	Tyrosine

28
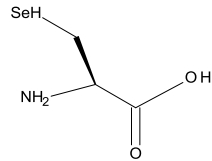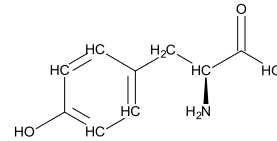
Negatively Charged		Positively Charged		
Aspartic Acid	Glutamic Acid	Lysine	Arginine	Histidine
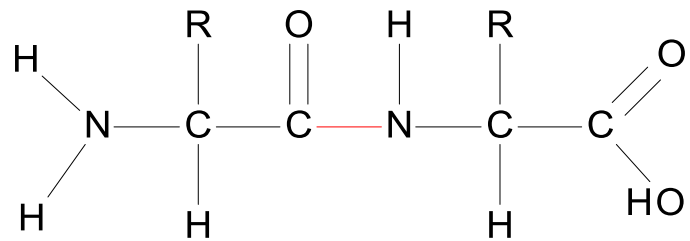
Figure 2.2: Amino acid linkage. The peptide bond is highlighted in red.

2.1.2 Secondary Structure

In addition to R group interactions, the protein chain also has possibility for hydrogen bonding. The two moieties either side of the peptide bond are capable of forming hydrogen bonds in both directions (Figure 2.3). Although these are relatively weak interactions, a large number of these non-covalent bonds can confer strength to the protein structure. Instances of such large scale hydrogen bonding frameworks are α-helices and β-sheets, which are prime examples of protein secondary structures .The α-helix is characterised by a protein chain completing a full 360 ^{o} spiral every 3.6 amino acid residues (Figure 2.3). This is the most common helical structure and forms a right handed turn. Helices can range in size from under 10 residues to over 50.

This structure commands a large degree of flexibility from the constituent amino acids, as the prerequisite backbone torsion angles between the neighbouring carbons are very specific, which only certain combinations of residues are capable of filling. Amino acids that are flexible such as glycine and alanine are often present in α-helices, where as proline would only ever be present at the terminus of the structure. Hydrogen bonding occurs between every 4th amino acid, which are oriented so that the side R groups face outward, as inward pointing side groups would be unlikely due to steric hindrance. Across the length of the α-helix, there will be multiple side chain groups pointing out of the same side. This allows for a large scale area, or side, of the protein structure to have a property such as hydrophobicity if all corresponding residues carry the same characteristic. This then allows for the whole region of the helix itself to have this property, enabling it to fulfil structural duties on a larger scale such as embedding in a membrane.
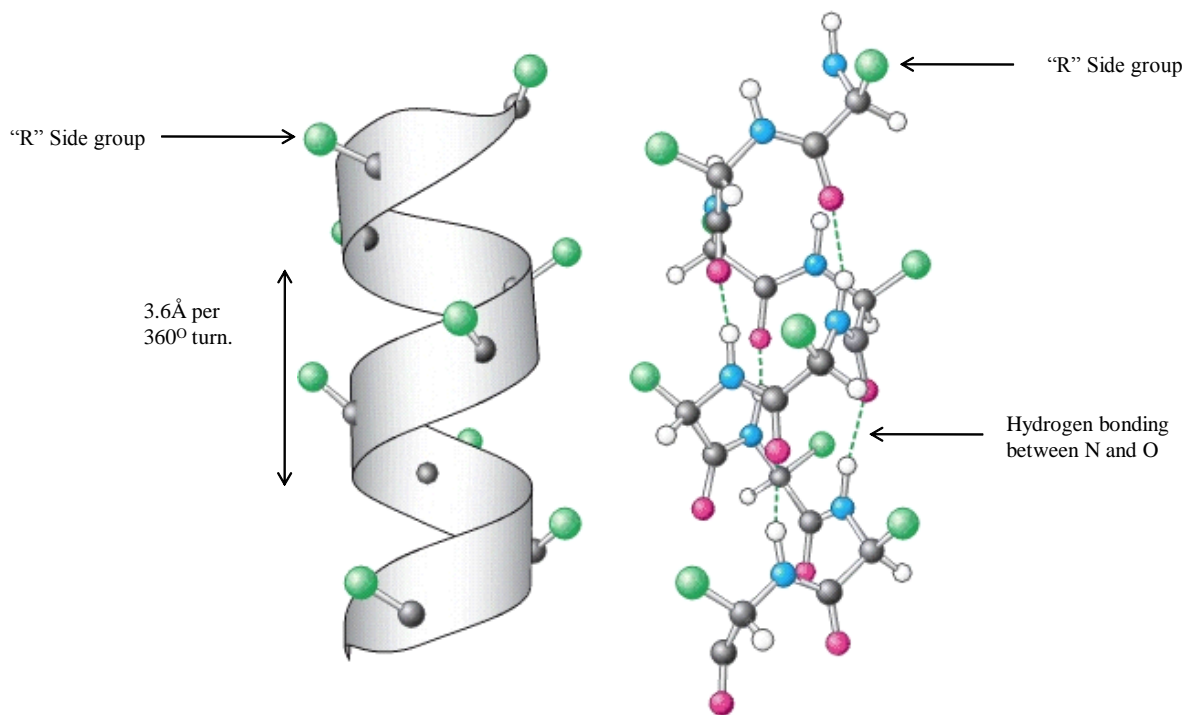
Figure 2.3: The structure of the α-helix. Adapted from [1].

A second most common example of the secondary structure of proteins is the β-sheet Figure 2.4, Figure 2.5). This consists of a helix-like strand of amino acids, with two residues for every turn. This allows for the residues to hydrogen bond to a second section of the protein chain that is running parallel to it. This second chain is also capable of hydrogen bonding to a third, and this behaviour can continue, with 2 to over 10 strands linked together in order to form a sheet like structure. There are two varieties of β-sheet, parallel and antiparallel; these names describe the orientations of the neighbouring strands; parallel sheets have identical atoms opposite each other (Figure 2.4), whilst in the antiparallel strands, the atoms opposite each other are non identical (Figure 2.5). Due to the orientations of the bonds, the parallel β-sheets are also less stable than their antiparallel counterparts. Similarly to the α-helices, there are some amino acids which are more commonly found in β-sheets than others. Prime examples of these amino acids in β-sheets include isoleucine, tyrosine and valine.

Hydrogen bonding
between N and O

"R" side groups
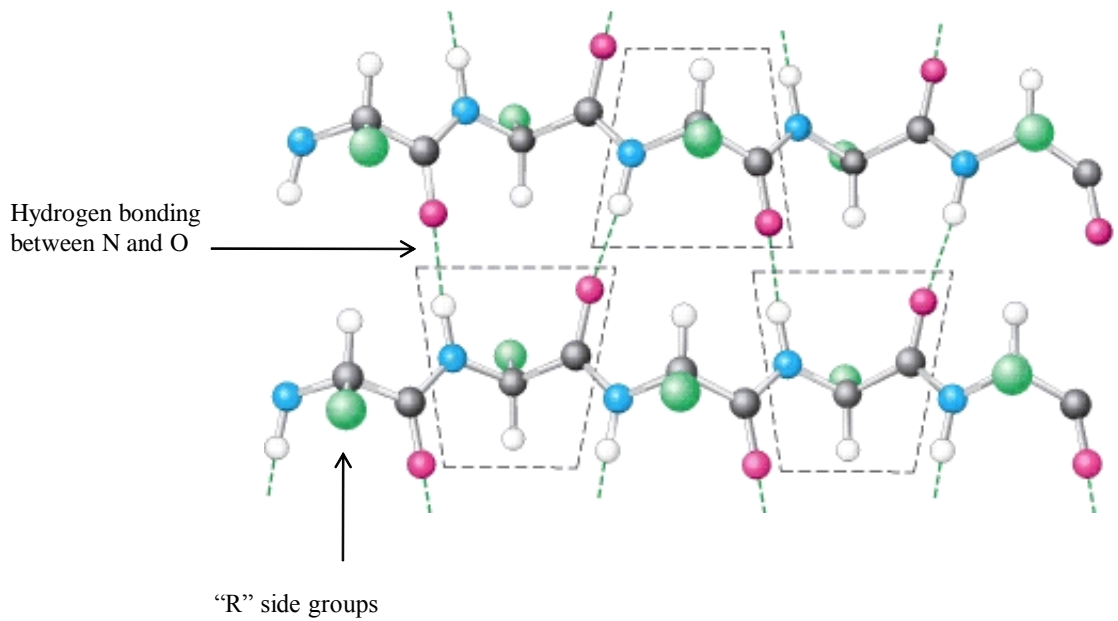
Figure 2.4: Parallel β-sheet structure. Atoms are colour coded: Carbon – black, Nitrogen – blue, Oxygen – pink, Hydrogen - white. Adapted from [1].

Hydrogen bonding
between N and O

"R" Side groups
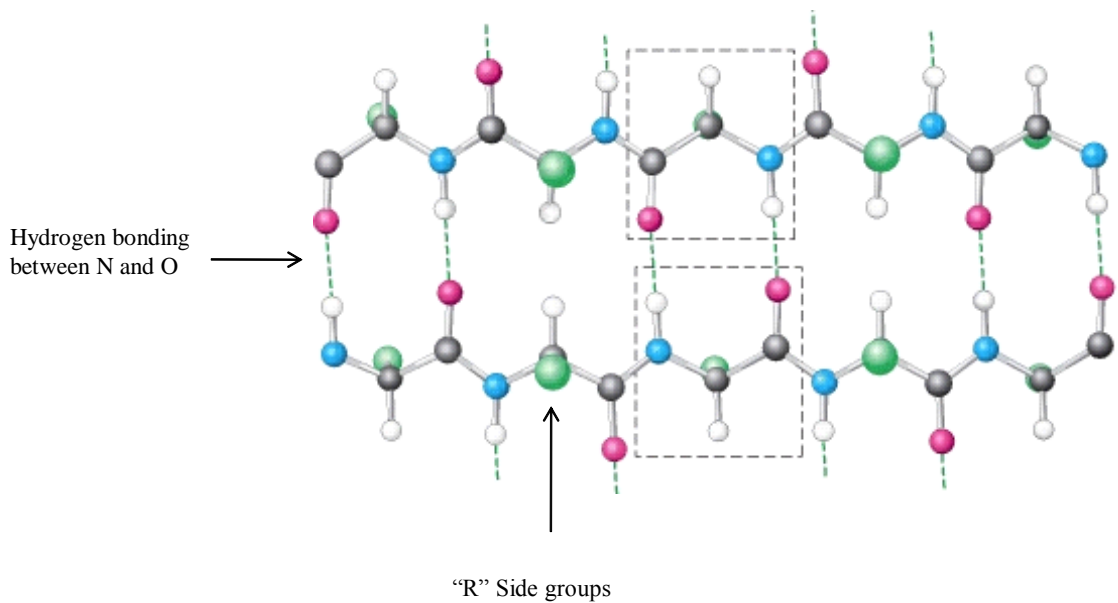
Figure 2.5:Antiparallel β-sheet structure. Carbon – black, Nitrogen – blue, Oxygen – pink, Hydrogen- white. Adapted from [1]

Although considered a sheet like structure, large β-sheets incorporate a right handed turn which is caused by the presence of the R group side groups. In order for these functional groups to achieve their most energetically favourable positions, the backbone must contort, resulting in the turn. This property allows them to form large β-barrel structures. An example of such a secondary structure feature can be seen in green

fluorescent protein (Figure 2.6). Here the beta sheet structure forms a protective barrel which allows the chromophore within the cylinder to fluoresce. Similarly to α-helices, the protruding side groups on β-barrels can confer properties such as hydrophobicity to the structure. Such a property would enable the β-barrel to be anchored into a membrane in order to form a transmembrane pore, which is a common utilisation of the β-barrel.
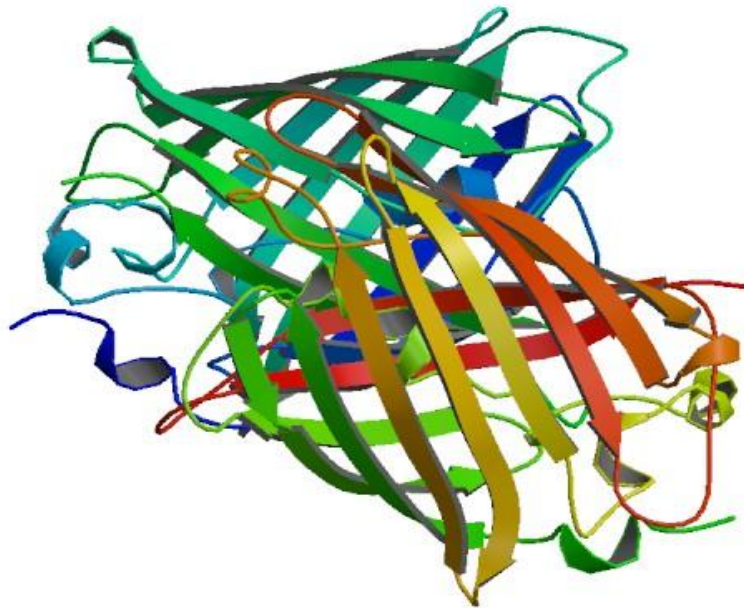
Figure 2.6: β-barrel structure of green fluorescent protein, protein data bank code: 1GFL [1].

These make up roughly 50 % of the total structure elements. Other elements include 3_{10}-helix, π-helix and the β-turn. These structural elements perform different roles within the protein structure, but are in essence very similar chemically. It is the order in which these chemical groups are arranged which determine these varying structures, and it is these structures present within a protein that contributes to its function.

2.1.3 Tertiary and Quaternary Structure

The tertiary structure represents the arrangements of the secondary structural elements in three dimensional space. These secondary structural elements may be brought together by hydrophobic nature of their respective surface residues, or long ranging disulphide bonds through distant cysteine residues. The interactions between the thiol moieties (Table 2.1) on the amino acid R groups can be long ranging. They are more prevalent in larger proteins, and are responsible for bringing structures from distant parts of a protein molecule closer together. An example of this is lysozyme (Figure 2.7) where both α-helices and β-sheets contain hydrophobic residues which congregate at the proteins core. In addition to this, there exists three disulphide bonds which bind the α-helices to the outlying polypeptide chain.
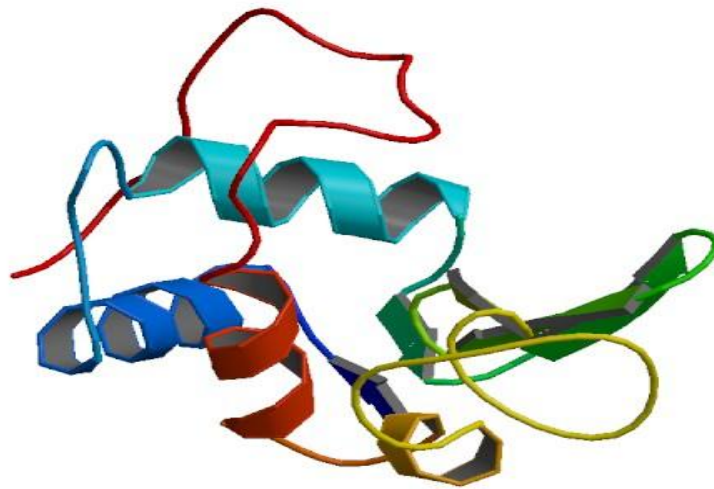
Figure 2.7: Structure of lysozyme from protein data bank code 2LYZ [1].

Quaternary structure is similar to tertiary in how the structural elements interact, but differs by the fact that quaternary structure is an interaction between separate

polypeptide chains to create the overall protein structure. This may be the assembly of a number of identical sub units to create a tetramer, or the association of separate protein subunits to create the overall globular protein. An example of such quaternary structure is bacteriorhodopsin, where 3 α-helical subunits associate to form the trans-membrane protein (Figure 2.8). The arrangement of the structure of the protein, which is primarily determined by the proteins amino acid sequence, enables it to perform its function as a protein pump which fuels ATP synthesis in bacteria.

Figure 2.8: Structure of bacteriorhodopsin containing 3 subunits. Protein data bank code 1FBB [1] .

.

2.2 Protein Folding

The general processes and mechanisms behind protein folding are well understood [2, 3]. As the previous section has shown, a protein's structure depends on the primary sequence of amino acids present within the chains. As such the primary sequence is also responsible for the folding of the protein to achieve the final protein structure. Understanding of the topic of protein folding has wide ranging implications from understanding the nature of diseases to further improving therapeutic protein production. Many *in vitro* experiments have been conducted in order to better understand the folding mechanisms of individual protein types and to improve understanding of their general formation [4-6].

2.2.1 Protein Folding *in vivo*

For a protein to fold properly as it emerges from the ribosome it must form a native, functional state. This native state essentially requires all the backbone torsion angles of the residues within the protein chain to achieve their individual set values. If this were to occur by the protein randomly adopting different states until it reached the native state, the time for a protein to fold would be impossibly large, even given the most generous estimations for adopting one state. This is known as Levinthal's paradox [7].

The specificity for protein structure is held within the primary structure; the amino acid sequence. This folding is driven through conformational entropy, the presence of hydrophobic and charged residues force the collapse of the protein structure in order to exclude water and achieve more energetically favourable states. The native state of the protein has a lower free energy than the unfolded state, and this is achieved through burying the hydrophobic residues within the protein core, and presenting hydrophilic R groups to the surface [2].

The hydrophobic driving force can be described in terms of entropy. The unfolded protein presents a large amount of hydrophobic residues to the surrounding water molecules. This does not allow opportunities for hydrogen bonding, and so the water molecules have a low degree of freedom; they are ordered and have low entropy. The folded protein presents only hydrophilic residues, these are polar and allow for hydrogen bonding to take place with the surrounding water molecules [8]. This enables the water to have a higher degree of freedom and thus higher entropy [9].

There also exists some opposing entropic forces to this folding, as the folded protein itself is more ordered than the unfolded protein. This leads to a negative change in entropy during folding. However, this entropy change is minute compared with the vast positive entropy change seen in from the exclusion of water from the proteins core, hence folding occurs [9].

In vivo, protein folding is assisted by molecules called chaperones [10]. Chaperones such as the hsp70 proteins are small molecules which bind to the unfolded hydrophobic areas of protein as it emerges from the ribosome [11]. This binding prevents regions of the protein from associating with themselves to form a non-native state before the full sequence of the peptide chain is produced. Additionally, the chaperone binds to, and releases, hydrophobic peptide regions which cause local unfolding of the protein, allowing it to overcome kinetic obstacles to achieve the native state. There also exist larger chaperones, known as hsp60-like proteins [12], which are larger and are used to facilitate the folding of mis-folded proteins by enveloping their entire structure. This prevents interactions with other proteins which may lead to aggregation, whilst also binding and releasing the protein until the natively folded structure is formed [13].

The structural formation described here occurs on the millisecond time scale [14], and as a result the protein is in a near-native state [15]. Following this, the protein may undergo more time consuming modifications in order to reach its native state. An example of this involves the packing of side chains [16]; the near-native protein exists in a molten globule state, where secondary structure is formed, but full tertiary structure is not. The remaining exposed side chains pack into energetically favourable positions in order to complete the tertiary structure [17]. This process occurs over a larger timescale than the first folding stages and can take some seconds to complete folding. A further example of large time scale structural modifications is proline isomerisation. A small percentage of proline residues exist in a cis orientation, whereas the majority of the amino acids in the peptide chain exist in the trans form. In order to amend this for the protein to form its native state, geometric isomerisation is performed by an peptidyl-prolyl-isomerase enzyme [18]. This is a process that occurs in the late stage of protein folding, much like the formation of disulphide bridges that finalises the native structure of the protein. These processes finalise the structure of the protein enabling its functionality.

2.2.2 The Molten Globule State

The general identifying features of the molten globule structure of a protein have been stated as follows: a moderately compact structure (much more so than the unfolded state, but less compact that the native), and mobile protein side chains in partially structured conformations [19]. The molten globule state of proteins proves difficult to characterise in its entirety, due to its changing structure and short existence period. There has been much research into the molten globule due to its important role in protein folding and the implications this holds for many research areas into protein structure and function [20]. The molten globule state can be identified as a statistically

sufficient number of the protein population which are in a stable state, as opposed to there being a large number of very unstable intermediates [21].
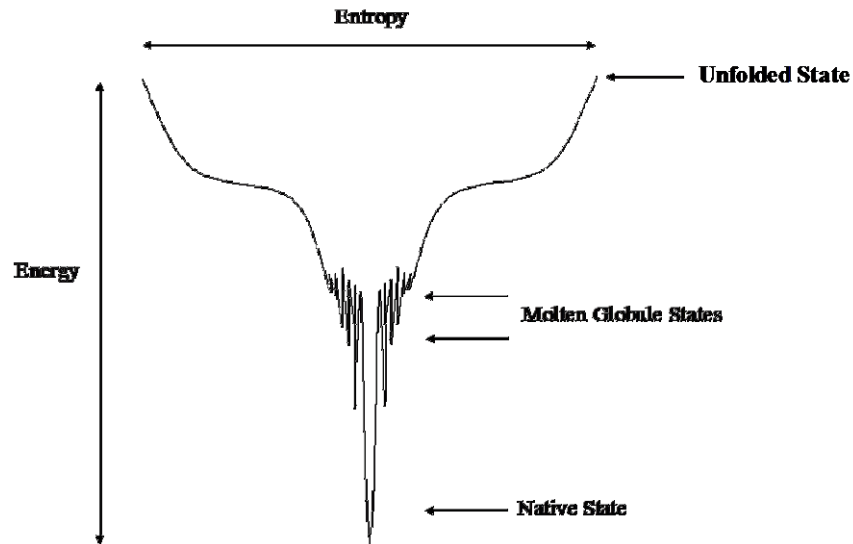
Figure 2.9: The protein folding funnel. Based on [9].

Elsewhere, from the researcher who coined the term, it is described as a compactness and presence of secondary structure, with no rigid tertiary structure and a little expanded in comparison to the native state [22]. The overall concept of the molten globule state is one whose secondary structure content varies little from the native state [23], but whose tertiary structure is fluid as the side chain positions are mobile [24].

In order to better illustrate the careful energy barriers which the molten globule state, and indeed the native state occupies, protein energy basin diagrams have been adopted. The energy basin illustrates protein folding in the form of a landscape, where instead of a small number of intermediates present between unfolded and native, there are a multitude of conformations that can be adopted in the path to the native state [25]. The energy basin diagrams show the double stage folding process and illustrate the energy required to overcome undesirable non-native states of the protein [26].

The initial hydrophobic collapse of the unfolded protein is represented by the smooth slope with a downward gradient. This is the first stage within Figure 2.9, where the intermediate states are more stable than the unfolded state. This has been proved in both acid unfolding experiments [27], and computer simulations [28]. The second stage represented is the molten globule state, with multiple metastable states that have energy barriers that must be overcome to reach the native, most energetically favourable state. *In vivo* this stage is assisted by chaperones.

Without sufficient energy, the protein will remain trapped in one of these local minima in a misfolded, or non-native state. Within the cell, presence of these misfolded proteins can carry out alternate physiological actions than their intended function, such as cytochrome C; where a misfolded conformation of the protein is capable of causing apoptosis in the cell [29]. More serious implications of the presence of non-native proteins within the cell include the formation of amyloid fibrils, which are responsible for diseases such as Alzheimer's, Parkinson's and Creutzfeldt–Jakob disease (CJD). This form of protein aggregate has been shown to form from molten globule intermediates [30], even to the extent where there is a large change in the secondary structure content of a protein from the native state [31] and these proteins have been shown capable of inducing a change to the non-native state from neighbouring proteins, increasing the proliferation of the non-native population [32]. This illustrates the importance of work into the molten globule state and protein folding pathways, as research shows that the native state has competition from protein aggregates as the most thermodynamically favourable minima on the folding landscape [33, 34]. The topic of protein aggregation will be discussed in detail further in this chapter.

2.2.3 *In vitro* protein folding

The vast majority of *in vitro* folding research focuses on recombinant proteins. Procedures for recombinant production of proteins focus on two areas: obtaining the highest yield of native protein structure and ensuring this sample is solely the native protein. In most cases the end product is the goal, rather than understanding the mechanisms that help achieve it. Recombinant protein produced from bacterial cells is in such a high concentration that it forms large aggregated masses, known as inclusion bodies [35]. These bodies are unfolded using high levels of denaturant and refolded by dilution or dialysis in order to achieve the native state [36]. Additives may be added to the refolding conditions in order to assist the protein in achieving the native state, such as metal ions around which the protein may fold. These conditions will differ depending on the protein. The refolded sample will not consist solely of the native state protein, and some aggregates will remain, therefore separation procedures such as ion exchange and size exclusion chromatography are required in order to achieve a sample of protein monomer [37]. Native structure determination can be performed by techniques such as circular dichroism, NMR and protein crystallography in order to determine that the native state of the protein has been formed [38, 39].

2.2.4 Protein Folding Research

Early protein folding research concentrated on some simple model systems. Proteins such as lysozyme and α-lactalbumin were often used in such experiments as they were able to be purified in large quantities for experimentation, and are simple α-helical proteins, which allowed the kinetics of their folding and secondary structure formation to be investigated with ease [40]. Many of the protein folding properties discussed previously in this chapter were elucidated from work on these proteins [41]. Additionally the small size of the proteins allowed them to be studied without the need for multiple kinetics rates for separate domain folding [42]. These model systems have been used in more recent work for probing the molten globule state, in the case of lysozyme its molten globule state has been shown to have a low affinity for calcium metal ions [43] which may affect its propensity to aggregate. A further example of these well characterised model systems is β-lactoglobulin, with similar characteristics to lysozyme and α-lactalbumin, the timescale of hydrophobic collapse has been elucidated for this protein [44]. Here the authors show that it takes only 100ms for the protein to collapse to 1.1 times its native size, and similar times are assumed for the timescale of folding for other model proteins. Further work on protein folding kinetics has extended to the use of computer simulations to investigate what factors contribute to the timescale of protein folding, such as the impact of the sequence properties on the speed of protein folding [45]. The history and ease of use of these proteins make them ideal for further studies on generic properties in the protein folding process.

In addition to assigning a timescale to the process of protein folding, much work has been performed to elucidate the mechanisms through which proteins fold and unfold. Examples include using both extreme temperature and pH in order to induce a molten globule like state from proteins in order to study the molten globule state [46]. The

rapid nature of the native folding process is too quick for the majority of characterising techniques, therefore techniques involving sub zero temperatures have been used to slow down the process of protein folding to better understand the mechanism by which β-lactoglobulin folds [47]. Complimentary to this, unfolding studies have been performed to sequentially break the bonds forming the structure of the native protein in the assumption that this action is the reverse of the mechanism of protein folding. Increasing concentrations of denaturant are added to the protein, and the binding energies of the protein backbone are elucidated [48] which covers a larger timescale than the folding process. This technique has also been used alongside size exclusion chromatography in order to quantify the sizes of protein molecules in multiple disordered molten globule states [49].

Most recent work in the area of protein folding has centred on the role protein folding plays in the formation of protein aggregate formation. It has been shown that the folding of lysozyme in a crowded environment results in the formation of lysozyme aggregates which lack the enzymatic activity of the native state [50]. Experimental methods have also been developed in order to monitor the extent of the aggregation during folding using fluorescent probes [51], and also into methods which can be used to reverse the aggregation process resulting from folding such as the use of denaturants and high pressure to return the protein to the native state [52], which has been shown possible for aggregates of a number of different proteins. A large focus of the work centres on prevention of formation of aggregates. If mild concentrations of denaturant are present within the refolding environment, the extent of aggregation is lower than if there is no denaturant present. This is thought to occur by stabilising the aggregation prone molten globule state by the denaturant [53]. Although fairly comprehensive, there is little in the literature which relates the mobility of the molten globule state to the likelihood of

aggregation, especially involving real time studies which are more informative on real life processes.

A large amount of work on preventing aggregation involves studying the effect chaperones have on protein folding, as in the cell they are known to facilitate this process [54]. Work has been performed to look at the interactions between protein (α-lactalbumin) and chaperone (GroEL) [55] and how they bind. Also investigating the effect of ATP dependence between a chaperone and protein [56], which shows how cycles of ATP binding are responsible for the facilitated folding of α-lactalbumin. Further to this, additional studies have monitored the states of the protein inside the chaperone, in order to understand the conformational changes that the protein undergoes whilst in contact with the chaperone [57], the chaperone appears to be responsible for binding to aggregation prone regions of the protein, and releasing them when the environment is suitable enough for them to fold into the stable native state in order to prevent aggregation [58]. Understanding the role chaperones play and their interactions with the various states of the protein can elucidate routes through which proteins interact to aggregate and therefore offer a route to prevention of aggregation.

2.2.5 Techniques for Studying Protein Folding

One of the most elementary techniques used to study the folding of proteins is tryptophan fluorescence. Tryptophan amino acid residues fluoresce under ultra violet light when they are exposed, such as when the protein is unfolded, but when the protein has collapsed and the tryptophan residues are buried and there is a significant difference in the fluorescence signal of the residue. These properties of tryptophan fluorescence have long been exploited as a method for determining the extent of protein folding [59]. Tryptophan fluorescence is also used extensively in the characterisation of protein

folding pathways, prime examples include α-lactalbumin and lysozme folding [60]. External effects on the extent of protein folding have also been probed using this technique, such as the effects of temperature and pH [46], or the effect of additives such as urea to the protein environment [53]. On its own, this technique is very qualitative and is often run to complement other techniques, such as those mentioned below [46].

Circular dichroism is a technique which allows determination of the secondary structure of a protein [40]. In terms of studying protein folding, research has attempted to study the sequences of secondary structure formation in specific proteins [61]. This technique however is not as capable at performing real time dynamic experiments, and requires the stages of folding to be slowed down, such as step-wise unfolding processes [62], or alternatively performing the refolding experiments at sub zero temperatures [47]. Examples of the techniques used include secondary structure characterisation of β-lactoglobulin [63] and cytochrome C [64], whilst more recently it has been used in conformational studies of recombinant therapeutic proteins [65]. Despite these restrictions, circular dichroism is an invaluable tool for understanding the state of protein secondary structure, and is often used in tandem with other techniques to verify the quality of the protein fold [60]. Additionally, synchrotron radiation circular dichroism is an improved technique which allows for further structural features to be characterised, and is capable of providing dynamic structural information [66].

A third commonly used method for monitoring protein folding is the use of 1-anilinonaphthalene-8-sulfonic acid (ANS) in folding experiments. This chemical binds to available hydrophobic regions on a protein and fluoresces. As the protein collapses, the ANS becomes unbound, and the fluorescence is lost. These features have been exploited in order to monitor the hydrophobic collapse of protein folding systems [62]. This method is ideal for looking at dynamic folding of a protein, and has often been

used to elucidate the molten globule state [53]. In addition to its more common use, ANS has also been shown to be capable of binding to protein aggregates formed from folding experiments, to give an indication of the aggregate quantity being produced [51]. Whilst useful in characterising hydrophobic collapse, the absence of ANS fluorescence does not indicate the absence of folding, and requires complimentary techniques to provide a full picture of folding.

In addition to these common techniques listed above, less orthodox methods have been used to complement these works. In order to better understand the contribution of different residues to a proteins fold, proteolysis experiments have been carried out where specific residues are cleaved from the protein, and the kinetics of the novel protein is then compared to that of the native [67]. Similar experiments have been undertaken where mutagenesis is used; selected residues within a protein are mutated and their contribution to the fold measured [68]. In addition, NMR has also been shown to be a capable tool for monitoring the structure of both the protein and its molten globule counterpart [46, 60]. Some techniques which are capable of monitoring protein monomer size have also been utilised to observe the hydrophobic collapse of the protein. Examples include small angle X-ray scattering [44], SDS page [61] and size exclusion chromatography has also been used for monitoring the unfolding states of proteins [49] where the timeframe of the collapse can be characterised by the protein hydration radius decreasing. Attempts have also been made at predicting protein folding kinetics and timescales through the use of computer simulations [45]. Finally, an emerging technique for monitoring the process of protein folding is single molecule Förster resonance energy transfer (FRET). As it is examining the fluorescence of just a single molecule, instead of the average of a whole population of protein folds, FRET promises to provide insight into exact timescales for protein folding processes [69], and

has been utilised to measure more complex systems such as interactions with chaperones [58]. Whilst the characterisation of protein folding offered by these techniques is vast, none characterise the extent of mobility in the molten globule state, or how this relates to the aggregation proneness of the protein molecule.

2.2.6 Unfolding Methods

Protein folding experiments in general require the protein to become unfolded in order to be able to emulate the folding process. Some of the most common methods include the use of denaturing additives which are added in a concentrated amount to unfold the protein and then diluted in order to initiate folding [48]. Acids have been known to be used to denature a protein [55]. The high charge present at low pH acts to destabilise any intra protein charge-charge interactions, as such the unfolded state becomes more stable at low pH than the native [70]. However this conformational behaviour is not present in all proteins [71], especially in those whose native environment is acidic.

Denaturants most commonly used are urea [44], and guanidine hydrochloride [43]. High concentrations of these molecules perturb the protein structure through interaction with hydrogen bonding groups which results in a decrease of the proteins entropy and enthalpy [72]. Urea at high concentrations is capable of maintaining the unfolded protein through direct interaction with the polar residues within the protein and additionally through reducing the hydrophobic effect the surrounding water exerts on the protein residues [73]. Whilst these are efficient at unfolding the protein, by diluting the solution there will always remain a residual amount of denaturant in solution.

In contrast to this, temperature has also been used to denature proteins either to study the strength of bonds [74], or in order to then initiate folding [46]. By providing the energy to overcome hydrogen bonding within the protein, a temperature jump to above

the protein's melting temperature will result in complete unfolding of the protein where it may adopt multiple intermediate conformations in the pathway to the unfolded state [75]. However this method may also initiate aggregation, where inter-protein interactions will contribute to any energetic characteristics obtained from these studies.

2.3 Protein Aggregation

Protein aggregation occurs when a protein is subjected to unfavourable conditions (generally non-native), which cause structural deformation in the protein structure. This structural change can lead to the molecule being prone to self association, where it may interact with a similarly perturbed protein molecule to form an aggregate, generally resulting in loss of function of the two protein monomers [76]. The protein aggregate may then act as a nucleation site for aggregation, interacting with additional structurally perturbed monomers to form a larger aggregate.

Protein aggregates are found in a number of shapes; fibrous protein aggregates are commonly seen in neurogenerative diseases such as Creutzfeld-Jakob and Alzheimer's diseases [77]. Amorphous aggregates are a second class of aggregate which are most notably present in inclusion bodies in cells where a large concentration of non-natively folded protein occupies a small volume [78] and particulate aggregates which are spherical and have been formed *in vitro* through thermal aggregation in conditions near the isoelectric point of the protein [79].

As can be seen in the variety of structures aggregates may form, the study of protein aggregation carries importance in many areas [80]. Protein aggregates have been shown to be precursors to a number of neurodegenerative and prion diseases [81]; where understanding the formation of these aggregates may elucidate a route to treatment. Prevention of protein aggregate formation also plays a pivotal role in the production of

biopharmaceuticals, where proteins are produced in such quantities that they form aggregates in the inclusion bodies. These aggregates are non functional and therefore represent a loss in yield during production [82]. In addition, any presence of protein aggregates in therapeutics have been shown to cause immunogenicity, where the presence of the aggregate can cause the body to raise antibodies to the natively folded protein, nullifying any therapeutic effect [83].

2.3.1 Thermal Unfolding

Temperature dependant unfolding of proteins occurs when a protein in its native state is subjected to a sufficient temperature increase to overcome the thermodynamic stability of the protein. This process is endothermic as energy input is necessary for overcoming barriers to unfolding such as disrupting salt bridges and breaking hydrogen bonds that maintain the proteins tertiary structure. As the temperature of the protein increases, it is seen that the strength of hydrogen bonds and hydrophobic interactions gradually decreases [21]. The strength of these bonds continues to weaken and after increased heating to above the melting temperature the protein is denatured. This causes a loss of function and tertiary structure, resulting from the protein becoming partially unfolded.

The thermal unfolding of proteins has been seen to begin with the loss of tertiary structure dependent contacts [84], forming an intermediate, partially unfolded state, although not mentioned in the quoted paper, it bears resemblance to the molten globule state. The order of this unfolding may depend on unfolding temperature and the temperature at which simultaneous breakage of all structural bonds occurs; tΩ. [74]. In this paper, Cieplak *et al*. suggest from a molecular dynamic model, that below tΩ, protein unfolding occurs similar to the reverse order of folding, which appears to agree with the findings from Li *et al*. [84] above, where loss of tertiary structure is the first

49

step in thermal perturbation of the protein. Additionally, this seems consistent with the findings that the molten globule state is a prominent phase in protein folding [22]. Above tΩ however, the bond breakages within a protein are seen to occur at once, showing no preferential order of unfolding [74]. This more comprehensive denaturation would include removal of the large number of weak intermolecular interactions that are responsible for the secondary structure and hydrophobic packing. These bonds contribute to the stability of the protein when in its molten globule state, so therefore this state will not be present when unfolding occurs above tΩ. It would appear this characteristic is important, especially when considering refolding of the protein, as it is seen that protein denaturation can be reversible, but it depends on the types of the bonds which are broken [21], where peptide chain hydrolysis is an irreversible process.

The thermal unfolding of a protein, including the strength of bonding during its unfolding transitions, can be characterised by measuring the change in specific heat capacity; ΔC_p [85]. The heat capacity indicates the amount of energy required to increase the temperature of on mole of the protein by one degree. Specifically in protein unfolding measurements; this indicates the amount of heat energy required to break the bonds within the protein. The larger this value, the more highly structured and stable the protein is likely to be. This value is seen to be a highly positive change during each stage of protein unfolding [21]. This is relative to the high number of bonds broken, including many of the weak secondary structure interactions that are an inherent feature of the structure of the proteins intermediate states. As the protein becomes unfolded, it exposes hydrophobic residues, which are responsible for the interactions leading protein aggregation [86] in addition to the higher flexibility in the backbone of the partly unfolded state.

A further method for characterising protein unfolding is through the calculation of the melting temperature; the temperature at which half of the protein within the sample has been unfolded. A large melting temperature corresponds to a strong and highly stable conformation which is unlikely to unfold. Furthermore, multiple melting points can be extracted from this method, indicating the strengths of individual domains of the protein. Papain was shown to have 2 transitions in melting temperature; this was interpreted to represent the two separate domains of papain melting to give the molten globule conformation [62]. Additionally, this can reveal the presence of any stable intermediates that are formed within the unfolding process through the presence of the multiple melting temperatures [86]. Both the heat capacity and the melting temperature are common properties of proteins measured commonly by differential scanning calorimetry [87]. However as previously noted, during protein heating aggregation may also occur, and separating the unfolding energies from the contribution from aggregation is not considered.

2.3.2 Thermal Aggregation

Protein aggregation is proposed to be a universal characteristic of polypeptide chains when the native interactions normally present in the protein are destabilised, leading to an increase in hydrophobic interactions occurring between protein molecules [88, 89]. It has been shown with bovine serum albumin that only a small change in tertiary structure is necessary before aggregation occurs [90]. The hydrophobic exposure from the loss in tertiary structure is responsible for the formation of aggregate nuclei, initiating aggregation [91]. These nuclei are capable of propagating aggregation by the conversion of additional perturbed protein monomers [92].

The aggregation stage consists of an exothermic irreversible process; which is dependent upon the initial rate of formation of aggregate nuclei to determine the kinetics of aggregation thereafter [93]. The aggregation process may not occur immediately after structural perturbation begins however [94].

2.3.3 Aggregate Nucleation

Aggregate nucleation takes place where the conformational change in a monomer results in the binding of a second or multiple monomers whose conformation also differs from that of the native state. It is the initiating step in the aggregation process, most relevantly shown in the thermal aggregation of α-chymotrypsin A [95] where it has been postulated that the presence of the non-native aggregate prone monomer may encourage conformational change within other native protein molecules as it provides a lower energy barrier for the conformational change to occur. This nuclei then acts as a seeding template from which further monomer addition is facilitated [96]. When a large number of nuclei are present throughout the sample, the aggregation process flourishes.

An alternative method of action proposed is that there is already a low energy barrier between native monomer and the varying intermediate conformations. The rate of aggregate nuclei formation would be proportional to the concentration of monomer, and it is solely the collisions between intermediate molecules that drives aggregate formation [97]. In order for a nucleus to seed aggregation, it must remain stable for sufficient time for growth to occur from it. As growth is considered to be a slower process than the nucleation phase, the nuclei must remain for a substantial period of time. However in situations where very rapid growth occurs even the formation of reversible nuclei may seed the aggregation process [98].

It has been suggested that the protein monomer in a partially folded conformation may be the most prone to form an aggregate nuclei; even more so than the completely unfolded protein [99]. The explanation is that the hydrophobic areas in a partially folded protein would be exposed, but still grouped, offering a more concentrated area for aggregate interaction. However in the unfolded state, the hydrophobic residues would be randomly oriented around the backbone [100]. Within this paper, denaturants are used to maintain the unfolded state of the protein. It must be noted that the increasing denaturant concentration will play some contributing role to the effects seen here, as the urea molecules will be present around hydrophobic areas of the unfolded protein. Here they compete with the monomers for interactions with the protein, probably leading to a decreased rate of aggregation over "pure" unfolded protein.

Opportunities for these aggregate prone monomer species to appear arise when considering the protein folding pathway. As the protein occupies multiple states within the folding energy basin, one or more of these states may exist in equilibrium with a nucleation prone intermediate state. These nucleation prone states are often seen to contain a higher level of β-sheet than that which is found in the monomer [95]. In addition to this, there are also multiple routes which the folding may take to occupy one of these conformations [101]. These routes can be considered in terms of the protein energy basin, where changing the conditions within a system will also have an effect on the profile of the energy basin. In extreme cases this may make non native conformations of the protein the most energetically favourable, promoting aggregation. The nucleating monomer can be present in many forms; within the same type of monomer across multiple conformations [98, 102] and during folding where aggregates begin to form even before stable unfolding intermediates are present [103]. Furthermore it has been shown that aggregation can be prevented through the use of high pressure

during the unfolding stage, it is postulated by the author that this is because the high pressure physically prevents the unfolding process occurring at comparable temperatures [93].

The number of partially unfolded monomers required to create an aggregate nucleus can vary between systems. The formation of dimers has been shown to be sufficient to nucleate aggregation [104-106], but also oligomers [107] and larger nucleation sites formed from micelles in the case of amyloid β protein [108]. Furthermore the smallest stable aggregate nucleus of a short polypeptide chain has been shown to be in the multimer range [109] and in these circumstances, formation of any reversible aggregates is considered a pre-nucleation step. Here the trend seems to be that larger molecules require fewer numbers of monomer create a stable aggregate as they offer more protein-protein interaction opportunities than smaller molecules, which therefore require a larger population.

In the aggregation model which consists of both nucleation and growth phases [110], the nucleation stage is considered to be the rate limiting step within the system [102]. However, in the system of particulate aggregation investigated here, all nucleation sites are thought to form simultaneously [79], where only a process in the monomer addition stage can be the rate limiting factor in growth.

Once nucleation has occurred, there are multiple ways in which aggregation might progress, depending on the concentration within system. There may be minimal growth whilst the aggregates remain in solution, there might be rapid growth to larger insoluble aggregate, or a formation of a gel phase may occur if the concentration is sufficiently high [98] [111].

2.3.4 Aggregate Growth and Structure

Protein aggregation can take a number of forms and is dependent upon the environmental conditions the proteins are subjected to during the aggregation process. The most commonly investigated of these is the amyloid fibril, but in addition to this there are also particulate and spherulite aggregates, which have been shown to be formed from proteins which also aggregate into amyloid fibrils [79, 112, 113]. These aggregation characteristics have been proven for multiple proteins whose structures and functions are unrelated, which leads to the assumption that these characteristics are properties of all proteins.

Fibril formation is formed from highly charged protein molecules. These charged molecules provide long ranging repulsion, which only allows for further proteins to bind in a small number of arrangements. This type of aggregation particularly favours the formation of β-sheets and leads to a one dimensional formation of aggregate growth [113]. The formation of ordered aggregates of BSA was discovered when forming amyloid fibrils away from the isoelectric point [90]. The mechanism of aggregation in BSA was additionally shown to include the transformation of alpha helices to β-sheets during partial unfolding of the tertiary structure.

Conversely to fibrils forming in conditions away from their isoelectric point, spherulites have been seen to form in a pH region closer to the isoelectric point of the protein. These protein structures are not well characterised; but they contain short runs of β-sheet, forming fibrils which emanate from a central nucleus to give a shaped akin to a "maltese cross" as described in the paper [112]. Insulin is shown to form spherulites readily [114] whereas β-lactoglobulin only forms them infrequently, generally forming at the same time as amyloid fibrils, which possibly shows its dependence on the

isoelectric point of the environment for spherulite formation. [113, 115] Spherulites can consist of a core of collapsed fibrils in the nucleus. This is one possible explanation for how they are seen to occur alongside amyloid fibrils. However, for them to be seen to exist alongside fibrils, they must be present in an environment away from the protein's isoelectric point, so it is possible that spherulite formation may be possible over a range of conditions.

Near the isoelectric point, there is a significant barrier to "ordered" aggregation, and unlike in fibril formation, there is no long range charge repulsion present. This decreases the likelihood of structural rearrangements, and because the aggregation occurs in a random and non-specific way, the particulates formed are spherical. [113] Proteins that have been proven to form particulate aggregates had all previously been shown to form amyloid fibrils under different aggregation conditions. This difference in aggregation is completely independent of amino acid sequence, and is determined by the general state of the protein in its environment. [79]

Particulate aggregation occurs around the isoelectric point, where there is a low overall amount of electrostatic interactions, therefore there are many arrangements in which aggregates can form, which leads to aggregation growth occurring randomly, causing the aggregates to be approximately spherical in shape. Particulates can have a size of around 250nm and contain some β-sheet content [113]. These particulates have a strong tendency to induce aggregation. After particulate aggregation, remaining monomeric protein within the solution becomes deposited between the particulates. This leads to the appearance of the particulates being connected [79].

Particulates, were found to have similar levels of secondary structure after heating when compared to the native state in various samples: BSA, Myoglobin, transthyretin; where

there was an increase in secondary structure content [79]. Here, aggregation of these samples proved that the proteins were not aggregating through disulphide bond formation as both myoglobin and transthyretin do not contain the cysteine residues required and the aggregation was more likely a result of partial unfolding of the proteins

The rate of heating is shown to affect particulate aggregation size [79]. The higher the heating rate, the smaller the aggregation of the particle, these results led the authors to the conclusion that particulate nucleation is determined by the heating temperature, where a higher temperature results in more nucleation sites forming, the subsequent growth results in a smaller diameter of particle where remaining protein concentration is uniformly distributed around each of the particulate nuclei. [116]

β-lactoglobulin subjected to an environment away from the protein's isoelectric point was shown to form clear and fibrillar gels. Conversely, when the β-lactoglobulin protein was present in an environment near the protein's isoelectric point, the gels formed after heating were shown to be opaque and consist of particulates [116, 117] Partial denaturation of the β-lactoglobulin occurs at its isoelectric point, this unfolding of the protein allows an area in which interactions between molecules can occur, and aggregation can be facilitated. This has been shown to be similar for other proteins which will at the least partially unfold.

These aggregation interactions in protein particulates have been shown to interact through (sometimes non-native) β-sheet structures [118]. Current understanding indicates that this is the same interactions occurring in amyloid fibril formation, therefore work on the early stages (i.e. nucleation) of particulate formation may also be applicable to that of amyloid formation. In addition, the kinetics and thermodynamics of the particulate system are not well characterised in the literature, understanding these

factors would not only improve our understanding of the system, but provide numerical benchmarks from which aggregate preventing additives may be tested.

2.3.5 Techniques for Measuring Protein Aggregation

The following section introduces the main techniques used in the literature to study protein aggregates. Though not exhaustive, it covers the techniques used to extract the main properties of an aggregate system, namely; shape, structure, size and concentration of the aggregates. Examples of their use in the field are included, and the advantages and limitations of each are discussed.

I. Size Exclusion Chromatography (SEC)

SEC separates protein molecules and aggregates according to their hydrodynamic volume [119]. The column used to perform the separation is populated with porous silica particles. The protein aggregate sample is passed into the column and the larger aggregate particles are excluded from passing through the pores; passing around the edges of the beads instead. This results in larger aggregates being eluted from the column quicker than smaller aggregates, or monomers which pass through the pores in the beads as they travel through the column. The concentration of the eluted protein can be analysed by a spectrophotometer and the size of the particles can be compared to elution times for standard particles of known size [120]. Gabrielson *et al.* showed that this technique could be successfully used to quantify the level of aggregation in antibody samples [121], whilst Grillo *et al.* showed how the SEC can be used to complement other techniques for biopharmaceutical proteins [122]. SEC however requires an extensive knowledge and experience with the technique in order to ensure the column has a suitable composition in order to separate the aggregate system, and

will generally involve diluting the aggregate sample which in turn may have some effect on the aggregation present within.

II. Static Light Scattering (SLS)

SLS measures the scattering intensity of light coming from the protein particles, which are illuminated by a single wavelength beam of light, and the scattered light is generally measured at multiple angles. The time averaged scattering of this data is deconvoluted to give information on the size, shape and concentration of the sample. This requires the sample to be of high enough concentration to produce sufficient scattering, whilst low enough to avoid multiple scattering. This technique can be utilised as the sample passes through a column or as a more static sample measurement from a cuvette, although the latter involves regular and thorough cleaning of the cell. In addition, for full analysis, multiple concentrations of the protein are required, which may cause error if the system is concentration dependent. This technique has been shown to be valuable in the detection of low levels of aggregate present in samples of immunoglobulin G [123] where SLS was used in conjunction with SEC.

III. Dynamic Light Scattering (DLS)

DLS is used in a similar manner to that of SLS, but the technique measures the fluctuations in the light intensity over time instead of a time averaged intensity across multiple angles (as is the case with SLS). These fluctuations are due to the random motion of the protein particles through the sample. This information can be deconvoluted to give information for a wide range of aggregate sizes and distributions [124]. One of the advantages of DLS is that there are no special sample requirements involved, and therefore it can be used on the direct sample which is especially useful in investigating the kinetics of aggregation. A prime example of the use of DLS is in a

study of β-lactoglobulin aggregates to form fibrils [125], where the technique is used alongside other scattering experiments to characterise the formation of the fibrils. As DLS requires a period of time over which to collect the data and ascertain the sample size, it is not suited to dynamic processes such as protein aggregation. Furthermore, complex data analysis and a highly refined setup are required in order to resolve separate but closely related size populations.

IV. Analytical Ultracentrifugation

This technique uses centrifugal force to separate aggregates based on size, and concentration through sedimentation equilibrium. This is then quantified using absorption or fluorescence spectroscopy. It requires a low amount of sample preparation, and therefore can be used to measure the protein aggregate sample directly [124]. This technique requires a detailed knowledge in order to separate the aggregates sufficiently, much like SEC. Additionally there requires some detailed analysis and complex separation issues that can arise from the aggregate solution [126].

V. Field Flow Fractionation (FFF)

FFF uses an external field (generally flow, electrical or magnetic) to separate protein aggregates under flow. Coupled with this is the simultaneous measurement of the mass, size and charge of the particles within the system, separating the particles due to the movement of solution components in different flow channels. FFF has a wide range of particle detection from between 1 nm to 100 µm [127] and can be utilised to give an excellent separation between particles. There are limited examples of its use in protein aggregation where it is only limited to the study of soluble aggregates, but it has been suggested that with increasing knowledge of how to extract data from this technique it is becoming more useful for analysis of aggregates [128].

VI. Gel Electrophoresis

Gel electrophoresis uses sodium dodecyl sulphate (SDS) as a denaturant to linearise the protein molecules. These molecules have an overall negative charge, as therefore migrate through a gel when a current is passed across it. The separation of the protein molecules is performed by molecular weight, with smaller proteins migrating further down the gel than the larger aggregates. The molecules are then stained by a visualising agent in order to quantify the amount of protein present, and their migration distance compared to that of a known standard. This is an often used and simple tool for both the measurement of size and quantity of protein aggregates, especially when separating large ranges of aggregate [129]. However the use of SDS results in only aggregates which are covalently linked to still be present within the gel, this may not elucidate the full extent of protein aggregation within the sample. This problem may be solved by the use of native-PAGE, but there is little evidence for it in the literature at present.

VII. Microscopy

Light microscopy may only be used on very large aggregates, where visual inspection of particle number and size can be ascertained from the sample by comparison with standard samples of known size. It is also possible to have computerised methods of particle counting utilising a flow setup for the aggregates [130]. In the sub visible range, both transmission and scanning electron microscopy have been used to evaluate aggregate size and shape to a very accurate degree. This requires a large amount of sample preparation in a vacuum and coating the samples with gold in order to get sufficient contrast to see the scanning image [120]. In addition to the lengthy setup requirements, microscopy generally deals with a small area of the sample, unlike most techniques which are an average representation of the sample as a whole. Despite this,

microscopy is often used for sizing of protein aggregates with many examples in the literature; Nielson *et al.* formed insulin fibrils and characterised the length and diameters using electron microscopy [131] and elsewhere the effect of bovine serum albumin on membrane fouling has been characterised using microscopy to examine the structural changes incurred [132].

VIII. Spectroscopy

Fluorescence spectroscopy utilises fluorescent dyes which are natively fluorescent, and when bound to a protein aggregate, have a shift in the fluorescent light which they produce [133]. Dyes such as 1-anilinonaphthalene-8-sulfonic acid bind to hydrophobic regions on the protein, and have been shown to bind preferentially to protein aggregates [134], where they interact through the available hydrophobic regions which are exposed on the aggregate. This fluorescence change can be monitored through the use of spectroscopy or fluorescence microscopy. A fitting example of spectroscopic dyes being used in protein aggregation is in the use of Nile red to monitor the thermal aggregation of β-galactosidase [135]. This technique also enables very low levels of quantification, however as it binds solely to hydrophobic patches within the protein, it is difficult to study aggregates formed through refolding studies as the probe will bind strongly to the unfolded protein.

Additionally, intrinsic tryptophan fluorescence can be used as a measure of proteins tertiary structure before unfolding that leads to aggregation begins. This is covered in the protein folding section of this work.

Circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR) both make use of the chiral nature of the protein molecules to analyse the secondary structure of the aggregates. The differences in light absorbtion can be translated into a secondary

structure content for the protein. These techniques are powerfully accurate, where circular dichroism has been used in secondary structure quantification between differing self association states of human insulin [136] and FTIR has been utilised to analyse the secondary structure content in the aggregates of the same molecule [131]. Whilst these techniques are very useful in quantifying secondary structure, they are an average of the entire sample, so aggregates present in monomer solutions may need separating before successful analysis can be undertaken.

IX. Ultra Violet Light Scattering Spectroscopy (UVLSS)

UVLSS is a light scattering technique which utilises a range of wavelengths of ultra violet light in a spectrophotometer to calculate the scattering produced from the sample. The theory of the technique will be discussed in more detail in the following chapter. UVLSS allows the diameters of particles above 50 nm (depending on refractive index of the protein) to be measured. A proof of concept experiment has been run where this technique accurately sized polystyrene spheres of set diameters [137, 138]. This allows for in-situ time resolved dynamic measurements of protein aggregation and from this the calculation of monomer depletion and kinetics [139] and unlike some techniques, is non-destructive. This is an underused technique but is ideal for monitoring the aggregation of proteins on the larger length scales as it does not detect particles of the size scale of monomers, therefore data deconvolution is not necessary [140]. However this also means that detection of small aggregates or aggregate nuclei is not possible using this technique.

X. Small Angle Scattering

Small angle scattering utilises a beam of either neutrons or X-rays which is directed towards a sample and scattered onto a detector. This scattering gives information on

particle size shape, concentration and interactions [141]. This will be covered in more detail in the theory section. This technique offers a wide range of sample environments for studying many systems, including proteins [142]. Details from the system can be extracted by modelling the system and comparing the modelled profile to the experimental. This can give much more detailed information such as the flexibility or subunit arrangement of proteins within the system [143].

Information provided by the small angle scattering techniques have been used to study the quality of the protein fold, including estimation of secondary structure levels [144]. This has been used to monitor the effect of pH on the conformational state of bovine serum albumin and the resulting protein-protein interactions that take place [145]. Experiments on dynamic systems such as protein folding have been shown to yield accurate results when compared to complimentary techniques such as tryptophan fluorescence and circular dichroism, where the formation of the globular protein is monitored [146].

There are further examples of how beamline small angle scattering can be used to take measurements on dynamic systems, owing to the short time period it takes for a single scattering pattern to be obtained. This was utilised using SAXS by *M.Kataoka et al.* to monitor the radius of gyration of cytochrome C over time to study the compactness of the protein as it passes through the molten globule state [147]. More relevantly this technique has also been used to monitor the dynamic aggregation of BSA after breakage of the disulphide bond [148]. In this early experiment however, the appearance of aggregation was sufficient, and a quantitative examination of the data was not undertaken.

Static samples of aggregates have additionally been monitored using this technique, such as aggregates of amyloid β-protein, where the number of molecules that form an aggregate, and their hydrodynamic radius were uncovered [149]. The aggregation of staphylococcal nuclease has been probed under small angle scattering, coupled with FTIR and circular dichroism to understand the transitions in structure that occur during aggregation [100]. More recently, the heat induced aggregation of β-lactoglobulin was studied in order to understand the nucleation phase of aggregation [150]. This allowed the initial nuclei shapes of tetramers and cyclic clusters to be elucidated.

2.4 Proteins

This section will cover the proteins utilised in these experiments and the relevant literature in the field of aggregation and refolding, depending on their use in experiments within this project.

2.4.1 Bovine Serum Albumin
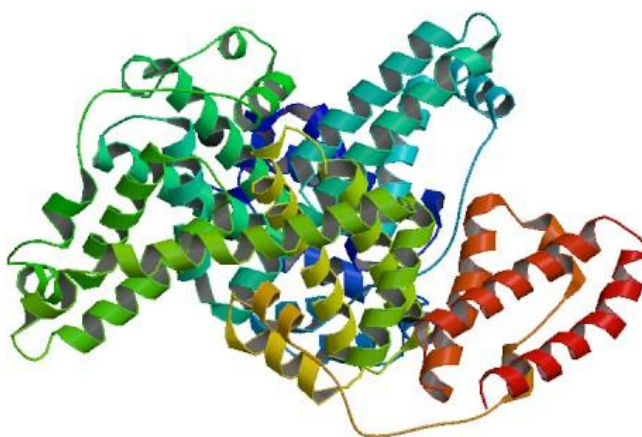
Figure 2.10: Structure of BSA molecule. PDB code 3V03 [1].

65

BSA is a 66kDa serum protein present in cows. Its structure has recently been solved (PDB code 3V03) [1], it consists of 583 amino acids and contains a largely helical α-helical structure. Due to the ease of obtaining the protein from cow serum; and its relative low cost, BSA is used as a standard in many protein assays and also used in protein aggregation studies (Figure 2.10)

When subjected to a temperature of 70 ⁰C or above, it is seen that there is a change in secondary structure content with an increase in the quantity of β-sheet structure whilst the quantity of α-helices decreases. Above 70 ºC the formation of β-sheets is irreversible [151]. The self association interactions occur through the non-native, more open domain I of neighbouring molecules. It is this initial structural change which is the initiating step for aggregation [152]. Additionally the amount of β-sheet content formed and the amount of aggregates increase as the protein environment is more distant from the isoelectric point [90]. This secondary structure formation may be altered by the presence of metal ions in the solution [153], however it does not affect the propensity of the molecules to aggregate.

2.4.2 β-Lactoglobulin
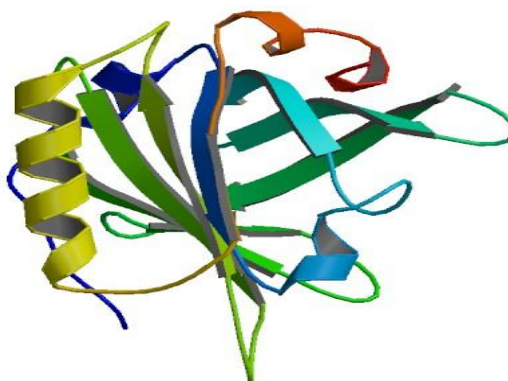
Figure 2.11: Crystal structure of β-lactoglobluin. PDB code 3NPO [1].

β-lactoglobulin is an 18kDa protein which consists mostly of β-sheets (PDB code 3NPO) [1]. It is present in the milk of cows, and as with BSA, due to the ease in isolating large quantities of the protein, it is often used in folding and aggregation studies, in addition to use as a standard in protein assays (Figure 2.11).

The thermal aggregation of β-lactoglobulin is a rapid process, due to the high hydrophobic nature of the residues within the protein. The thermal unfolding of the molecule has been shown to be immediately followed by aggregation of the monomer [154]. In later studies it was elucidated that the lactoglobulin protein undergoes a slight structural arrangement before the aggregation is seen to occur [155]. This appears to agree with early work that suggests that the β-strands present in lactoglobulin are transformed to antiparallel β-sheeets before aggregation [156]. The monomers associate through a combination of hydrophobic and disulphide interactions. Additionally, the levels of secondary structure present within the aggregate molecules can also be dependent on concentration, with higher concentrations leading to a larger proportion of secondary structure remaining [157]. This appears to be a stability conferred by the presence of other native protein molecules.

There has been some interest in the refolding of β-lactoglobulin as during folding, there is an α-helical intermediate present, before the formation of the native β-strand protein [158]. The transformation to native β structure occurs in the late stages of the folding, although there is a significant amount of β-sheet already present within the intermediate [159] .

2.4.3 Myoglobin
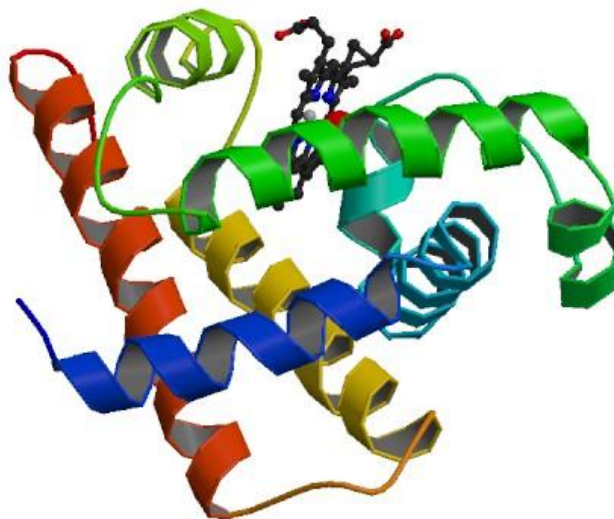
Figure 2.12: Crystal structure of myoglobin. PDB code 1MBN [1].

Myoglobin is an 18kDa protein which consists of many α-helices. Myoglobin is 153 residues in length and is an oxygen carrying protein which exists in the muscle, including the heart [1, 160] Figure 2.12.

Early studies of myoglobin showed that there is a decrease in α-helical content when the protein is solubilised in water [161] therefore the molecule may be less stable than the native. More recent aggregation studies have identified the presence of an aggregation prone intermediate which retains the core structure of the myoglobin molecule, resembling the early folding intermediate [162]. Additionally, this paper showed that there is a formation of β-sheets when the protein tends to aggregate at it is in conditions away from the isoelectric point, which is comparable with studies seen on other proteins mentioned here. When the aggregates of myoglobin are formed at a low temperature (50

ºC), there is a large quantity of the native structure still present within the aggregate subunits, however the process is irreversible [163].

During refolding of myoglobin the protein has a tendency to form a non native dimer with exposed hydrophobic residues [164]. Secondary structure formation takes place in two stages which appear to form at different rates [165]. The initial stage is the formation of around 40 % of the secondary structure whereupon the remaining secondary and tertiary structure assemble around in the second stage. Here the molten globule state of myoglobin may exist [166]. Succeeding this, minor structural rearrangement results in the final native structure.

2.4.4 Cytochrome C
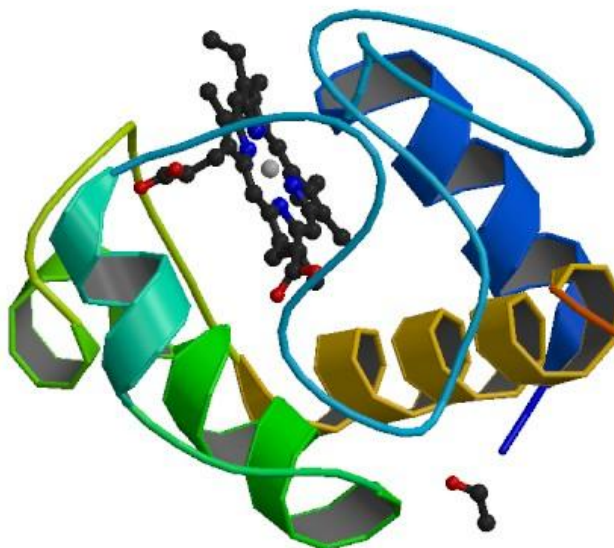
Figure 2.13: Cytochrome C crystal structure. PDB code 3CYT [1].

Cytochrome C is an electron carrying protein found in cells for production of ATP [1]. It is 24kDa in size and contains a secondary structure which consists solely of α-helices

69

(Figure 2.13). Folding of cytochrome C involves formation of the protein structure around the heme molecule, within 1 millisecond of initiating folding the structure of cytochrome C can be collapsed around the heme moiety [167]. The heme is not thought to determine the folding behaviour of cytochrome C, which is driven through protein-protein interactions [168]. Rapid rearrangement of exposed side groups follows to achieve energetically favourable conformations. Whilst in this molten globule state cytochrome C has also been shown capable of dimerising [169], whilst partially folded states of cytochrome C are the most likely to be aggregate prone [167]

2.4.5 Ribonuclease A
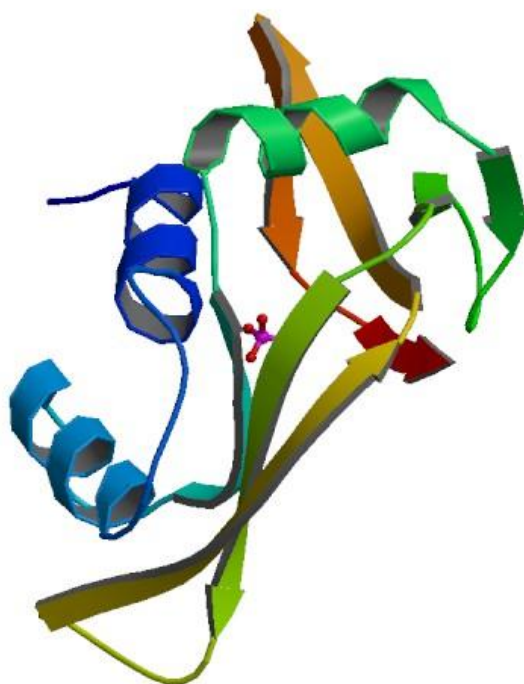
Figure 2.14: Crystal structure of ribonuclease. PDB code 5RSA [1].

Ribonuclease A is one enzyme responsible for degrading RNA. It consists of 124 amino acids and is 13.8kDa in size. It consists of both α-helices and β-sheets and is a very stable molecule (Figure 2.14). Ribonuclease was one of the first proteins used in folding studies [170], where amino acid sequence was shown to determine the folded structure

of the protein. During folding ribonuclease forms a number of intermediates which utilises at least 2 disulphide bonds to retain proximity of separate sections of the amino acid chain to complete folding [171].

2.4.6 Summary

The proteins used here for both aggregation and refolding experiments are model proteins in their respective fields. They represent a cross section of secondary structural contents, and isoelectric points. During experimentation this will allow for any contributing effect of starting structure, or properties, to be contrasted. In refolding experiments, proteins with a lack or (or minimal) disulphide bonds were favoured in order to achieve the most highly unfolded state possible, without using reductants and oxidants to moderate these bonds.

Using this range of proteins, generic questions relating to aggregation could be asked, with the varied systems representing general trends of all proteins:

a) How does particulate aggregate nucleation occur? Is it as simple as suggested from microscopy studies by *Krebs et al.*(2.3.4)?

b) Does the model for particulate aggregate growth (2.3.4) agree with dynamic measurements of the model proteins aggregating? Is it possible to elucidate kinetic and thermodynamic values for this process?

c) Does the mobility of a protein monomer affect its propensity to aggregate? How mobile must a protein be to form aggregates?

2.5 References

1. Berman, H., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
2. Alberts, B., et al., *Molecular Biology of the Cell.* 4th ed. 2002: Garland Science.
3. Berg, J.M., J.L. Tymoczko, and L. Stryer, *3.6 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure*, in *Biochemistry*. 2002.
4. Hartl, U. and M. Hayer-Hartl, *Converging concepts of protein folding in vitro and in vivo.* Nature Structural & Molecular Biology, 2009. **16**(6): p. 574-581.
5. Bartlett, A. and S. Radford, *An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms.* Nature Structural & Molecular Biology, 2009. **16**(6): p. 582-588.
6. Fersht, A.R., *Nucleation mechanisms in protein folding.* Current opinion in structural biology, 1997. **7**: p. 3-9.
7. Karplus, M., *The Levinthal paradox: yesterday and today.* Folding and Design, 1997. **2**(Supplement 1): p. S69-S75.
8. Privalov, P. and G. Makhatadze, *Contribution of Hydration to Protein Folding Thermodynamics.* Journal of Molecular Biology, 1993. **232**(2): p. 660-679.
9. Pickett, S. and M. Sternberg, *Empirical Scale of Side-Chain Conformational Entropy in Protein Folding.* Journal of Molecular Biology, 1993. **231**(3): p. 825-839.
10. Hartl, U., *Molecular chaperones in cellular protein folding.* Nature, 1996. **381**(6583): p. 571-580.
11. Mayer, M.P. and B. Bukau, *Hsp70 chaperones: Cellular functions and molecular mechanism.* Cellular and Molecular Life Sciences, 2005. **62**(6): p. 670-684.
12. Bukau, B. and A.L. Horwich, *The Hsp70 and Hsp60 chaperone machines.* Cell, 1998. **92**(3): p. 351-366.
13. Gething, M.-J. and J. Sambrook, *Protein folding in the cell.* Nature, 1992. **355**(6355): p. 33-45.
14. Fersht, A.R. and V. Daggett, *The present view of the mechanism of protein folding.* Nature Reviews, Molecular Cell Biology, 2003(497-502).
15. Guo, Z. and D. Thirumalai, *Kinetics of protein folding: Nucleation mechanism, time scales, and pathways.* Biopolymers, 1995. **36**(1): p. 83-102.
16. Ohgushi, M. and A. Wada, *'Molten-globule state': a compact form of globular proteins with mobile side-chains.* FEBS Letters, 1983. **164**(1): p. 21-24.
17. Veitshans, T., D. Klimov, and D. Thirumalai, *Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties.* Folding and Design, 1997. **2**(1): p. 1-22.
18. Schmid, F. and R. Baldwin, *Acid catalysis of the formation of the slow-folding species of RNase A: Evidence that the reaction is proline isomerization.* Proceedings of the National Academy of Sciences, 1978. **75**(10): p. 4764-4768.
19. Daggett, V. and M. Levitt, *A Model of the molten globule state from molecular dynamics simulations.* Proceedings of the National Academy of Science, 1992. **89**: p. 5142-5146.
20. Ptitsyn, O.B. and V.N. Uversky, *The molten globule is a third thermodynamical state of the protein molecules.* FEBS Letters, 1994. **341**: p. 15-18.
21. Jaenicke, R., *Protein stability and molecular adaptation to extreme conditions.* European Journal of Biochemistry, 1991. **202**: p. 715-728.

22.   Ptitsyn, O.B., *How the Molten Globule became.* Trends in Biochemical Sciences, 1995. **20**: p. 376-379.

23.   Peng, Z. and P.S. Kim, *A protein dissection study of a molten globule.* Biochemistry, 1994. **33**: p. 2136-2141.

24.   Arai, M. and K. Kuwajima, *Rapid formation of a molten globule intermediate in refolding of a-lactalbumin.* Folding & design, 1996. **1**: p. 275-287.

25.   Onuchic, J.N., Z. Luthley-Schulten, and P.G. Wolynes, *Theory of protein folding: the energy landscape perspective.* Annual Reviews of Physical Chemistry, 1997. **48**: p. 545-600.

26.   Arai, M. and K. Kuwajima, *Role of the molten globule state in protein folding.* 2000, Elsevier. p. 209-282.

27.   Barrick, D. and B. R.L., *The molten globule intermediate of apomyoglobin and the process of protein folding.* The Protein Society, 1993. **2**: p. 869-876.

28.   Pande, V. and D. Rokhsar, *Is the molten globule a third phase of proteins?* Proceedings of the National Academy of Sciences, 1998. **95**(4): p. 1490-1494.

29.   Santucci, R., F. Sinibaldi, and L. Fiorucci, *Protein folding, unfolding and misfolding: role played by intermediate States.* Mini reviews in medicinal chemistry, 2008. **8**(1): p. 57-62.

30.   Skora, L., S. Becker, and M. Zweckstetter, *Molten Globule Precursor States Are Conformationally Correlated to Amyloid Fibrils of Human β-2-Microglobulin.* Journal of the American Chemical Society, 2010. **132**(27): p. 9223-9225.

31.   Gerber, R., et al., *Oligomerization of the Human Prion Protein Proceeds via a Molten Globule Intermediate.* Journal of Biological Chemistry, 2007. **282**(9): p. 6300-6307.

32.   Soto, C. and L. Estrada, *Protein Misfolding and Neurodegeneration.* Arch Neurol, 2008. **65**(2): p. 184-189.

33.   Gazit, E., *The correctly folded state of proteins: Is it a metastable state?* Angewandte Chemie International Edition, 2002. **42**: p. 257-259.

34.   Bhattacharyya, A., A. Thakur, and R. Wetzel, *Polyglutamine aggregation nucleation: Thermodynamics of a highly unfavorable protein folding reaction.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15400-15405.

35.   Kopito, R.R., *Aggresomes, inclusion bodies and protein aggregation.* Trends in cell biology, 2000. **10**: p. 523-530.

36.   Middelberg, A., *Preparative protein refolding.* Trends in biotechnology, 2002. **20**(10): p. 437-443.

37.   Li, M., Z.-G. Su, and J.-C. Janson, *In vitro protein refolding by chromatographic procedures.* Protein Expression and Purification, 2004. **33**(1): p. 1-10.

38.   Eftink, M., *Fluorescence Techniques for Studying Protein Structure.* Methods of Biochemical Analysis, 2006: p. 127-205.

39.   Wuethrich, K., *The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination.* Accounts of Chemical Research, 1989. **22**(1): p. 36-44.

40.   Kuwajima, K., et al., *Comparison of the transient folding intermediates in lysozye and a-lactalbumin.* Biochemistry, 1985. **24**: p. 874-881.

41.   Radford, S. and C. Dobson, *Insights into Protein Folding Using Physical Techniques: Studies of Lysozyme and α-lactalbumin.* Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 1995. **348**(1323): p. 17-25.

42. Demetrius, L., *Thermodynamics and kinetics of protein folding: an evolutionary perspective.* Journal of theoretical Biology, 2002. **217**(397-411).

43. Nakatani, H., et al., *Equilibrium and kinetics of the folding and unfolding of canine milk lysozyme.* Biochemistry, 2007. **46**: p. 5238-5251.

44. Arai, M., *Kinetic refolding of β-lactoglobulin. Studies by synchrotron X-ray scattering, and circular dichroism, absorption and fluorescence spectroscopy.* Journal of Molecular Biology, 1998. **275**(1): p. 149-162.

45. Veitshans, T., D.K. Klimov, and D. Thirumalai, *Protein folding kinetics:timescales,pathways and energy landscapes in terms of sequence-dependent properties.* Folding and Design, 1996. **2**: p. 1-22.

46. Polkar, N., et al., *pH and Temperature-induced molten globule-like denatured states of equinatoxin II: A study by UV-melting, DSC, Far- and near-UV CD spectroscopy, and ANS fluorescence.* Biochemistry, 1997. **36**: p. 14345-14352.

47. Qin, Z., *Refolding of β-lactoglobulin studied by stopped-flow circular dichroism at subzero temperatures.* FEBS Letters, 2001. **507**(3): p. 299-302.

48. Bolen, W. and G. Rose, *Structure and Energetics of the Hydrogen-Bonded Backbone in Protein Folding.* Annual Review of Biochemistry, 2008. **77**(1): p. 339-362.

49. Uversky, V.N., *Use of fast protein size exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule.* Biochemistry, 1993. **93**: p. 13288-13298.

50. Berg, B.v.d., R.J. Ellis, and C.M. Dobson, *Effects of macromolecular crowding on protein folding and aggregation.* The EMBO Journal, 1999. **18**(24): p. 6927-6933.

51. Finke, J.M. and P.A. Jennings, *Early Aggregated States in the Folding of Interleukin-1β.* Journal of Biological Physics, 2001. **27**(2): p. 119 ppl=-131.

52. St.John, R.J., J.F. Carpenter, and T.W. Randolf, *High pressure fosters protein refolding from aggregates at high concentrations.* Proceedings of the National Academy of Science, 1999. **96**(23): p. 13029-13033.

53. Edwin, F., Y.V. Sharma, and M.V. Jagannadham, *Stabilization of molten globule state of papain by urea.* Biochemical and Biophysical Research Communications, 2002. **290**: p. 1441-1446.

54. Rothman, J. and R. Schekman, *Molecular Mechanism of Protein Folding in the Cell.* Cell, 2011. **146**(6): p. 851-854.

55. Makio, T., M. Arai, and K. Kuwajima, *Chaperonin-affected refolding of α-Lactalbumin: Effects of nucleotides and the co-chaperonin GroES.* Journal of Molecular Biology, 1999. **293**: p. 125-137.

56. Makio, T., E. Takasu-Ishikawa, and K. Kuwajima, *Nucleotide-induced transition of GroEL for the high affinity to the low-affinity state for a target protein: Effects of ATP and ADP on the GroEL affected refolding of α-lactalbumin.* Journal of Molecular Biology, 2001. **312**: p. 555-567.

57. Sharma, S., et al., *Monitoring Protein Conformation along the Pathway of Chaperonin-Assisted Folding.* Cell, 2008. **133**(1): p. 142-153.

58. Hofmann, H., et al., *Single-molecule spectroscopy of protein folding in a chaperonin cage.* Proceedings of the National Academy of Sciences, 2010. **107**(26): p. 11793-11798.

59. Chen, Y. and M. Barkley, *Toward Understanding Tryptophan Fluorescence in Proteins†.* Biochemistry, 1998. **37**(28): p. 9976-9982.

60. Radford, S. and C. Dobson, *Insights into protein folding using physical techniques: studies of lysozyme and a-lactalbumin.* Philosophical Transactions: Biological Sciences, 1995. **348**: p. 17-25.

61. Hattori, M., et al., *Complete refolding of bovine β-lactoglobulin requires disulfide bond formation under strict conditions.* Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics, 2005. **1752**(2): p. 154-165.

62. Edwin, F. and M.V. Jagannadham, *Sequential unfolding of papain in molten globule state.* Biochemical and Biophysical Research Communications, 1998. **252**: p. 654-660.

63. Kuwajima, K., et al., *Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism.* FEBS Letters, 1987. **221**(1): p. 115-118.

64. Elove, G., et al., *Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy.* Biochemistry, 1992. **31**(30): p. 6876-6883.

65. Bertucci, C., M. Pistolozzi, and A. De Simone, *Structural Characterization of Recombinant Therapeutic Proteins by Circular Dichroism.* Current Pharmaceutical Biotechnology, 2011: p. 1508-1516.

66. Wallace, B.A., *Protein characterisation by synchrotron radiation circular dichroism spectroscopy.* Quarterly Reviews of Biophysics, 2009. **42**(04): p. 317-370.

67. Laureto, P.P.d., et al., *Probing the molten globule state of alpha lactalbumin by limited proteolysis.* Biochemistry, 1995. **34**: p. 12596-12604.

68. Nishimura, C., et al., *Sequence Determinants of a Protein Folding Pathway.* Journal of Molecular Biology, 2005. **351**(2): p. 383-392.

69. Schuler, B. and W. Eaton, *Protein folding studied by single-molecule FRET.* Current Opinion in structural biology, 2008. **18**(1): p. 16-26.

70. Yang, A.-S. and B. Honig, *Structural Origins of pH and Ionic Strength Effects on Protein Stability.* Journal of Molecular Biology, 1994. **237**(5): p. 602-614.

71. Fink, A., et al., *Classification of Acid Denaturation of Proteins: Intermediates and Unfolded States.* Biochemistry, 1994. **33**(41): p. 12504-12511.

72. Makhatadze, G. and P. Privalov, *Protein interactions with urea and guanidinium chloride.* Journal of Molecular Biology, 1992. **226**(2): p. 491-505.

73. Bennion, B. and V. Daggett, *The molecular basis for the chemical denaturation of proteins by urea.* Proceedings of the National Academy of Sciences, 2003. **100**(9): p. 5142-5147.

74. Cieplak, M. and J.L. Sulkowska, *Thermal unfolding of proteins.* The Journal of Chemical Physics, 2005. **123**: p. 194908.

75. Day, R., et al., *Increasing temperature accelerates protein unfolding without changing the pathway of unfolding.* Journal of Molecular Biology, 2002. **322**(1): p. 189-203.

76. Cohen, F. and J. Kelly, *Therapeutic approaches to protein-misfolding diseases.* Nature, 2003. **426**(6968): p. 905-909.

77. Koo, E., P. Lansbury, and J. Kelly, *Amyloid diseases: Abnormal protein aggregation in neurodegeneration.* Proceedings of the National Academy of Sciences, 1999. **96**(18): p. 9989-9990.

78. Przybycien, T., et al., *Secondary structure characterization of β-lactamase inclusion bodies.* Protein Engineering, 1994. **7**(1): p. 131-136.

79. Krebs, M.R.H., G.L. Devlin, and A.M. Donald, *Protein particulates: another generic form of aggregation?* Biophysical Journal, 2007. **92**: p. 1336-1342.

80. Bondos, S., *Methods for Measuring Protein Aggregation.* Current Analytical Chemistry, 2006. **2**(2): p. 157-170.

81. Ross, C. and M. Poirier, *Protein aggregation and neurodegenerative disease.* Nat Med, 2004. **10**: p. S10-S17.

82. Berkowitz, S., *Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals.* The AAPS Journal, 2006. **8**(3): p. E590-E605.

83. De Groot, A. and D. Scott, *Immunogenicity of protein therapeutics.* Trends in Immunology, 2007. **28**(11): p. 482-490.

84. Li, A. and V. Daggett, *Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2.* Proceedings of the National Academy of Science, 1994. **91**: p. 10430-10434.

85. Ladbury, J.E. and B.Z. Chowdhry, *Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions.* Chemistry & Biology, 1996. **3**: p. 791-801.

86. Wang, W., *Instability, stabilisation and formulation of liquid protein pharmaceuticals.* International Journal of Pharmaceuticals, 1999. **185**: p. 129-188.

87. Robic, S., et al., *Role of residual structure in the unfolded state of a thermophilic protein.* Proceedings of the National Academy of Sciences, 2003. **100**(20): p. 11345-11349.

88. Patra, A., et al., *Optimization of Inclusion Body Solubilization and Renaturation of Recombinant Human Growth Hormone from Escherichia coli.* Protein Expression and Purification, 2000. **18**(2): p. 182-192.

89. Fandrich, M. and C. Dobson, *The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation.* The EMBO Journal, 2002. **21**: p. 5682-5690.

90. Militello, V., et al., *Aggregation kinetics of bovine serum albumin studied by FTIR spectroscopy and light scattering.* Biophysical Chemistry, 2003. **107**: p. 175-187.

91. Come, J.H., P.E. Fraser, and P.T. Lansbury, *A kinetic model for amyloid formation in the prion diseases: importance of seeding.* Proceedings of the National Academy of Sciences, 1993. **90**(13): p. 5959-5963.

92. Serio, T., et al., *Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant.* Science, 2000. **289**(5483): p. 1317-1321.

93. Dzwolak, W., et al., *Aggregation of bovine insulin probed by DSC/PPc calorimetry and FTIR spectroscopy.* Biochemistry, 2003. **42**: p. 11347-11355.

94. Beck, C., X. Siemens, and D.L. Weaver, *Diffusion-collision model study of misfolding in a four-helix bundle protein.* Biophysical Journal, 2001. **81**(6): p. 3105-3115.

95. Weiss, W., et al., *Nonnative protein polymers: structure, morphology, and relation to nucleation and growth.* Biophysical Journal, 2007. **93**(12): p. 4392-4403.

96. Griffith, J.S., *Nature of the Scrapie Agent: Self-replication and Scrapie.* Nature, 1967. **215**(5105): p. 1043-1044.

97. Bolton, D.C., M.P. McKinley, and S.B. Prusiner, *Identification of a protein that purifies with the scrapie prion.* Science (New York, N.Y.), 1982. **218**(4579): p. 1309-1311.

98. Andrews, J. and C. Roberts, *A Lumry–Eyring Nucleated Polymerization Model of Protein Aggregation Kinetics: 1. Aggregation with Pre-Equilibrated Unfolding.* The Journal of Physical Chemistry B, 2007. **111**(27): p. 7897-7913.

99. Vendruscolo, M., et al., *Structures and relative free energies of partially folded states of proteins.* Proceedings of the National Academy of Sciences, 2003. **100**(25): p. 14817-14821.

100. Uversky, V., et al., *Association of partially-folded intermediates of staphylococcal nuclease induces structure and stability.* Protein Science, 1999. **8**(1): p. 161-173.

101. Dill, K. and H. Chan, *From Levinthal to pathways to funnels.* Nature Structural & Molecular Biology, 1997. **4**(1): p. 10-19.

102. Weiss, W., T. Young, and C. Roberts, *Principles, approaches, and challenges for predicting protein aggregation rates and shelf life.* J. Pharm. Sci., 2009. **98**(4): p. 1246-1277.

103. Finke, J.M., et al., *Aggregation events occur prior to stable intermediate formation during refolding of interleukin 1beta.* Biochemistry, 2000. **39**(3): p. 575-583.

104. Speed, M., J. King, and D. Wang, *Polymerization mechanism of polypeptide chain aggregation.* Biotechnol. Bioeng., 1997. **54**(4): p. 333-343.

105. Andrews, J. and C. Roberts, *Non-Native Aggregation of α-Chymotrypsinogen Occurs through Nucleation and Growth with Competing Nucleus Sizes and Negative Activation Energies†.* Biochemistry, 2007. **46**(25): p. 7558-7571.

106. Pallitto, M.M. and R.M. Murphy, *A mathematical model of the kinetics of beta-amyloid fibril growth from the denatured state.* Biophysical Journal, 2001. **81**(3): p. 1805-1822.

107. Roberts, C., R. Darrington, and M. Whitley, *Irreversible aggregation of recombinant bovine granulocyte-colony stimulating factor (bG-CSF) and implications for predicting protein shelf life.* J. Pharm. Sci., 2003. **92**(5): p. 1095-1111.

108. Lomakin, A., et al., *Kinetic theory of fibrillogenesis of amyloid β-protein.* Proceedings of the National Academy of Sciences, 1997. **94**(15): p. 7942-7947.

109. Top, A., K. Kiick, and C. Roberts, *Modulation of Self-Association and Subsequent Fibril Formation in an Alanine-Rich Helical Polypeptide.* Biomacromolecules, 2008. **9**(6): p. 1595-1603.

110. Gsponer, J. and M. Vendruscolo, *Theoretical Approaches to Protein Aggregation.* Protein and Peptide Letters, 2006. **13**(3): p. 287-293.

111. Li, Y. and C. Roberts, *Lumry–Eyring Nucleated-Polymerization Model of Protein Aggregation Kinetics. 2. Competing Growth via Condensation and Chain Polymerization.* The Journal of Physical Chemistry B, 2009. **113**(19): p. 7020-7032.

112. Krebs, M.R.H., et al., *The formation of spherulites by amyloid fibrils of bovine insulin.* Proceedings of the National Academy of Science, 2004. **101**(40): p. 14420-14424.

113. Krebs, M.R.H., K.R. Domike, and A.M. Donald, *Protein aggregation: more than just fibrils.* Biochemistry Society Transactions, 2009. **37**(4): p. 682-686.

114. Rogers, S.S., et al., *Optical microscopy of growing insulin amyloid spherulites on surfaces in vitro.* Biophysical Journal, 2006. **90**: p. 1043-1054.

115. Domike, K.R. and A.M. Donald, *Thermal dependence of thermally induced protein spherulite formation and growth: kinetics of B-lactoglobulin and insulin.* Biomacromolecules, 2007. **8**: p. 3930-3937.

116. Bromley, E.H.C., M.R.H. Krebs, and A.M. Donald, *Mechanisms of structure formation in particulate gels of B-lactoglobulin formed near the isoelectric point.* The European Physical Journal E, 2006. **21**: p. 145-152.

117. Bromley, E.H.C., M.R.H. Krebs, and A.M. Donald, *Aggregation across the length-scales in B-lactoglobulin.* Faraday Discussions, 2004. **92**: p. 1336-1342.

118. Krebs, M., G. Devlin, and A. Donald, *Amyloid Fibril-Like Structure Underlies the Aggregate Structure across the pH Range for β-Lactoglobulin.* Biophysical Journal, 2009. **96**(12): p. 5013-5019.

119. Krueger, S., D. Ho, and A. Tsai, *Misbehaving Proteins*, in *Small-Angle Neutron Scattering as a Probe for Protein Aggregation at Many Length Scales*, R.M. Murphy and A. Tsai, Editors. 2006, Springer.

120. Hiemenz, P. and R. Rajagopalan, *Principles of Colloid and Surface Chemistry (Undergraduate Chemistry Series).* 1997: CRC Press.

121. Gabrielson, J., et al., *Quantitation of aggregate levels in a recombinant humanized monoclonal antibody formulation by size-exclusion chromatography, asymmetrical flow field flow fractionation, and sedimentation velocity.* J. Pharm. Sci., 2007. **96**(2): p. 268-279.

122. Grillo, A., et al., *Conformational Origin of the Aggregation of Recombinant Human Factor VIII†.* Biochemistry, 2000. **40**(2): p. 586-595.

123. Ahrer, K., et al., *Analysis of aggregates of human immunoglobulin G using size-exclusion chromatography, static and dynamic light scattering.* Journal of Chromatography A, 2003. **1009**(1-2): p. 89-96.

124. Arakawa, T., et al., *Aggregation Analysis of Therapeutic Proteins, Part 2.* Bioprocess International, 2007: p. 36-47.

125. Aymard, P., et al., *Static and Dynamic Scattering of β-Lactoglobulin Aggregates Formed after Heat-Induced Denaturation at pH 2.* Macromolecules, 1999. **32**(8): p. 2542-2552.

126. Gabrielson, J., et al., *Common excipients impair detection of protein aggregates during sedimentation velocity analytical ultracentrifugation.* J. Pharm. Sci., 2009. **98**(1): p. 50-62.

127. Giddings, J.C., *Field-flow fractionation: analysis of macromolecular, colloidal, and particulate materials.* Science, 1993. **260**(5113): p. 1456-1465.

128. Liu, J., J. Andya, and S. Shire, *A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation.* The AAPS Journal, 2006. **8**(3): p. E580-E589.

129. Bence, N., R. Sampat, and R. Kopito, *Impairment of the Ubiquitin-Proteasome System by Protein Aggregation.* Science, 2001. **292**(5521): p. 1552-1555.

130. Sharma, D., et al., *Micro-flow imaging: flow microscopy applied to sub-visible particulate analysis in protein formulations.* The AAPS Journal, 2010. **12**(3): p. 455-464.

131. Nielsen, L., et al., *Studies of the structure of insulin fibrils by Fourier transform infrared (FTIR) spectroscopy and electron microscopy.* J. Pharm. Sci., 2001. **90**(1): p. 29-37.

132. Salgın, S., S. Takaç, and T. Özdamar, *A Parametric Study on Protein‐Membrane‐Ionic Environment Interactions for Membrane Fouling.* Separation Science and Technology, 2005. **40**(6): p. 1191-1212.

133. Hawe, A., M. Sutter, and W. Jiskoot, *Extrinsic Fluorescent Dyes as Tools for Protein Characterization.* Pharmaceutical Research, 2008. **25**(7): p. 1487-1499.

134. Roy, S., et al., *Temperature Dependence of Benzyl Alcohol- and 8-Anilinonaphthalene-1-sulfonate-Induced Aggregation of Recombinant Human Interleukin-1 Receptor Antagonist†.* Biochemistry, 2006. **45**(12): p. 3898-3911.

135. Sutter, M., et al., *Sensitive Spectroscopic Detection of Large and Denatured Protein Aggregates in Solution by Use of the Fluorescent Dye Nile Red.* Journal of Fluorescence, 2007. **17**(2): p. 181-192.

136. Uversky, V., et al., *Prediction of the association state of insulin using spectral parameters.* J. Pharm. Sci., 2003. **92**(4): p. 847-858.

137. Cox, A.J., A. Deweerd, and J. Linden, *An experiment to measure Mie and Rayleigh total scattering cross sections.* American Journal of Physics, 2002. **70**(6): p. 620-625.

138. Chen, K., *Nanoparticle sizing with a resolution beyond the diffraction limit using UV light scattering spectroscopy.* Optics Communications, 2003. **228**(1-3): p. 1-7.

139. Belton, D.J. and A.F. Miller, *Thermal Aggregation of Protective Antigen Monitored by Light Scattering Spectroscopy*, in *8th World Congress of Chemical Engineering (WCCE8)*. 2009: Palais des Congrès de Montréal.

140. Murphy, R.M. and C.C. Lee, *Laser Light Scattering as an indispensible tool for probing protein aggregation*, in *Misbehaving Proteins - Springer ebook*. 2006.

141. Roe, R.-J., *Methods of X-ray and Neutron Scattering in Polymer Science (Topics in Polymer Science)*. 2000: Oxford University Press, USA.

142. Lipfert, J. and S. Doniach, *Small-Angle X-Ray Scattering from RNA, Proteins, and Protein Complexes.* Annual Review of Biophysics and Biomolecular Structure, 2007. **36**(1): p. 307-327.

143. Jacques, D. and J. Trewhella, *Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls.* Protein Science, 2010. **19**(4): p. 642-657.

144. Makowski, L., et al., *Characterization of Protein Fold by Wide-Angle X-ray Solution Scattering.* Journal of Molecular Biology, 2008. **383**(3): p. 731-744.

145. Barbosa, L., et al., *The importance of protein-protein interactions on the pH-induced conformational changes of bovine serum albumin: a small-angle X-ray scattering study.* Biophysical Journal, 2010. **98**(1): p. 147-157.

146. Semisotnov, G., *Protein Globularization During Folding. A Study by Synchrotron Small-angle X-ray Scattering.* Journal of Molecular Biology, 1996. **262**(4): p. 559-574.

147. Kataoka, M., et al., *Molten Globule of Cytochrome c Studied by Small Angle X-ray Scattering.* Journal of Molecular Biology, 1993. **229**(3): p. 591-596.

148. Ueki, T., et al., *Aggregation of bovine serum albumin upon cleavage of its disulfide bonds, studied by the time-resolved small-angle x-ray scattering technique with synchrotron radiation.* Biophysical Chemistry, 1985. **23**(1-2): p. 115-124.

149. Yong, W., et al., *Structure determination of micelle-like intermediates in amyloid β-protein fibril assembly by using small angle neutron scattering.* Proceedings of the National Academy of Sciences, 2002. **99**(1): p. 150-154.

150. Vogtt, K., et al., *Tracing nucleation pathways in protein aggregation by using small angle scattering methods.* Soft Matter, 2011. **7**(8): p. 3906-3914.

151. Murayama, K. and M. Tomida, *Heat-Induced Secondary Structure and Conformation Change of Bovine Serum Albumin Investigated by Fourier Transform Infrared Spectroscopy.* Biochemistry, 2004. **43**(36): p. 11526-11532.

152. Militello, V., V. Vetri, and M. Leone, *Conformational changes involved in thermal aggregation processes of bovine serum albumin.* Biophysical Chemistry, 2003. **105**(1): p. 133-141.

153. Navarra, G., et al., *Influence of metal ions on thermal aggregation of bovine serum albumin: Aggregation kinetics and structural changes.* Journal of Inorganic Biochemistry, 2009. **103**(12): p. 1729-1738.

154. Laligant, A., et al., *Surface hydrophobicity and aggregation of .beta.-lactoglobulin heated near neutral pH.* Journal of Agricultural and Food Chemistry, 1991. **39**(12): p. 2147-2155.

155. Vetri, V. and V. Militello, *Thermal induced conformational changes involved in the aggregation pathways of beta-lactoglobulin.* Biophysical Chemistry, 2005. **113**(1): p. 83-91.

156. Casal, H., U. Köhler, and H. Mantsch, *Structural and conformational changes of β-lactoglobulin B: an infrared spectroscopic study of the effect of pH and temperature.* Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1988. **957**(1): p. 11-20.

157. Matsuura, J. and M. Manning, *Heat-Induced Gel Formation of .beta.-Lactoglobulin: A Study on the Secondary and Tertiary Structure As Followed by Circular Dichroism Spectroscopy.* Journal of Agricultural and Food Chemistry, 1994. **42**(8): p. 1650-1656.

158. Hamada, D., S.-i. Segawa, and Y. Goto, *Non-native α-helical intermediate in the refolding of β-lactoglobulin, a predominantly β-sheet protein.* Nature Structural & Molecular Biology, 1996. **3**(10): p. 868-873.

159. Forge, V., et al., *Is folding of β-lactoglobulin non-hierarchic? intermediate with native-like β-sheet and non-native α-helix.* Journal of Molecular Biology, 2000. **296**(4): p. 1039-1051.

160. *The RSCB Protein Data Bank.* Available from: http://www.rcsb.org/pdb. Accessed Jan 2012

161. Epand, R. and H. Scheraga, *The influence of long-range interactions on the structure of myoglobin.* Biochemistry, 1968. **7**(8): p. 2864-2872.

162. Meersman, F., L. Smeller, and K. Heremans, *Comparative Fourier transform infrared spectroscopy study of cold-, pressure-, and heat-induced unfolding and aggregation of myoglobin.* Biophysical Journal, 2002. **82**(5): p. 2635-2644.

163. Yan, Y.-B., et al., *Two-dimensional infrared correlation spectroscopy study of sequential events in the heat-induced unfolding and aggregation process of myoglobin.* Biophysical Journal, 2003. **85**(3): p. 1959-1967.

164. Eliezer, D., et al., *Evidence of an associative intermediate on the myoglobin refolding pathway.* Biophysical Journal, 1993. **65**(2): p. 912-917.

165. Chiba, K., et al., *Kinetic study on myoglobin refolding monitored by five optical probe stopped-flow methods.* Proteins, 1994. **19**(2): p. 110-119.

166. Bismuto, E., I. Sirangelo, and G. Irace, *Salt-induced refolding of myoglobin at acidic pH: Molecular properties of a partly folded intermediate.* Archives of Biochemistry and Biophysics, 1992. **298**(2): p. 624-629.

167. Lyubovitsky, J., H. Gray, and J. Winkler, *Mapping the Cytochrome c Folding Landscape.* Journal of the American Chemical Society, 2002. **124**(19): p. 5481-5485.

168. Weinkam, P., et al., *The Folding Energy Landscape and Free Energy Excitations of Cytochrome c.* Accounts of Chemical Research, 2010. **43**(5): p. 652-660.

169.   Segel, D., et al., *Transient Dimer in the Refolding Kinetics of Cytochrome c Characterized by Small-Angle X-ray Scattering†.* Biochemistry, 1999. **38**(46): p. 15352-15359.

170.   Anfinsen, C.B., *ADVANCES IN PROTEIN CHEMISTRY VOL 29, Volume 29 (v. 29).* Vol. 29. 1975: Academic Press.

171.   Claudia, T., et al., *Analysis of RNase A refolding intermediates by electrospray/mass spectrometry.* FEBS Letters, 1994. **352**(3): p. 301-306.

Chapter 3 **: Theory**

3.1 Introduction

This chapter contains details of the experimental techniques carried out on this project coupled with a justification for their use and underpinning theory. The techniques included here are: fluorescence anisotropy, small angle neutron scattering and ultra violet light scattering spectroscopy.

3.2 Fluorescence Anisotropy

Fluorescence anisotropy uses polarised light to probe the rotational motion of protein fluorophore side groups. The protein is subjected to light polarised in the vertical plane, before the emission of horizontally polarised light from the protein is measured. This technique can be used to monitor whether fluorescent side chains are compact within the protein or exposed and mobile. Compaction of side chains is one of the latter stages in the protein folding process [1] therefore this technique will indicate the extent of mobility present in the latter stages of the structural arrangement.

Fluorophores will only absorb light at specific wavelengths. In respect to protein fluorophores, there are two main residues; tryptophan and tyrosine. These two have separate peak absorbances at different wavelengths; the optimum for tyrosine is 276 nm, whilst for tryptophan it is 280 nm [2]. The overall range over which these absorb overlap, but selecting a wavelength greater than 276 nm will excite almost exclusively the tryptophan residues. This property is desirable as the tryptophan is a hydrophobic residue, therefore when the protein is in its folded state it will be buried inside the hydrophobic core of the protein and therefore is less mobile. An additional useful characteristic of probing for this residue is seen by the fact that tryptophan fluorescence is shifted by up to 20 nm when buried within the protein as compared to on the surface of a protein [3], which can be studied directly by fluorescence spectroscopy.

When the vertically polarised excitation light strikes the protein sample, only the fluorophores situated in the same plane as that of the light beam will absorb photons [4]. There is a significant period of time between the photon of light being absorbed by the fluorophore, and a photon being emitted at a larger wavelength. During this period of time, if there is movement within the protein fluorophore, it will alter the plane in which the emitted photon is released. If there is no movement within the side group, the photon will be emitted in the same plane in which it was excited, in this case the vertical plane.
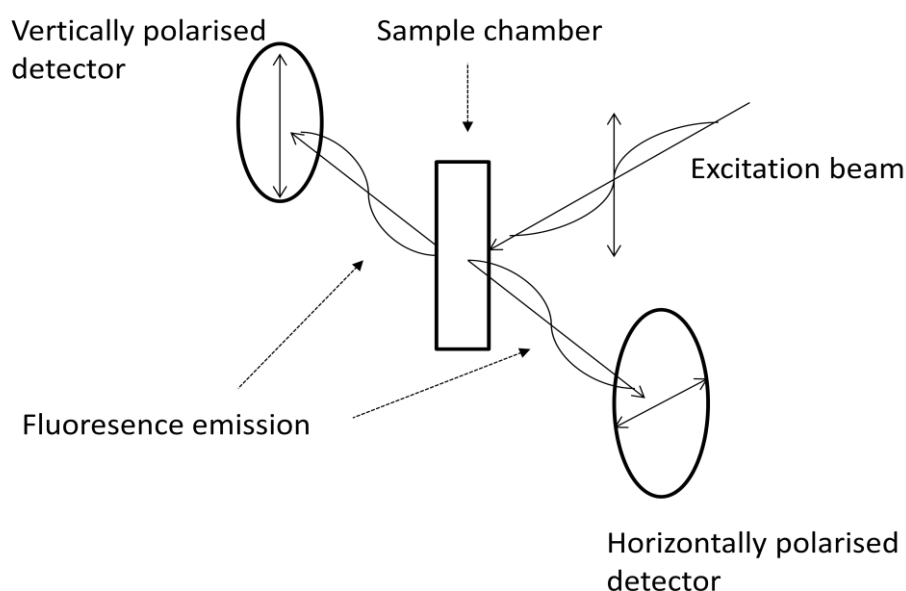
Figure 3.1: Planes of polarisation for beam excitation and emission detection within the fluoresence anisotropy setup

The fluorescence anisotropy setup utilises two photomultipliers (Figure 3.1), a separate photomultiplier for detecting the individual planes: vertically, and horizontally polarised light [5]. This allows for detection of the light emitted in the same plane as the excitation beam, and light whose polarisation has been changed through mobility of the tryptophan residues. This is the basis of the fluorescence anisotropy theory used to study the structural aspects of the protein. The fluorescence anisotropy is calculated using Equation 3.1;

$$r = \frac{I_V - I_H}{I_V + 2I_H}$$

Equation 3.1: Defining fluorescence anisotropy

Where the fluorescence anisotropy r is calculated from: I_v; the fluorescence intensity of the vertically polarised emission and I_H ; the fluorescence intensity of the horizontally polarised emission. This ratio gives an anisotropy value, where higher values of r relate to a more stable fluorophore, and lower r values relate to fluorophores which are more mobile [6]. This can be read on the millisecond timescale, therefore hundreds of readings can be taken over the course of a second, and these measurements are taken for many minutes, which is ideal for both the rapid dynamic process and the slower structural rearrangements that are seen in protein folding.

3.3 Small Angle Scattering

There is a large amount of literature present on the theory and principles behind small angle scattering [7-10]. This section aims to cover an overview of the general theories can be applied within this work on proteins and protein aggregates.

Neutrons are uncharged particles with a mass of 1.678×10^{-24}g and a spin of ½. The uncharged nature of the neutron allows it to interact deep within the scattering matter, with the nuclei of atoms through the strong nuclear force [7]. The spin of the neutron allows it to be used to probe magnetic properties of the scattering structure, however this use in not covered within this body of work.

Unlike X-rays which have energies of around 10keV, neutrons have energies of around 10meV. This difference in properties means that when X-rays are scattered there is a minimal energy change, however inelastic scattering of neutrons delivers a measurable

change in energies. The use of the small angle techniques here covers only the elastic scattering from the particles.

The strength with which each atom scatters the neutron does not vary linearly with atomic number (Figure 3.2). The nucleus of an atom occupies only a fraction of the total atom volume, therefore the majority of neutrons may pass through without interaction. Therefore a large number of neutrons are required to produce significant measurable interaction with the matter. For this reason, large scale facilities are required to produce the flux of neutrons capable of creating measurable data sets within a satisfactory timescale. Thermal neutrons are used in small angle experiments and generally have a wavelength of between 1-6 Å. This length scale is of similar magnitude to the separation distance between atoms in matter, and so is ideal to probe detailed structure.
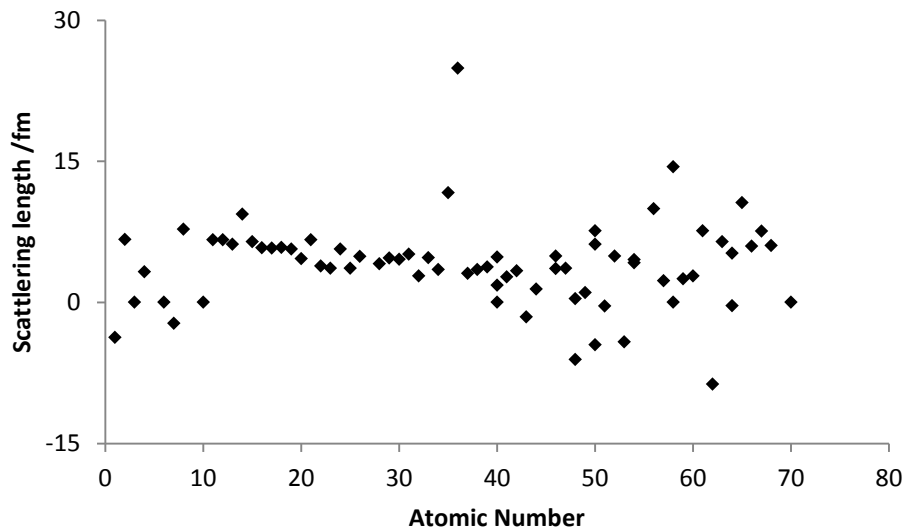
Figure 3.2: Scattering length value against atomic number. Values include isotopes.[11]

Neutrons are produced either by nuclear fission in a reactor or spallation. Examples of both are methods of production at large scale sources are; the Instituit Laue-Langevin,(ILL) in Grenoble, France as a source of neutrons provided by fission and

ISIS at the Rutherfod Appleton Laboratory, Oxfordshire, UK for neutrons produced by spallation.

In a spallation facility, protons are extracted from a hydrogen plasma source and accelerated by a linear accelerator into a larger circuit of bending magnets. Here, the bending magnets accelerate the ions to 70 MeV before being directed towards a heavy metal target. The nuclei in the atoms of the heavy metal target can absorb protons which will result in each nuclei emitting a neutron. Neutrons emitted by the target are slowed by moderators in order to achieve a suitable wavelength i.e thermal neutrons with a wavelength around the size of 1 Å.

Nuclear fission facilities work from the fission of a radioactive element such as uranium 235. Within the nuclear reactor, the uranium undergoes radioactive decay releasing neutrons which propagate the further decay of the isotope. The fission process is mediated through the use of boron control rods which absorb neutrons to limit the propagation of the radioactive decay. As in spallation, the neutrons emitted from the reactor are slowed by moderators in order to obtain thermal neutrons of useable energy.

Small angle scattering equipment is designed to treat the flow of photons or neutrons in order to maximise the flux at the sample whilst maintaining accuracy in the beam in order to correctly calculate the scattering angle, *2θ*. This essentially involves removing any divergence from the beam by passing the beam through collimation and ensuring the desired wavelength and size of the beam is directed onto the sample (Figure 3.3).

As small angle scattering probes length scales of matter in the range of above 10 Å, we can see using Braggs law [12] that the scattering from the beam will be present at a low angle near to the un-scattered beam (Equation 3.2).

$$\sin\theta = \frac{\lambda}{2d}$$

Equation 3.2: Braggs law

Where *d* represents the repetitive length scales within the structure. As the angles at which the scattering is detected are so low, it is imperative that there is no divergence from the beam; which would mask the scattering produced from the sample and therefore collimation is of great importance. This further shows why such high flux sources are required for these experiments as the accurate collimation reduces the intensity of the beam.

The scattered beam which hits the detector is produced from the volume of the sample which is irradiated by the beam. The scattering is a result of variations in the "density" of the sample. As X-rays and neutrons are scattered from different components of the atom, these "density" attributes vary between SAXS and SANS techniques, allowing each to probe different features of the structure. The scattering length density is responsible for neutron scattering.
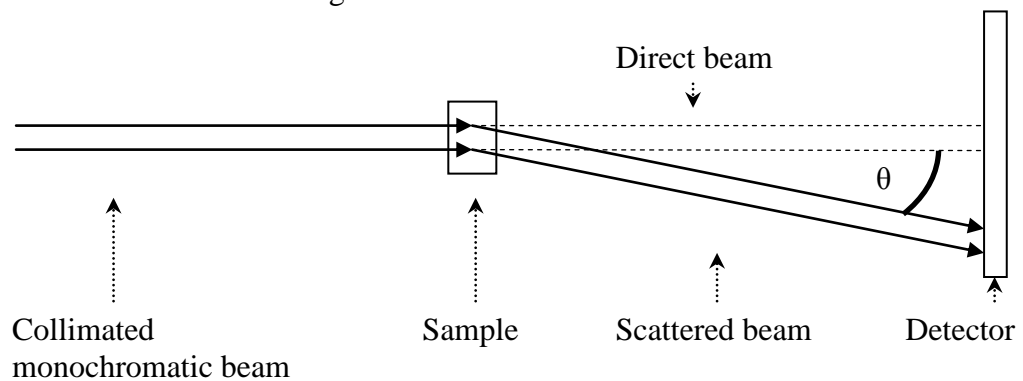
Figure 3.3: Overview of a scattering experiment.

In small angle scattering, some assumptions are applied to the scattering from the sample in order to extract information from the scattering pattern. The first is that the sample is of a low enough concentration that the scattered beam is the product of a single scattering interaction, and that none of the beam is scattered from multiple scattering points. The second is that there is neither a gain nor loss in the frequency in

the beam i.e the scattering is elastic, and that the incident beam suffers no distortion from the sample.

Examining Figure 3.4, it can be seen that the scattered neutrons at wave vector k_s originating from beam k_i (of which the majority of neutrons pass through the scattering medium) land on the detector at solid angle $d\Omega$. Figure 3.5 and Equation 3.3 illustrate how the scattering vector, q can be ascertained from this. Models of the neutron scattering are given as a function of q.
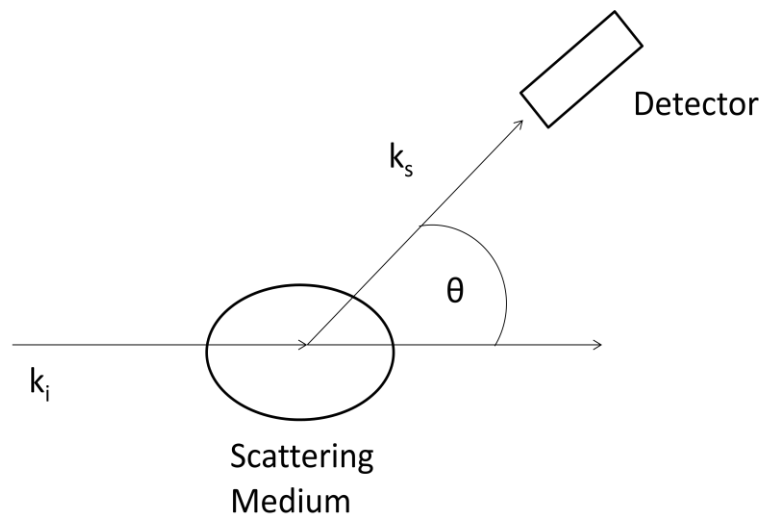
Figure 3.4: The scattering geometry of a SANS experiment. A detector at solid angle dΩ counts the number of neutrons scattered. Adapted from [9].
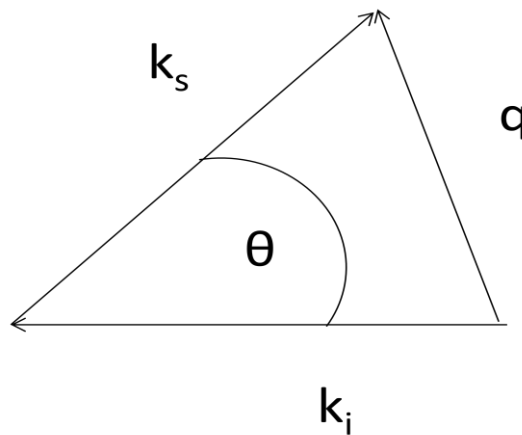
Figure 3.5: Vector diagram for calculation of scattering vector [9].

$$q = k_s - k_i$$

Figure 3.6: Equation for definition of q vector.

The distribution of the scattering, as measured on the detector, is given in terms of the scattering vector q, which relates both the wavelength λ of the incident beam, and the scattering angle by Equation 3.3.

$$q = \frac{4\pi}{\lambda} \frac{\sin \theta}{2}$$

Equation 3.3: Definition of the scattering vector

The scattering intensity measured from the detector can be related by the sum of the scattering from individual particles which are a product of scattering length; the distribution of material within a particle, and the phase factor determined by the position of the particle in the sample.

The instrument used in these experiments was D22 at the ILL. This is a small angle neutron scattering station which is designed for the study of biological macromolecules. Due to the large amount of flux (1.2×10^8 n cm^{-2} s^{-1} at the sample) produced by the reactor and the 55x40 mm beam cross section, it is ideal for studying dynamic samples. The wavelength range it is available to select is between 4.5-40 Å. Detector distance can range from 1.4 m to 17.6 m, and the detector has an area of 102.4 cm by 98 cm, allowing for a large q scale to be obtained (4×10^{-4} Å^{-1} to 0.44 A^{-1}). Pixels on the detector are 8x8 mm in size which allows for a high resolution in q vector to be studied.
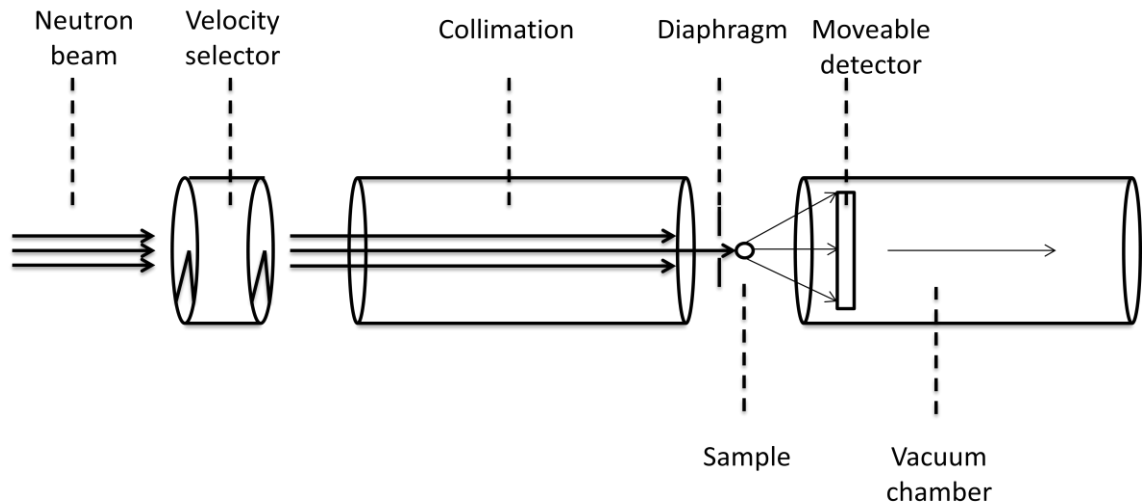
Figure 3.7: Diagram of D22 instrument small angle neutron scattering beamline [13]

The scattering from a solution; *I(q)*, can be described in terms of both the form factor *P(q)*; describing the shape of the particles, and the structure factor *S(q)*; describing the scattering arising from the interactions between these particles (Equation 3.4) .

$$I(q) = (\Delta\rho)^2 v^2 P(q)S(q)$$

Equation 3.4: Model of the total scattering from a solution.

Where $\Delta\rho$ is the difference in scattering length density between the protein and the surrounding medium (buffer) and v is the volume of sample illuminated by the beam. In this study of protein aggregates, the concentrations were kept low enough so that the effect of structure factor was negligible (any scattering arising from this would be inside the beamstop) therefore *S(q)*=1.

This can be simplified for dilute solutions as the position of the scatterer within the sample can be anywhere and is unaffected by the position of other particles and in the dilute system the average scattered intensity is the sum of the averaged intensities by the scattered particles. For identical particles the scattering can be calculated solely as the form factor *P(q)* of the particles.

As the scattering produced on the detector is a result of all matter in the path of the beam, i.e the sample holder and the buffer, it is required to make background subtractions in order to analyse solely the scattering coming from the sample. This is simply performed by measuring the sample cuvette and buffer for the same period of time as the sample, and performing a subtraction of the resultant *I* versus *q* plots. Alternatively the intensity from background samples can be scaled to match the intensities of the sample plots at high *q*, where there is no contribution of scattering from the sample.

The scattering length density ρ, of the protein is calculated from b_i (Equation 3.5); the coherent scattering length of the individual atoms, which can be seen in Figure 3.2.

$$\rho = \frac{\sum_i x_i b_i}{v}$$

Equation 3.5: Scattering length density calculation.

Where *i* is the number of identical atoms, *x* the number of atoms, and *v* the molecular volume of the protein [14]. This can also be calculated experimentally by monitoring the scattering from a protein in solutions with different H_2O:D_2O ratios. These solutions would have a known scattering length density (SLD), therefore the SLD of the protein could be obtained. This scattering length is not dependant on the wavelength of the scattering neutrons. For the scattering length of electrons the scattering length b_i can be substituted for scattering length of the number of electrons within the atom, therefore this value increases with increasing size of the atom. The scattering lengths of individual atoms for both neutrons and electrons are well documented [15].

For the homogenous spheres of radius *R*, the form factor *P(q)* can be related to scattering vector by Equation 3.6.

$$P(q) = \left[\frac{3}{qR^3} \left(\sin qR - qR \cos qR\right)\right]^2$$

Taking these known equations for the form factor of a sphere, a novel equation can be created for aggregates. In the case of the protein aggregate system, there are two populations of spheres; the large aggregates and the monomer. In order to model the form factor of this, the volume and concentration of the monomer have to be taken into account, and also these same values for the aggregate (these had to be expressed as a product of the two for fitting purposes, as neither value is known).

$$P(q) = \left(\left(V_m C_m \left[\frac{3}{qR_m{}^3} \left(\sin qR_m - qR_m \cos qR_m\right)\right]\right) + \left(V_a C_a \left[\frac{3}{qR_a{}^3} \left(\sin qR_a - qR_a \cos qR_a\right)\right]\right)\right)^2$$

Equation 3.7: Form factor of two sphere populations

Where V_m and V_a are the volumes of monomer and aggregate respectively, C_m and C_a are the volume fractions of both the aggregate and the monomer and R_m and R_a the radii.

The form factor of any homogenous particle can be estimated by assuming it to be made up of a number of small spherical particles. Using the form factor for a particle made up of spherical objects, which is averaged over all orientations in three dimensional space, the form factor can be expressed by the Guinier approximation (Equation 3.8):

$$P(q) = exp\left[-\frac{q^2 Rg^2}{3}\right]$$

Equation 3.8: Guinier equation

This is the Guinier approximation which gives a well defined measure of the object scattering [17]. This is defined in terms of the radius of gyration Rg, which describes the mean square mass distribution around its centre. The mean square radius of gyration

must be obtained from the scattering coming from low q (where $q \times Rg < 1.3$), as Equation 3.8 is an approximation of a Taylor series expansion where it is assumed q is very small (< 0.003 Å^{-1}). Furthermore, as the Guinier approximation is based on only the particle scattering, the sample must be sufficiently dilute as to not introduce inter particle scattering.
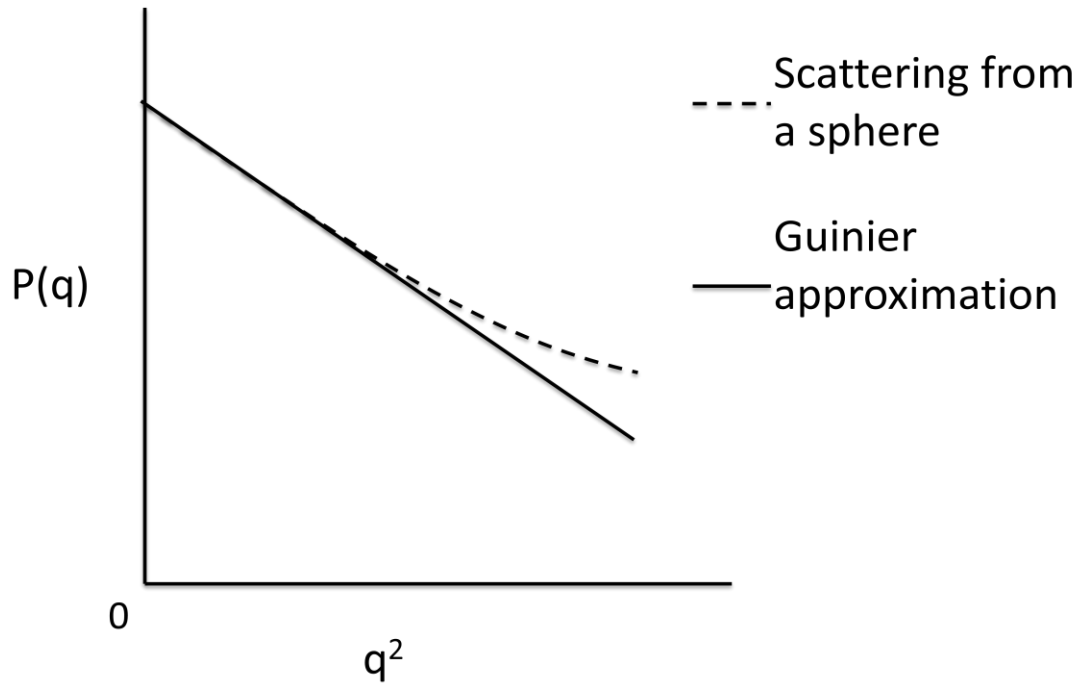
Figure 3.8: Accuracy of the Guinier approximation for scattering from dilute homogenous spheres. Values depend on normalisation and length scales, separation between the two plots begins at approximately $q \times Rg = 1.3$.

Additionally, for a sphere, the Rg can be related to the radius; R by Equation 3.9.

$$Rg = \sqrt{\frac{3}{5}} R$$

Equation 3.9: Relationship between radius of gyration for a sphere and physical radius.
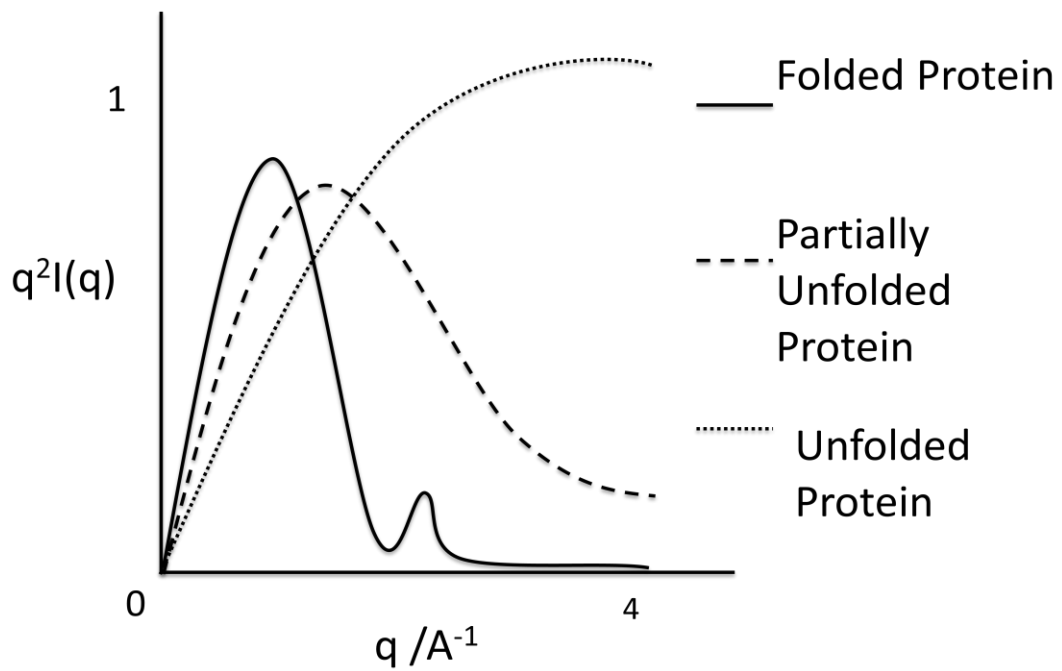
Figure 3.9: Illustration of Kratky plots for protein in different conformations from folded to unfolded. Adapted from [18]

Kratky plots offer a qualitative measure of the relative extent of folding or unfolding present within a protein sample (Figure 3.9). This is seen in the scattering as a small change and the small length scales and so is seen as a deviation in the scattering at the large q region. This change is small on the plot of *I versus q*, therefore using a Kratky plot of q^2I *versus q* will exaggerate any deviations seen at the higher q value. For a folded protein, there is a well defined peak in the Kratky plot (Figure 3.9) As the extent of unfolding within the protein increases, this peak will become less prominent as the intensity at high q increases. An unfolded protein will show no peak on a Kratky plot [18]. The Kratky plot measures the degree to which a molecule is compact, therefore it may not be able to differentiate between native and non-native protein conformations if they are equally compact.

The Porod law shows that for a sharp interface the intensity from scattering should decrease as q^{-4} for large q [8] (Equation 3.10), where S is the surface area of the matter and V the volume.

95

$$I(q) = \frac{2\pi(\Delta p^2)}{q^{-4}} \frac{S}{V}$$

Equation 3.10: Porod Equation

Equation 3.10 is limited to large q values, which can extend to infinity as the length scales from which the scattering occurs are uncorrelated to any other surface. This also holds for concentrated systems, as at large q it is only dealing with the surface of the particle. As the scattering intensity from a sphere is seen to decrease as q^{-4}, as is a thin rod at q^{-3} and a thin disk at q^{-2} [7].

In irregular shapes, this value can be considered as a fractal dimension, where values in the range between 1 and 3, the larger the value the more compact the structure is. At lower q values it is also possible to see evidence of surface fractals, where the scattering decrease is seen to occur at values between 2 and 3. A value of 2 represents a completely smooth surface, whereas a value of 3 would represent a totally rough, folded surface.

3.4 Ultra Violet Light Scattering Spectroscopy (UVLSS)

UVLSS is used to characterise the size of particles in the range of 50-1000 nm. Unlike either dynamic or static light scattering this technique utilises a range of wavelengths to determine the size of the particles.

Ultraviolet light backscattering intensity versus wavelength follows a power law (Equation 3.11).

$$Is \propto \lambda^{\beta(d)}$$

Equation 3.11: Relationship between back scattered intensity and particle diameter.

Equation 3.11 illustrates the wavelength dependence of the scattering is relative to the size of the particle scattering. I_s is the back scattered intensity, λ the wavelength and

$\beta(d)$ is the scattering exponent which is a function of the diameter, d. The relationship between β and d is covered by Mie theory [19], which determines the scattering from a single population of particle sizes within the sample. This generally requires that the sample be thin and of low concentration in order to prevent multiple scattering from occurring. However in a system such as the dynamic aggregation of proteins as studied here, such conditions are not possible where concentrated systems are required to instigate the protein aggregate formation. It has been shown though, that single scattering events can be calculated from multiple forward scattering values at absorbance values below 4.3 [20]. Therefore by monitoring the forward scattering across a range of wavelengths, it is possible to elucidate the size of protein aggregates over a time period.

In order to relate the experimentally obtained scattering exponent to a particle size, both the refractive index of the protein and of the solution is required. These values are wavelength dependent, and therefore refractive index value for the protein at each wavelength in the range is required. To calculate the refractive index for the protein, the Lorentz-Lorenz molar refraction [21] was used to sum the contributions to the refractive index from all residues and bonds within the molecule.

Table 3.1 (overleaf) lists the values used to calculate the refractive index of the protein for a series of wavelengths.

Table 3.1: Refractive increments for the amino acids and bonds present in the protein

Residue	Refractive increments for wavelength /cm^3 mol^{-1}			
	656.3	486.1	434.0	589.3
Glycine	1.092	1.115	1.121	1.100
Alanine	5.689	5.783	5.832	5.718
Valine	14.883	15.119	15.254	14.954
Leucine	19.480	19.787	19.965	19.572
Isolecine	19.480	19.787	19.965	19.572
Phenylalanine	29.593	30.343	30.788	29.825
Proline	13.791	14.004	14.133	13.854
Methionine	22.573	22.809	22.944	22.644
Aspartic Acid	13.499	13.819	13.991	13.605
Lysine	23.973	24.385	24.575	24.094
Glutamic Acid	18.096	18.487	18.702	18.223
Arginine	30.615	31.337	31.589	30.831
Serine	7.211	7.311	7.369	7.243
Asparagine	13.692	13.950	14.057	13.769
Cysteine	13.379	13.473	13.522	13.408
Threonine	11.808	11.979	12.080	11.861
Glutamine	19.975	20.442	20.654	20.120
Tryptophan	41.361	42.546	43.183	41.729
Tyrosine	31.115	31.871	32.325	31.350
Histidine	24.862	25.595	25.879	25.080
Backbone groups	13.363	13.740	13.897	13.482
N and C terminus	11.709	11.925	12.004	11.776
Water	3.706	3.758	3.779	3.725

The Lorentz-Lorenz molar refraction of the entire protein molecule R_{LL} can then be calculated as the sum of the molar refraction from individual residues R_i.

$$R_{LL} = \sum R_i$$

Equation 3.12: Lorentz-Lorenz summation

The R_{LL} value can then be related to the refractive index of the protein using the following equation.

$$R_{LL} = \frac{n^2 - 1}{n^2 + 2} \frac{M}{r}$$

Equation 3.13: Lorentz-Lorenz molar refraction equation.

Where M is the molecular weight of the protein (g/mol) and r is the density of the molecule (g/cm^3) and n is the refractive index.

The density of the protein was calculated using structural information from the Protein Data Bank [22]

$$\rho = \frac{M_w}{V A_n}$$

Equation 3.14: Density equation for proteins

Where A_n represents Avogadro's number, M_w is the molecular weight of the protein and V is the volume calculated using VADAR [23]. VADAR is a web service which takes the structural information of the protein from the PDB and calculates the volume of the molecule based on the position and angles of the residues within the protein. Below are listed the values for the volumes of proteins using VADAR.

Protein	PDB Accession Number	Volume Å^3
BSA	3V03	86654.65
β-lactoglobulin	1CJ5	21017.68
Horse heart myoglobin	1YMB	20876.5

Table 3.2: Volumes of protein calculated in VADAR.

The refractive index of each protein across a range of wavelengths was then determined. From this a plot of refractive index relative to wavelength can be made, and the Cauchy equation can be fitted to it [21]. The simplified form of the Cauchy equation shows:

$$n = A_c + \frac{Bc}{\lambda^2}$$

Equation 3.15: Cauchy equation

Equation 3.15 shows that a plot of *n versus $1/\lambda^2$* would give a straight line whose equation can be used to calculate the value of the refractive index at each wavelength in the range.
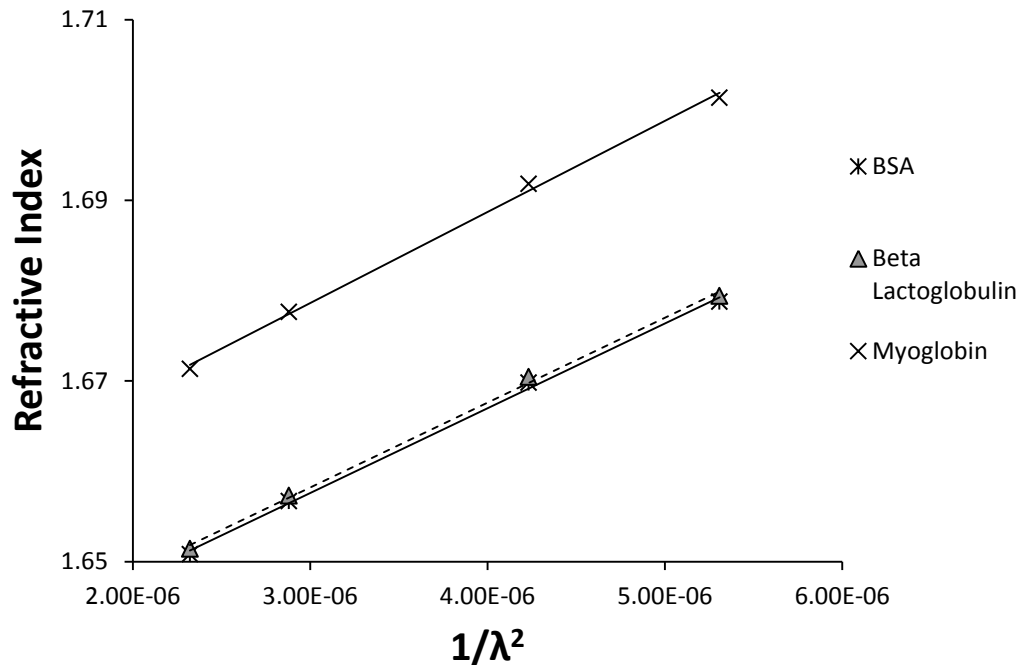
Figure 3.10 Wavelength dependence of refractive indices of candidate proteins. B-lactoglobulin and BSA show barely any difference at each wavelength value, but a slight difference in best fit lines. B-lactoglobulin best fit line is indicated by the dotted line.

Values for the refractive indices of water were obtained from the International Association for the Properties of Water and Steam, with the data obtained for water at 20 °C [24]. Furthermore this paper illustrates that the refractive index of water barely changes with increasing temperature, which is applicable to the methods described here. This can also be assumed to be the case for the protein over the same temperature range [25].

The refractive indices of the protein and the solution could now be used to populate a programme to calculate the values for the Mie theory solution to light scattering for a range of sphere sizes. The programme used is an adapted version of the programme CALLBHMIE [26]. This programme takes the values of the refractive index of the

protein, the medium, the wavelengths and the size of the spheres in a text file, and calculates the extinction efficiency for each value. The wavelengths covered the range used in the UVLSS experiments, depending on which protein was used, and the sphere range covered a size of 10-1000 nm. The programme then populated a text output file with values for extinction efficiency for each combination of wavelength and sphere size.

Extinction efficiency Q_{ext}, is related to the absorbance A by Equation 3.16, where ρ is the number of particles per unit volume and l is the sample pathlength with a being the sphere radius.

$$A = \frac{\rho l \pi a^2 Q_{ext}}{ln(10)}$$

Equation 3.16: Extinction efficiency equation [27]

Therefore, as the extinction efficiency is directly related to the absorbance, the theoretical scattering exponent can be obtained directly from the output of the CALLBHMIE programme. This allows for the theoretical sphere size to be related to the scattering exponent.
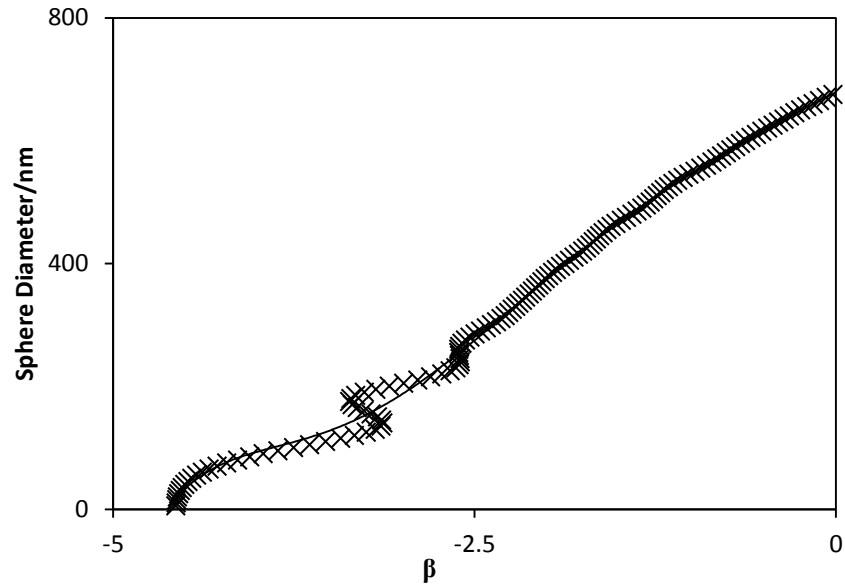
Figure 3.11: Theoretical diameter against scattering exponent for β-lactoglobulin.

A sixth order polynomial is fit to this graph (Figure 3.11) which gives a large R^2 value (>0.95) for each protein despite the deviation at low diameters. This has also shown good correlations with data obtained in the later sections. The fit within this plot is a sixth order polynomial which gives an equation to convert between sphere diameter and the scattering exponent β.

This theory relies on some assumptions, those mentioned here are addressed and related to the literature to show that they are fair. Further included in this theory is that the calculation of the number of monomers in an aggregate does not take into account and hydration levels within the aggregate. This may be a limitation of the technique mentioned here, however this may be improved by using the hydration levels of aggregates created by SANS experiment

3.5 References

1. Alberts, B., et al., *Molecular Biology of the Cell*. 4th ed. 2002: Garland Science.
2. Berg, J.M., J.L. Tymoczko, and L. Stryer, *3.1 Proteins Are Built from a Repertoire of 20 Amino Acids*, in *Biochemistry*. 2002, W.H.Freeman.
3. Vivian, J.T. and P.R. Callis, *Mechanisms of tryptophan fluorescence shifts in proteins*. Biophysical Journal, 2001. **80**: p. 2093-2109.
4. Mann, T.L. and U.J. Krull, *Fluorescence polariszation spectroscopy in protein analysis*. The Analyst, 2003. **128**: p. 313-317.
5. Photophysics, A., *Fluorescence Polarisation Section*, in *Applied Photophysics Stopped Flow Polarisation Manual*, Applied Photophysics.
6. Steiner, R.F., *Fluorescence Anisotropy: Theory and Applications*, in *Topics in Fluorescence Spectroscopy: Volume 2*, J.R.Lakowicz, Editor. 1991, Plenum Press: Baltimore.
7. Roe, R.-J., *Methods of X-ray and Neutron Scattering in Polymer Science (Topics in Polymer Science)*. 2000: Oxford University Press, USA.
8. Glatter, O., *{A new method for the evaluation of small-angle scattering data}*. Journal of Applied Crystallography, 1977. **10**(5): p. 415-421.
9. Lindner, P. and T. Zemb, *Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter (North-Holland Delta Series)*. 2002: North Holland.
10. Krueger, S., D. Ho, and A. Tsai, *Misbehaving Proteins*, in *Small-Angle Neutron Scattering as a Probe for Protein Aggregation at Many Length Scales*, R.M. Murphy and A. Tsai, Editors. 2006, Springer.
11. Sears, V., *Neutron scattering lengths and cross sections*. Neutron News, 1992. **3**(3): p. 26-37.
12. Myers, H.P., *Introductory Solid State Physics, Second Edition*. 1997: CRC Press.
13. *ILL D22 Information*. Available from: http://www.ill.eu/instruments-support/instruments-groups/instruments/d22/. Accessed Jan 2012
14. *Soft-Matter Characterization*. 2008: Springer.
15. Munter, A. *The National Institue for Standards and Technology Scattering Length Density Calculator*. 2012; Available from: http://www.ncnr.nist.gov/resources/sldcalc.html. Accessed Jan 2012
16. Placzek, G., *The Rayleigh and Raman Scattering*. 1959: Lawrence Radiation Laboratory.
17. Rice, S., *Small angle scattering of X-rays. A. Guinier and G. Fournet. Translated by C. B. Wilson and with a bibliographical appendix by K. L. Yudowitch. Wiley, New York, 1955. 268 pp. $7.50*. J. Polym. Sci., 1956. **19**(93): p. 594-594.
18. Mertens, H. and D. Svergun, *Structural characterization of proteins and complexes using small-angle X-ray solution scattering*. Journal of Structural Biology, 2010. **172**(1): p. 128-141.
19. Chen, K., et al., *Nanoparticle sizing with a resolution beyond the diffraction limit using UV light scattering spectroscopy*. Optics Communications, 2003. **228**(1-3): p. 1-7.
20. Swanson, N.L. and D.B. Billard, *Multiple scattering efficiency and optical extinction*. Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics, 2000. **61**(4 Pt B): p. 4518-4522.

21.   *Physical Methods of Organic Chemistry (Techniques of Organic Chemistry)*:
         John Wiley & Sons Inc.
22.   *The RSCB Protein Data Bank*. Available from: http://www.rcsb.org/pdb
         Accessed Jan 2012.
23.   Willard, L., et al., *VADAR: a web server for quantitative evaluation of protein
         structure quality.* Nucleic Acids Research, 2003. **31**(13): p. 3316-3319.
24.   P., et al., *Refractive index of water and steam as function of wavelength,
         temperature and density.* Journal of Physical and Chemical Reference Data,
         1990. **19**: p. 677-717.
25.   Julio Atkins, P. and P. De, *Atkins' Physical Chemistry*: Oxford University Press.
26.   Draine, B.T. *FORTRAN Mie Theory computer code*. Available from:
         http://www.scattport.org/index.php/light-scattering-software/mie-type-
         codes/list/320-callbhmie Accessed Jan 2012.
27.   Bohren, C. and D. Huffman, *References.* Absorption and Scattering of Light by
         Small Particles, 2007: p. 499-519.

Chapter 4 : **Methods and Materials**

4.1 Ultra Violet Light Scattering Spectroscopy

Protein Solutions

Protein samples for both absorbance measurements and refolding experiments were made up similarly, but in separate buffers depending on application. Lyophilized protein samples were purchased from Sigma-Aldrich Co. and weighed out using a Precisa 120A 4 figure balance, and dissolved in the appropriate buffer. The protein samples and their supplier details are listed in Table 4.1.

Table 4.1:Table of studied proteins with associated details

Protein	Details	Sourced From	Product Code
Bovine Serum Albumin	Fravtion V approx 99 %	Sigma	H3059-109
B-Lactoglobulin	From Bovine Milk 90 %	Sigma	L3908
Myoglobin	From Equine Heart	Sigma	M1882

Buffers

For UVLSS protein aggregation experiments, a sodium phosphate based buffer, also containing sodium chloride, was made. The sodium phosphate buffer contains two elements; sodium phosphate monobasic (NaH_2PO_4), and sodium phosphate dibasic (Na_2HPO_4). These two elements have pH values of 5 and 10 respectively; therefore a combination of the two will result in a buffer of a pH between these values, depending on the proportions. Approximate volume values are given for the different pH buffers created. The purpose of the sodium chloride was to ensure that the protein aggregation occurred over a temperature range that could be performed within the experimental setup, as sufficient levels of sodium chloride increases a proteins propensity to

aggregate. The advantage of this buffer is that multiple pH ranges can be probed whilst the surrounding environment remains constant across experiments. The following buffer was assembled:

Table 4.2:Components of sodium phosphate buffer complete with sourcing details

Buffer	Components	Details	Sourced From	Product Code
Sodium Phosphate Buffer	100 mM NaH$_2$PO$_4$	Sodium Phosphate Monobasic	Sigma	S0751
	100 mM Na$_2$HPO$_4$	Sodium Phosphate Dibasic	Sigma	S0876
	100 mM NaCl	+80 %, reagent grade $\geq$ 98 %	Sigma	310166

Table 4.3:Approximate volumes of constituents for 1L sodium phosphate buffer.

pH	Volume NaH$_2$PO$_4$	Volume Na$_2$HPO$_4$
5.4	20	980
5.8	80	920
7.2	680	320

These reagents were dissolved in ultrapure water from the ELGA UHQ-PS Ultra pure water system, and the pH adjusted through the use of a SCHOTT probe to monitor the pH, adjusting the amount of each buffer component to obtain the desired pH.

UVLSS procedures

Absorbance spectra were obtained from the protein samples using a Shimadzu UV-1601 spectrophotometer between 320 nm and 420 nm for bovine serum albumin and β-lactoglobulin. Myoglobin, due to its red/brown colour absorbs light at these wavelengths, and therefore the aggregation was measured between 700-800 nm, which is still capable of monitoring aggregate growth. The protein samples were placed in a UV 10mm light path Hellma macro silica cuvette with a volume of 3.3 mL (Figure 4.1). The temperature of both the sample and the reference chamber was controlled through the use of a water heated temperature block, connected to a Grant GD150 5L water bath. These were connected using reinforced rubber tubing, and secured in place using jubilee clips around entry and exit nozzles. This set up allowed the protein sample to be heated from an initial 25 °C to approximately 95 °C. The temperature was monitored using a platinum HEL-705 RTD temperature sensor integrated into the rubber cuvette lid. This was calibrated so that it was sufficiently submerged in the sample to give an accurate temperature whilst not entering the window for UV absorbance measurement. This temperature sensor was then wired to a Pico PT100 data logger. Both this device and the spectrophotometer were connected to a PC that ran the Shimadzu UVProbe, Pico picolog PLW recorder, and Grant Labwise 1.0 water bath control software to enable data collection, and remote programming of the water bath temperature stages.
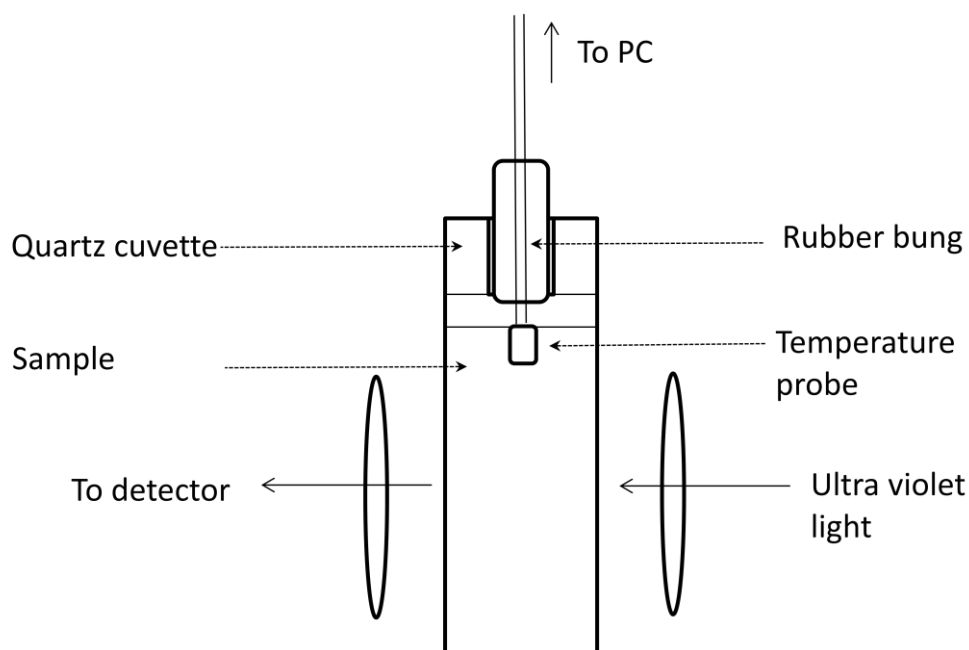
Figure 4.1:Sample setup for UVLSS inside a water bath controlled heating block.

The spectra were recorded between 420 nm and 320 nm at a medium scan rate, an entire scan taking 13.3 seconds to complete. UV probe software allowed for 100 scans to be taken when performing isothermal runs, before requiring a reset to continue measurement. To accommodate this; the scan intervals were adjusted according to the temperature being measured, in order obtain sufficient spectra to properly characterise the aggregate growth.

Between experimental runs, the protein sample and cuvette required meticulous cleaning to ensure no protein aggregate remained within the equipment. This was performed by rinsing with hot water, then soaking the equipment in hot detergent (Decon-90) this was then followed by rinsing with deionised water, and allowing the equipment to dry in ambient conditions.

Protein concentration: After initial sample preparation, the Shimadzu spectrophotometer was used to measure the protein absorbance at a single wavelength at 280 nm, to ensure the protein concentration was consistent between protein samples.

Temperature ramp: In order to ascertain the temperature range over which a protein aggregates, temperature ramp measurements were run. The Grant water bath control Labwise 1.0 was used to program a temperature ramp that would relate to a heating rate such as 0.1 ˚C/min. The Pico data logging software picolog PLW recorder was used to monitor the temperature within the cell at set time intervals, identical to that of the scan intervals from the Shimadzu. UV probe was utilised to give an absorbance value at 320 nm at set time intervals in "Kinetics" mode. These pieces of recording software were initiated simultaneously in order to relate the absorbance and temperature values.

Isothermal spectra: Isothermal spectra were collected using the "Spectrum" mode on the UV probe software. This recorded the sample cells absorbance between 320 nm and 420 nm at set time intervals. The Pico data logger was set to measure the temperature within the cuvette at time intervals equal to that of the intervals on the spectrophotometer. Each run started at 25 ˚C and was ramped up to the desired temperature at the heating rate of 2 ˚C/min.

4.2 Stopped Flow Refolding

The following section details the methods used for measurements of fluorescence anisotropy, absorbance a hydrophobic probe fluorescence of protein refolding in a stopped flow system.

Solutions

In protein refolding experiments of proteins listed in Table 4.4, the proteins were unfolded in 8 M urea (reagent grade 98 %, Sigma s30498-448) made up with ultra pure water.

Table 4.4:Refolding protein information complete with sourcing details.

Protein	Details	Source	Product Code
Cytochrome C	From bovine heart heart ≥95 % based on molecular weight	Sigma	C2037
β-lactoglobulin	From bovine milk ≥90% from PAGE	Sigma	L3908
α-lactalbumin	From bovine milk type III, calcium depleted ≥85 % from PAGE	Sigma	L6010
Ribonuclease A	From bovine pancrease type 1-A 20-100 Kunitz units/mg	Sigma	R5503

Each protein required an individual buffer to obtain a high standard of protein folding. The refolding buffer used in for each protein refolding experiment is listed in Table 4.5. Refolding buffers vary depending on the properties of the protein to be refolded. Most proteins require a favourable pH to aid refolding, additionally; some proteins require a salt concentration for to prevent aggregation. Cytochrome C aggregation assays are also performed with additional NaCl present within the buffer. In experiments where salt is present in the buffer this is stated. The constituents of each buffer are listed in Table 4.6, where the sodium phosphate buffer make up is identical to that stated in the UVLSS method section (Table 4.2).

Table 4.5:Refolding proteins complete with buffer pH and components

Protein	pH	Components
Cytochrome C	10.6	100 mM Tris
β-lactoglobulin	7.8	100 mM Tris, 10 mM NaCl
α-lactalbumin	7.8	100 mM Tris, 1 0mM NaCl
Ribonuclease A	7.8	100 mM NaPO4, 20 mM NaCl

Table 4.6: Buffer components and sourcing details.

Components	Details	Sourced From	Product Code
Tris	ACS Regenz	Sigma	S87203-289
NaH_2PO_4	Sodium Phosphate Monobasic	Sigma	S0751
Na_2HPO_4	Sodium Phosphate Dibasic	Sigma	S0876
NaCl	+80 %, reagent grade $\geq$ 98 %	Sigma	310166

Anisotropy Procedures

The stopped-flow fluorescence anisotropy measurements were taken using the Applied Photophysics SX-20 stopped flow fluorescence anisotropy instrument. This set up includes the power supply, bulb housing, beam width control, light polariser, sample cell, photomultipliers, stopped flow syringes, loading syringes, electronics unit, and PC. Drive syringes were installed into the device to give a desired refold ratio. Syringe sizes were 2.5 mL, 500 µL, 250 µL and 100 µL. This allowed for mixing ratios 1:25, 1:10, 1:5 and 1:2. The larger of these stopped flow syringes was loaded with the refolding buffer using a loading syringe, injected in from the top. The smaller was loaded with the protein sample in urea.

The device was calibrated for the individual protein, its buffer, and the wavelength at which the anisotropy would be measured; 280 nm or 297 nm. 280 nm was used to measure the signal fluorescence especially as CytC contains a single tryptophan residue. This involved ensuring the voltage of the photomultipliers was sufficient enough to give an accurate reading for the protein sample, zeroing any offset signals produced when no chromophore is present, and accounting for the G factor; the difference in gain between the two detectors.

Data was collected through the PC using the Pro-Data software. This software was run in "Kinetics" mode, using the "Fluorescence Polarisation" option for taking measurements. The source of the drive syringe was also determined using this program; manually or internally using the gas pump, depending on application. When using an internal trigger, nitrogen gas at a pressure of 8 bar was used to power the exit syringe, whilst a pressure of 4 bar was used to drive the separate mixing syringes.

Using the same stopped flow mixing setup, this equipment is also capable of measuring fluorescence and absorbance of the kinetics within the stopped flow cell. This is achieved by fitting an absorbance detector opposite the light source side of the cell, with a photomultiplier tube present at a right angle to the incoming light beam. Absorbance was measured using light emission at 320 nm. Fluorescence studies were carried out using the fluorescent probe ANS (40122-0050, Acros Organics, Fisher). ANS was present in solutions for fluorescence measurements in excess at 1×10^{-5} M. Fluorescent studies of the probe binding were carried out at 370 nm.

In experiments where mixing ratios were altered, the initial protein concentration was also altered to scale in order that the final concentration of protein was maintained. This is especially important in the aggregation studies in order that intensities be comparable between runs. In order for the final concentration of 0.77 mg/mL (w/v) protein to be maintained, initial solutions of protein were prepared following Table 4.7.

Table 4.7: Initial protein concentrations for each refold ratio.

Refold Ratio	Initial Protein Concentration (mg/mL)
1 to 5	4.6
1 to 10	8.46
1 to 25	20

4.3 Small Angle Neutron Scattering

The objective of this experiment was to use small angle neutron scattering to understand the formation and overall structure of spherical aggregates of bovine serum albumin (BSA). The following experiments were performed at the Instituit Laue-Langevin, Grenoble, France, on beamline D22; a small angle neutron scattering beamline.

In order to characterise the aggregation process, two categories of experiments were run: static measurements (where pre-aggregated samples were placed in the beam for analysis) and dynamic measurements (using the temperature control feature on the sample rack to aggregate samples and analyse the aggregation process in real-time).

Sample Creation

Static samples were created prior to the experiment using a 2 mg/mL concentration of BSA (sourced in Table 4.1) with 100mg of protein being measured into 50 mL of a buffer containing 100 mM Tris, 100 mM NaCl pH 7.4 (adjusted with HCl) buffer:

Table 4.8:Component list of Tris buffer for static experiments

200 mL H$_2$O Tris Stock Buffer	Reagent	Mass
	Tris(hydroxymethyl)aminomethane	2.4228g
	NaCl	1.1688g

Samples were then incubated at set temperatures for specific times to allow the aggregation process to be completed:

Temperature (°C)	Time (h:m)
70	7:00
72	3:30
74	2:00
76	1:30
78	0:55

After the designated duration, samples were then placed on ice, diluted by 75 % with D_2O (375 µL aggregate sample combined with 1125 µL of D_2O), and placed in 2 mm pathlength cells ready for insertion into the beamline.
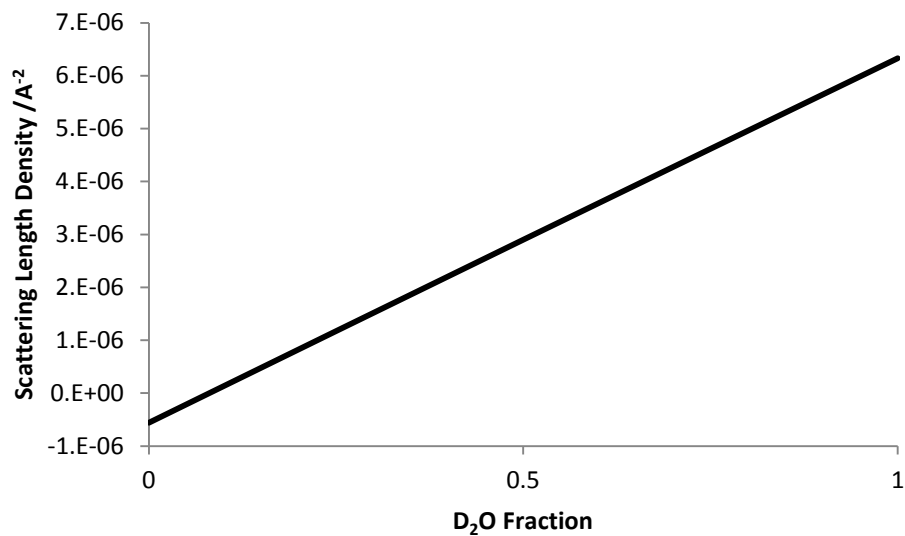
Figure 4.2:Scattering length density of buffer with varied H2O:D2O.[1]z

D_2O was added to the aggregate solution to ensure good contrast between the protein and the solution, whilst a high D_2O concentration is also required to reduce incoherent scattering and ensure each sample could be scanned in an appropriate time. The scattering length density of BSA was calculated at 2.97×10^{-6} $Å^{-2}$ which appears comparable with other calculations [2]. Using the data from Table 4.2 a D_2O fraction of 0.75 has a scattering length density of 4.62×10^{-6} $Å^{-2}$.

Dynamic samples were created using 2 mg/mL concentration of bovine serum albumin, starting with an initial stock solution of 20 mg/mL, created from 200 mg of BSA into 10 mL solution of ultrapure H_2O. This was then diluted into in a 100 mM Tris, 100 mM NaCl (sourced Table 4.4) , buffer made up in D_2O, pD7.4. This was diluted at a 1:9 ratio, with 70 μl of protein sample to 630 μl of buffer to give a final concentration of 2 mg/mL BSA in a 90 % D_2O buffer solution with a scattering length density of $5.62 \times 10^{-6} \text{ Å}^{-2}$.

Table 4.10: Component list of Tris buffer for dynamic experiments

	Reagent	Mass
50 mL D_2O Tris Stock Buffer	Tris(hydroxymethyl)aminomethane	0.6050g
	NaCl	0.2922g

This was put directly in quartz cells of 2 mm pathlength in preparation to be placed in the beam.

Data Collection

Data from static samples were collected at two separate camera lengths: 2 m (run for 10 min per sample) and 11.2 m (run for 20 min per sample). The collimation for the samples was the same distance as the camera length.

Dynamic samples were run at 3 camera lengths: 2 m, 4 m, and 11.2 m, with the collimation distance being identical to the camera length. Samples were run at 2 temperatures; 75 ˚C and 80 ˚C. At each temperature and camera length, the data was divided into 1 min increments. At 80 ˚C, this was performed for 60 minutes, and at 75 ˚C for 120 minutes to allow adequate time for aggregation to occur. In addition to the

aforementioned temperatures, a final run of 70 $^\circ$C was compiled with the detector distance set at 4 m. This was allowed to run overnight.

Over 1100 spectra were collected in total. GRASP [3] was used for data handling, including calculation of the beam centre, background and transmission corrections and radial averaging. For the dynamic runs this was completed for each individual spectrum, rather than an average of the set in order to analyse minute by minute changes associated with the aggregation process.

4.4 Rotor Gene Fluorescence

The Qiagen Rotor-Gene Q is primarily a real time PCR machine capable of rapidly and accurately heating samples whilst monitoring the UV absorbance or fluorescence of the sample. This also enables the Rotor-Gene Q to be used to monitor the unfolding of a protein with the use of a fluorescent probe which binds to the unfolded protein and fluoresces. Here the hydrophobic probe used was 1-anilinonaphthalene-8-sulfonic acid. This technique was used to specifically examine the thermal stability of bovine serum albumin in a 90 % D_2O solution. In order to achieve rapid and accurate heating, small volumes are required; therefore a PCR tube containing 100 μL of solution was used. In order to mimic the conditions of the dynamic SANS experiments, identical reagents were used to create a 90 % D_2O solution. This was created using 90 μL D_2O (Sigma), 5 μL of 300 μg/mL ANS (Fisher) and 200 μg of BSA (Sigma) to create a 2 mg/mL final protein concentration. This procedure was run in triplicate with a 90 % D_2O 15 μg/mL ANS containing no BSA used as the reference. The melt was carried out at 6 ºC/min between 35 ºC and 95 ºC, where the fluorescence was measured under the "Blue" channel which provides Excitation at 365 nm and detection at 460 nm, allowing the change in fluorescence of the probe to be detected.

4.5 Environment Scanning Electron Microscopy (ESEM)

Protein aggregate samples were measured using a Quanta 250 SEM in ESEM mode, using a Peltier stage for sample control and gaseous secondary electron detector (GSED) for detection. Aggregate samples produced using the UVLSS procedures were extracted from the cuvette after cooling to room temperature. Aggregate samples were placed into 2 mL eppendorf tubes and centrifuged at 14000 rpm for 60 s in an Eppendorf Minispin Plus centrifuge. Superfluous liquid was drained from the sample, and replaced with ultrapure water, and the aggregates were resuspended, before centrifuging again at 14000 rpm for 60 s. This process was performed three times to remove any salt from the sample, and prevent formation of salt crystals during microscopy. The aggregate sample was then transferred to an aluminium stub for loading onto the Peltier stage. The sample was allowed to equilibriate for 30 minutes at 5 °C, before the pressure was reduced to 2.84 Torr and imaging performed at 15 kV.

4.6 References

1.   Munter, A. *Scattering Length Density Calculator*.   2012; Available from: http://www.ncnr.nist.gov/resources/sldcalc.html. Accessed Jan 2012
2.   Koutsioubas, A., et al., *Slow and remanent electric polarization of adsorbed BSA layer evidenced by neutron reflection.* Soft Matter, 2012. **8**(9): p. 2638.
3.   Dewhurst, C., *GRASP: Graphical Reduction and Analysis SANS Program for Matlab.*   Institut   Laue   Langevin<   http://www.   ill. eu/fileadmin/users_files/Other_Sites/lss-grasp/grasp_main.   html,   2001. **2007**.Accessed Jan 2012

Chapter 5 : **Aggregate Nucleation**

5.1 Introduction

This chapter details the treatment and analysis of data from small angle neutron scattering experiments carried out at D22 at the Insituit Laue-Langevin, Grenoble. The aim of these experiments was to characterise the formation and growth of particulate aggregates of bovine serum albumin on a nanometre scale, and elucidate nucleation kinetics from these findings.

5.2 Experimental Overview

Bovine serum albumin at a concentration of 2 mg/mL was subjected to a temperature increase and then held at that temperature for a period of time. This was performed at three temperatures; 70 ℃, 75 ℃ and 80 ℃, and repeated with scattering data collected at three detector distances; 2 m, 4 m and 11.2 m, (except the 70 ℃ experiment in which 3 samples were simultaneously measured at 4 m), in order for the scattering intesity changes over the full q range to be analysed.

5.3 Data Treatment

Raw data from these experiments was collected in the form of a two dimensional scattering pattern (Figure 5.1). Initially a mask was applied over the area of the beam stop to negate any bad data pixels registering intensity values on an area of detector which is behind the beam stop. A number of corrections were applied to the data to extract the two dimensional scattering produced by the protein alone. Factors that needed accounting for were: sample container scattering, buffer scattering, the transmissions of these elements (including the sample itself), and residual scattering produced from the surrounding environment.
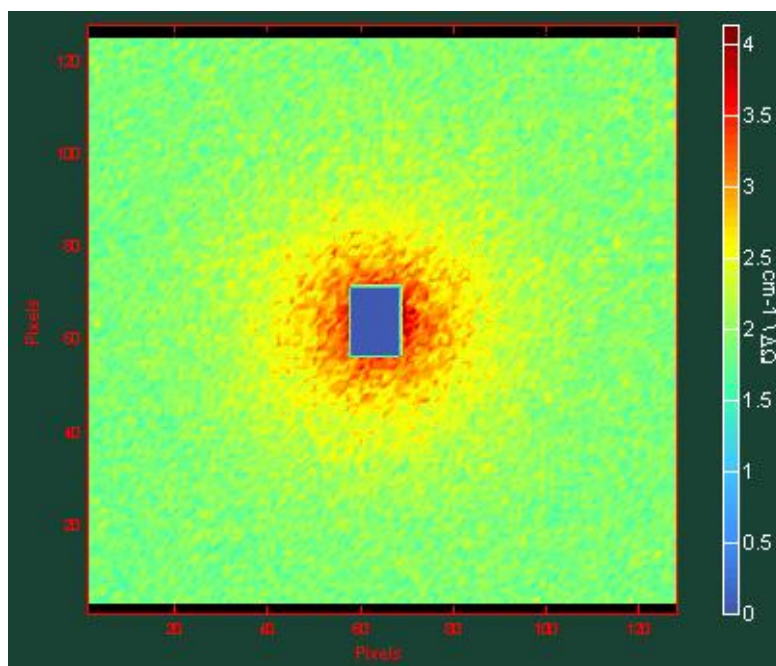
Figure 5.1: Raw data scattering pattern obtained from BSA at 75 °C at 4 m. A mask has been applied over the beamstop.

In order to correct for the scattering from external sources (those other than the protein sample) scattering from the sample container containing solely buffer was collected, and for the environmental scattering, cadmium was placed in the beam in order to absorb the full forward transmission. These data sets were collected for each temperature run and at each detector distance and collection time was identical to the collection time of the corresponding data set.

In order to background subtract the data sets, transmission values for both the sample (T_s) and background sample (T_e) were also obtained [1]. Where T_s (Equation 5.1), and T_e (Equation 5.2) can be calculated from counts of the neutron beam before and after the sample.

$$T_S = \frac{Transmitted\ neutrons\ through\ the\ sample + Empty\ cell}{Transmitted\ neutrons\ through\ the\ empty\ cell}$$

Equation 5.1: Equation for calculation of sample transmission.

$$T_e = \frac{Transmitted\ neutrons\ through\ empty\ cell}{Transmitted\ neutrons\ in\ the\ empty\ beam}$$

Equation 5.2: Equation for calculation of empty cell transmission.

The values of these transmissions and the background scattering intensities can then be used to obtain the corrected value for the scattering from the sample (Equation 5.3).

$$I_{Corrected} = \frac{1}{T_S T_e}(I_S - I_{cd}) - \frac{1}{Te}(I_{Bk} - I_{cd})$$

Equation 5.3: Calculation of corrected intensity from transmission and background scattering values.

Where $I_{corrected}$ is the corrected scattering from the sample, and I_s, I_{cd} and I_{Bk} the initial scattering from the sample, cadmium, and empty cell respectively. This process, applied through use of the programme GRASP [1], enabled the scattering pattern of the sample to be obtained (Figure 5.2). It can be seen that when compared with the raw data sample, the corrected data contains a flat background at the higher angles; illustrated by the blue on the scale, whilst the scattering from the sample is still prominent at the lower angles.
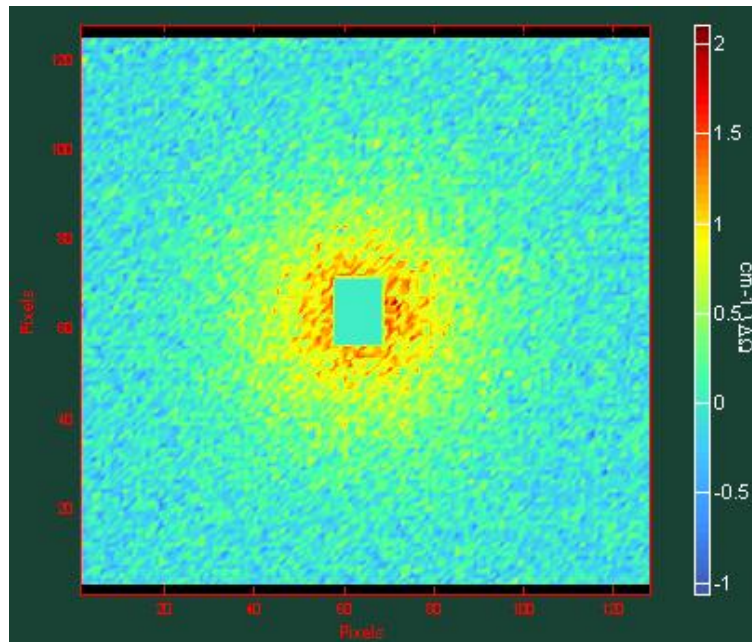
Figure 5.2: Corrected scattering pattern for BSA at 75 ℃ at 4 m.

121

In order to define the q range of the scattering pattern, first the beam centre was determined so that the central pixel could be used as to calculate the distance from the centre to each individual pixel. GRASP will determine the central pixel from the intensity of the beam profile (Figure 5.3). This central pixel value is also used as the base point from which the radial averaging of the scattering pattern is performed.
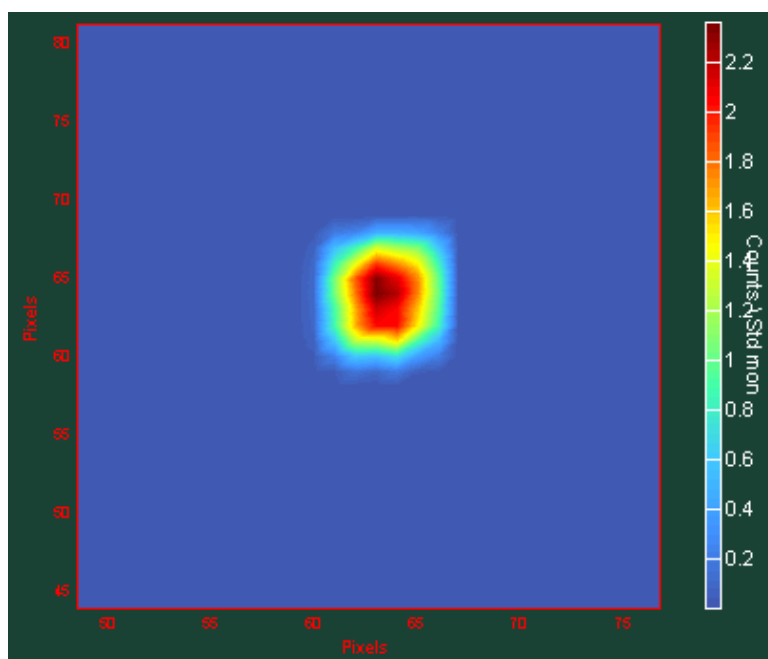
Figure 5.3: Attenuated beam profile on 4 m detector. Note: this is a magnified view when compared to previous scattering figures.

As the protein samples were isotropic, each scattering pattern could be radially averaged, and intensity versus scattering vector could be obtained. Radial averaging takes the intensity of each set of pixels at the same distance from the centre, and averages the pixels at every distance. The resultant plot is shown in Figure 5.4. The radial averaging does not include any pixels whose average would be calculated from pixels behind the beam stop or mask, as there are not sufficient pixels to give reliable statistics on the intensities so close to the beam stop (Figure 5.5). The resultant data was then exported as a .csv file where it could be manipulated in the data handling program; Mircosoft Excel.
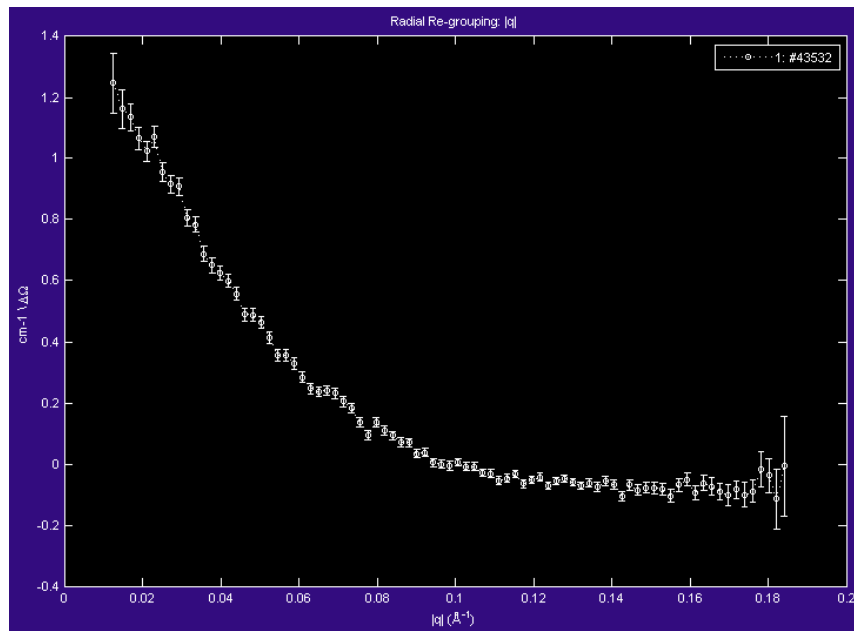
Figure 5.4: Intensity versus scattering vector plot of BSA at 75 ℃ at 4 m.

Beamstop

Detector area omitted from averaging

Detector area where radial averaging takes place.
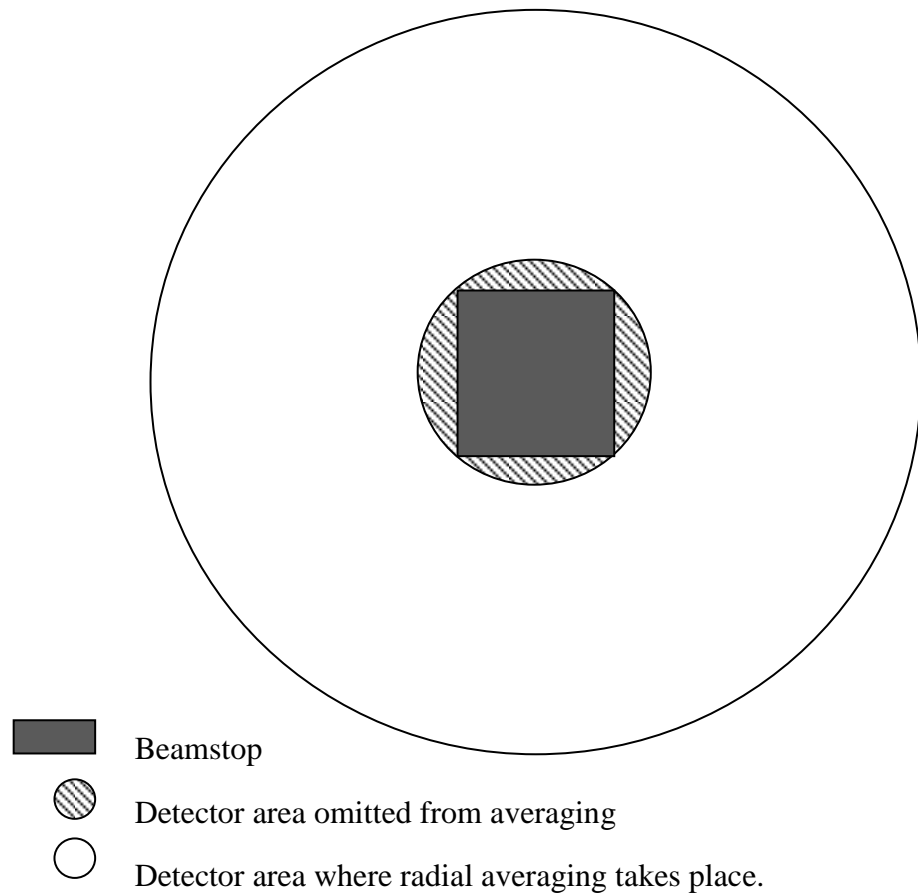
Figure 5.5: Diagram to illustrate omitted detector area from radial averaging process.

The data in its current form, although individually corrected, was not normalised across the 3 detector distances, therefore the intensities were not comparable (Figure 5.6).
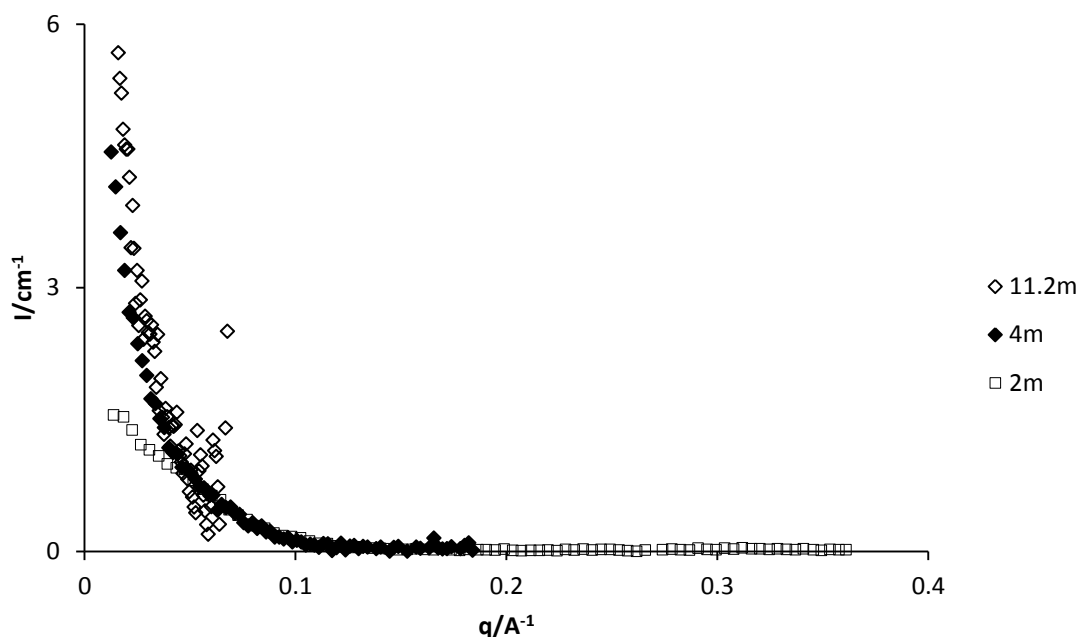
Figure 5.6: Corrected scattering data for the first minute of BSA at 75 ℃ at 3 detector distances.

In order to make the three data sets combine into one comparable set of data, both the 4 m and 11.2 m data were normalised using a scaling factor so that the intensities of the data at each scattering vector were of similar values. In order to perform this, the data was normalised to the 2 m data, as that had a sufficient flat background from which to start from (Figure 5.7). This normalisation was performed for both the 80 ℃ and 75 ℃ data sets, with each normalisation being performed across the entire time course of data. Points at which the intensities from different detectors overlap showed good agreement with where the errors on the data points began to increase (the error bars in Figure 5.6 and Figure 5.7 have been omitted for clarity). Additionally, as the intensity of the 4 m and 11.2 m data were scaled, the same scaling was applied to the magnitude of the errors in order to maintain the proportions between data magnitude and error bar size. The data normalisation maintained its integrity throughout the time course of each experiment (Figure 5.8)

This completed the handling of the raw data, which was now in a format suitable for analysis.
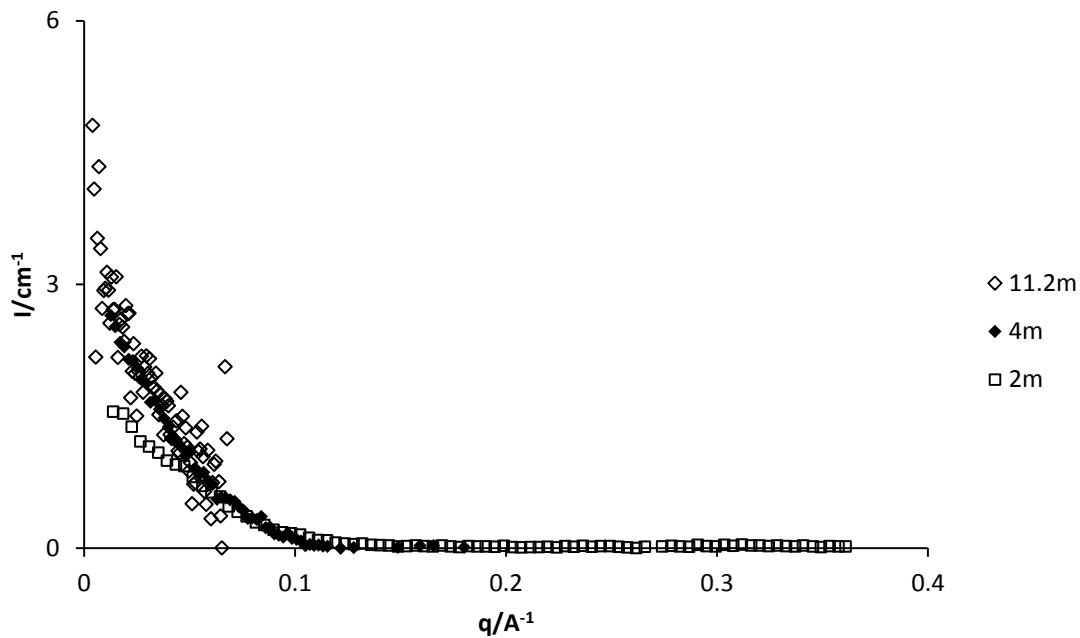
Figure 5.7:Normalised scattering data for the first minute of BSA at 75 ºC at the 3 detector distances.
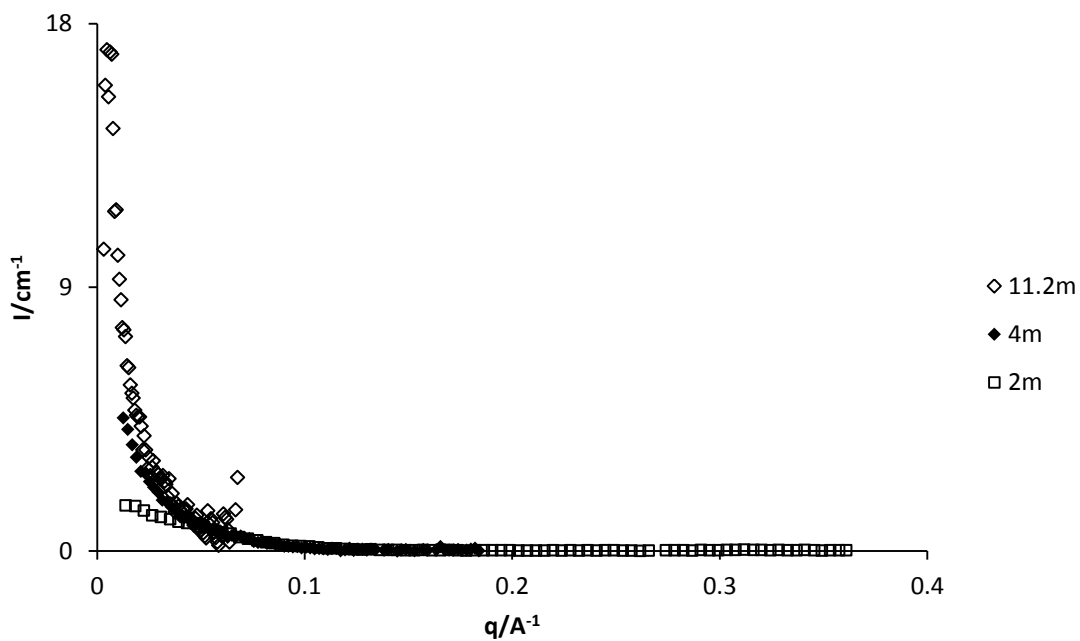
Figure 5.8: Normalised data of BSA at 75 ºC, after 228 minutes

5.4 Data Analysis – Native Protein

An initial scattering plot of bovine serum albumin (Figure 5.9) before heating was produced from the data collected across the 3 detectors ranges complete with errors. The magnitude of error bars increases at the inner and outer limits of each data set, owing to the lower number of pixels contributing to the intensity value at the specific q value. Where the error bars are large the overlapping data set contains more reliable data, therefore the majority of the range retains good statistics.
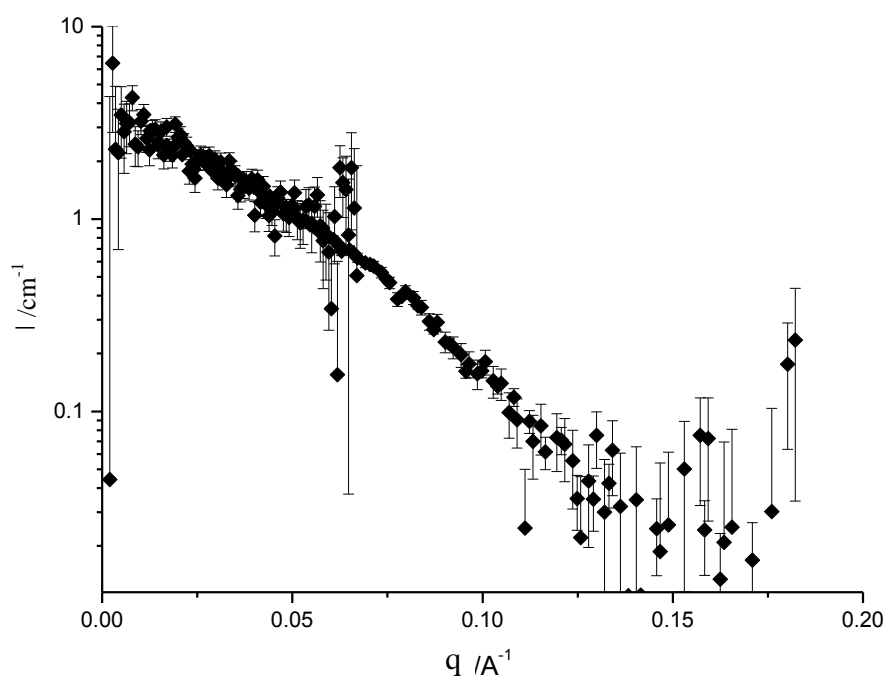
Figure 5.9: Log intensity against scattering vector plot for native BSA.

When both axes are placed on a logarithmic scale, the surface parameter of the native protein can be elucidated (Figure 5.10). This figure shows that as the gradient is seen as q^{-4} the protein can be seen as spherical, which is to be expected from a folded protein in

its native state such as bovine serum albumin is here. Some of the more erroneous data points have been omitted for clarity.
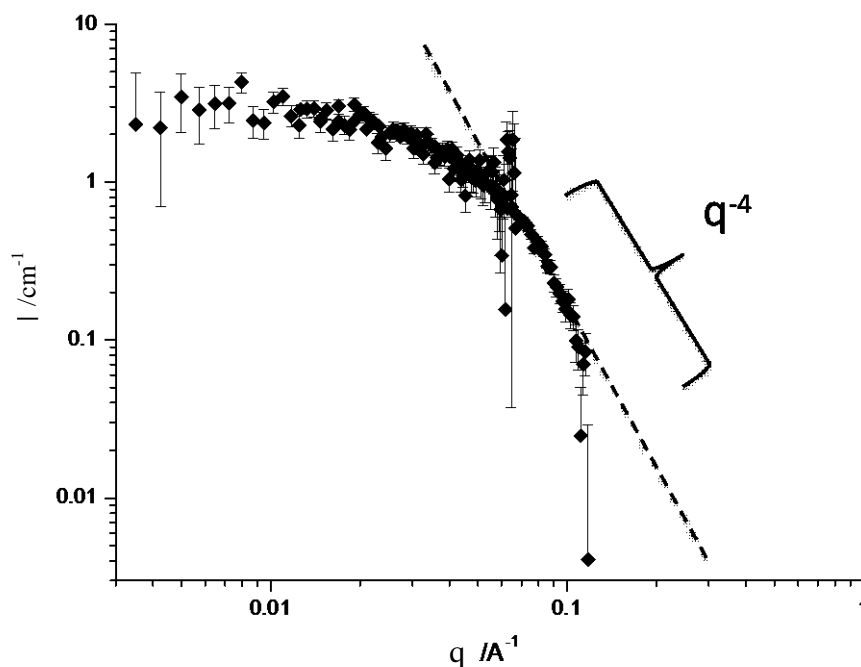
Figure 5.10: Log-log plot of native BSA.

The size of the particles was also estimated. The monomer was homogenous and in a dilute solution therefore the Guinier equation could be used to obtain a radius of gyration for the particle.

$$y = I0 \times e^{\left(\frac{r^2 x^2}{3}\right)}$$

Equation 5.4: Guinier equation.

Equation 5.4 states the Guinier equation where *r* is given as the radius of gyration. This was applied to the scattering pattern of the BSA in solution (Figure 5.11). Equation 5.4 was applied to the data only over the applicable Guinier region, where $q \times Rg > 1.3$ [2]. The extracted radius of gyration value was 33.9 ±0.9 Å. This appears in agreement with other values obtained for the radius of gyration of BSA [3-5], where the value is

generally considered to be around 30 Å. The value obtained here appears to be marginally larger, which may be attributed to the destabilising effect that D_2O may exert on the protein. This reported value can be stated with confidence however, as all native BSA samples reported near similar radius of gyration values.
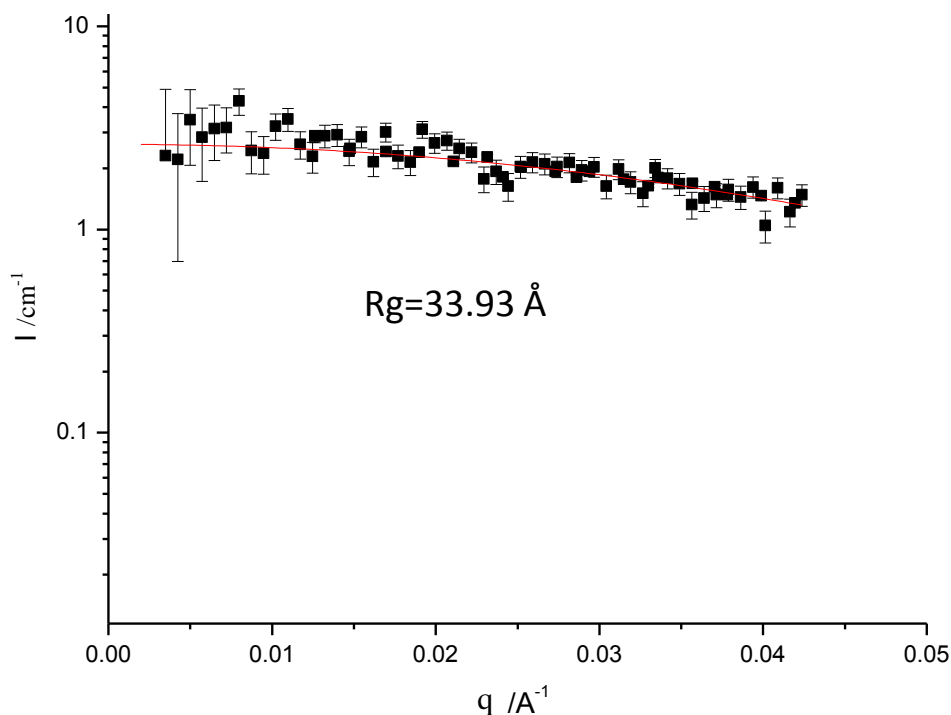
Figure 5.11: Guinier fit of native BSA.

As the native structure of BSA is solved [6] (PDB code 3V03) it is possible to fit the theoretical solution scattering of the structure to the obtained scattering pattern. This can be performed using the programme CRYSON [7]. This programme takes the data input in the form of the PDS file and creates a spherically averaged scattering pattern which takes into account the hydration shell and extent of deuteration present in the sample. CRYSON attempts to minimise the difference in the chi-squared value between the predicted and experimental data to give values for the hydration levels and envelope size of the protein. This was performed for BSA, however: the PDB structure of BSA is in the dimeric form, and as the studied solution of BSA should be monomeric a PDB

file was created involving solely the monomer chain (this function is not available in CRYSON, but in the near identical father programme CRYSOL [8], available in the same ATSAS package).
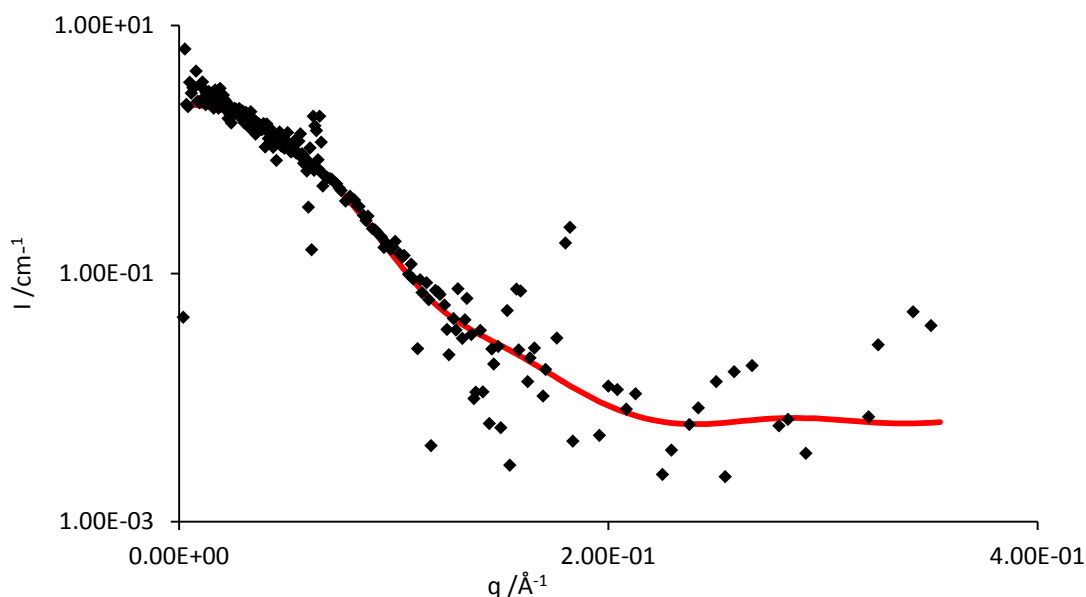
Figure 5.12: CRYSON fit of BSA crystal structure to obtained scattering data. Error points are identical to Figure 5.9, and omitted to show the quality of the fit.

The fit of the BSA theoretical structure (red line) shows an excellent fit to the obtained scattering data in Figure 5.12. This agreement was reached assuming an 80 % deuteration of available surface exchange atoms, and produced values for the volume of BSA at 82666 $Å^3$ and a radius of gyration of the envelope of 32 Å. The deuteration appears in agreement with the proposed levels of deuteration from BSA molecules [9, 10] whilst the radius of gyration is in good agreement with results from the Guinier plot (Figure 5.11).

To assess the quality of the native protein fold, a Kratky plot was produced of the native BSA sample (Figure 5.13). The bell shape with a decrease to the baseline of this plot is a prime example of a Kratky plot of a globular folded protein sample [11].
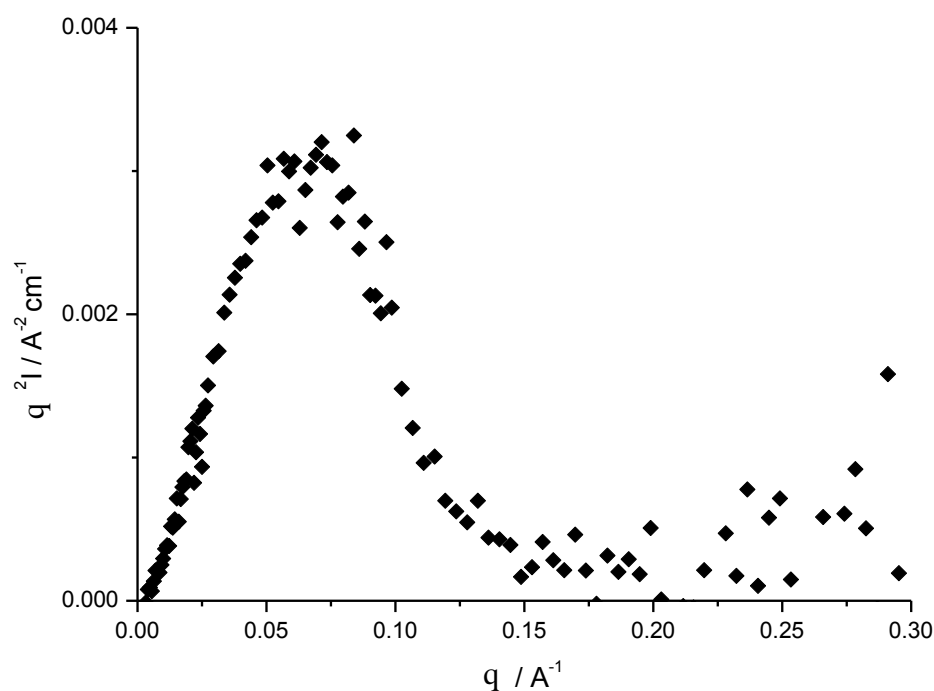
Figure 5.13: Kratky plot of native BSA. Error points are identical to Figure 5.9, and omitted to show the quality of the fit.

These initial data treatments show that the 2 mg/mL bovine serum albumin in a 90 % D_2O solution is in a monomeric form with a radius comparable to that seen in other experiments. It is in a natively folded, globular state, with 80 % of deuterated surface hydrogens. All results shown here are comparable with the literature. Thus the experiments carried out here can be considered to probe the behaviour of the native BSA protein molecule under heating.

5.5 Data Analysis – Heat Increase

Similar methods can be applied to the protein during the heating phase of the experiment to probe any structural changes that occur during the heating of the protein solution.
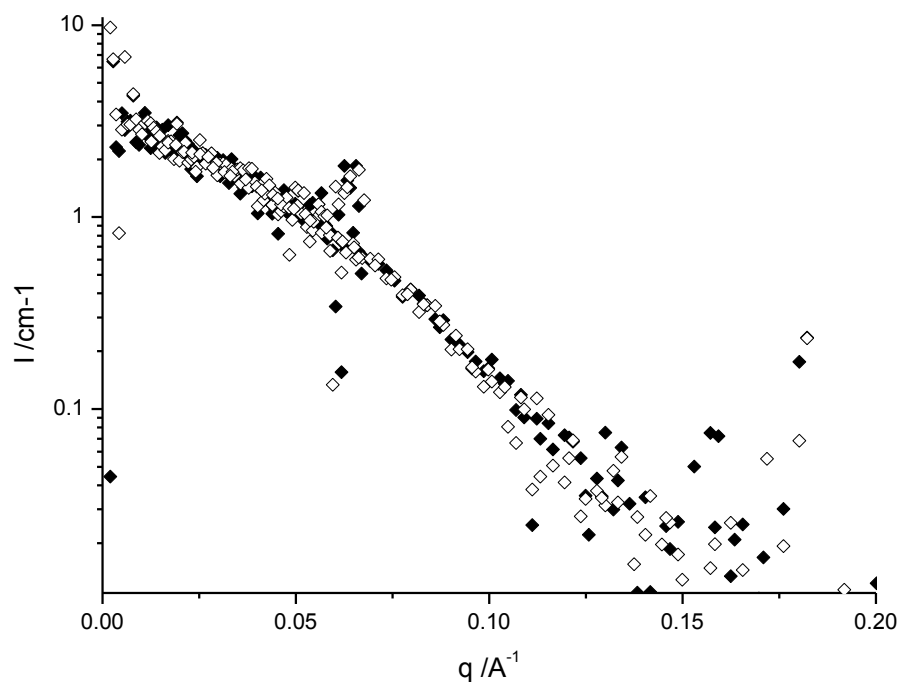
Figure 5.14: Intensity versus scattering vector plots of BSA at 25 ºC (closed diamonds) and after just reaching 80 ºC (open diamonds) taking 7 minutes. Error points are omitted to allow comparison.

During the heating phase of the experiment, it can be seen that there is very little difference in the intensity plot of the native BSA sample when compared to that of the sample which has been heated (Figure 5.14). In this example the frame is taken as the one preceding aggregation detection. In this case there are only seven frames leading up to the aggregation, however the same can be seen to be true of both the experiments at 70 ºC and 75 ºC where 25 frames are present, representative of 25 minutes (Figure 5.15). Both 75 ºC and 70 ºC exhibit similar initiation times, however only the 75 ºC data is shown here as it is presented over a larger q range than the 70 ºC data. As can be seen in both figures, there is little change in the intensities at low q, therefore there are now larger structures appearing (i.e. aggregates) and there appears to be minimal difference in the scattering present at high q.
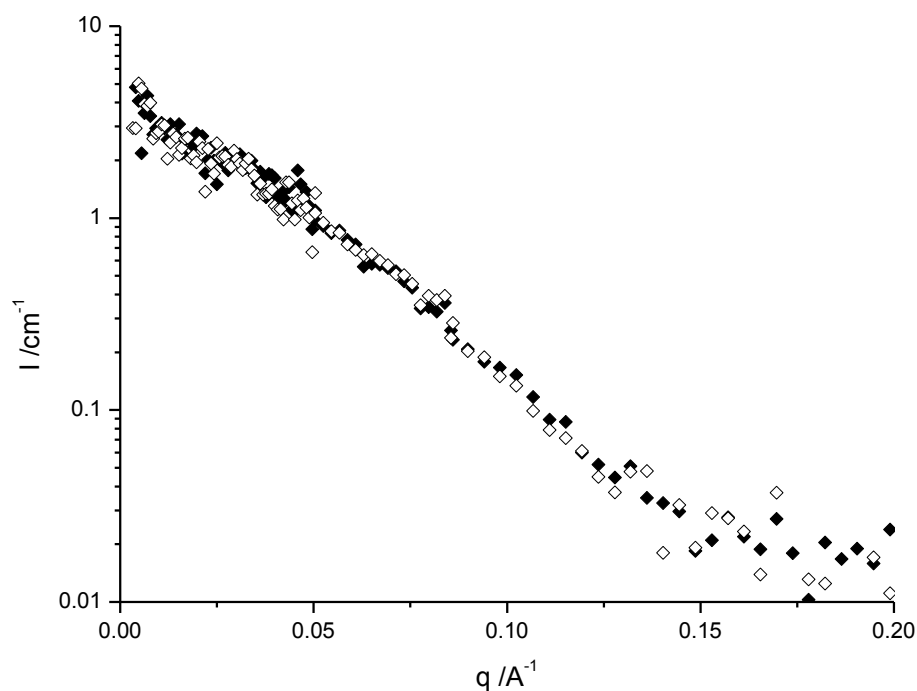
Figure 5.15: Intensity versus scattering vector plots of BSA at 25 ℃ (closed diamonds) and after just reaching 75 ℃ (open diamonds) after 25 minutes. Error points are omitted to allow comparison.

The y axis of a Kratky plot factors the intensity by q^2, which results in any small changes in the high q intensity to be amplified. This gives a good indication of any differences in the plots during the heating stage, and will also assess the quality of the protein fold during heating. The ramp up to 75 ℃ heating is used as a prime example (Figure 5.17). In the initial stage at 25 ℃, the Kratky plot peaks and drops down to the baseline as was seen in the Kratky plot of the native protein (Figure 5.13). As the temperature rises to 45 ℃, there is still very little deviation from the baseline at the higher q values. At a temperature of 65 ℃, there is a significant increase in the baseline of the Kratky plot at high q. This is indicative of the protein being in the partially folded state and a loss of globular shape. This transition in the protein structure in the range leading up to 65 ℃ appears to correspond to a melting temperature seen in a differential scanning calorimetry study of BSA [12] in which the melt range was seen to be between

56 ℃ and 69 ℃. At this temperature the loss of intra protein interactions must be sufficient enough to change the overall form of the protein to a slight random coil-like structure [13].

The conditions for these experiments require the protein to be present in a 90 % D_2O solution. In order to understand the effect this may have upon the unfolding characteristics of the BSA molecule, a fluorescence study was undertaken. The fluorescence probe 1-anilinonaphthalene-8-sulfonic acid was used to probe hydrophobic exposure on the protein during a heat ramp. This was performed for increasing concentrations of D_2O in the buffer solution.
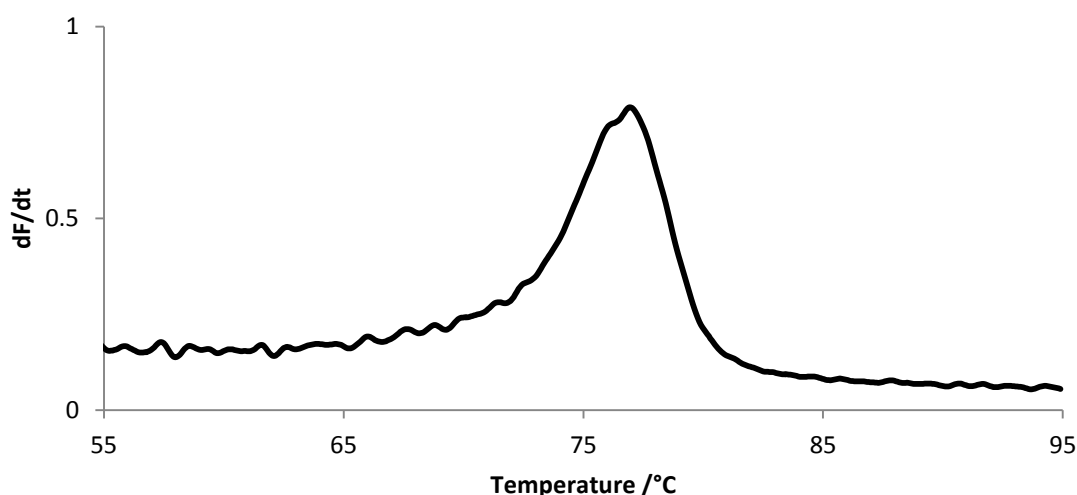
Figure 5.16: Change in fluorescence over time of ANS with BSA in 90 % D_2O solution.

It was found that increasing concentrations of D_2O in the sample buffer led to a slight decrease in the melting temperature of BSA. The melt temperature for BSA in 0 % D_2O was calculated at 83.8 ℃ whilst for BSA in 90 % D_2O it is 78.4 ℃. These values also correlate with the figures given from calorimetry results [12]. Figure 5.16 illustrates the unfolding characteristics of BSA at 90 % D_2O concentration, comparable to that of the SANS experiments detailed here.
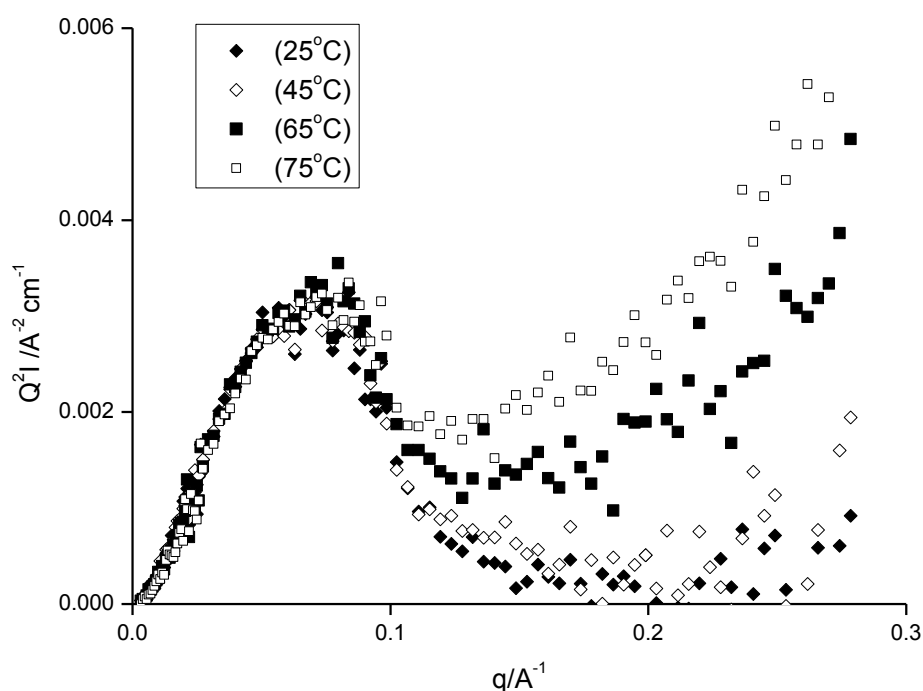
Figure 5.17: Kratky plot of BSA heated up to 75 ℃, at 25 ℃ (closed diamonds), 45 ℃ (open diamonds), 65 ℃(closed squares), 75 ℃ (open squares). Error bats are omitted to show difference in the plots.

The deviation from the baseline at high Q is seen to increase further as the protein in solution reaches 75 ℃. This indicates further unfolding of the protein structure. At this stage, the protein is still largely globular, as is evident in the remaining peak presence in the Kratky plot at 75 ℃. This however, is the total extent of the protein unfolding seen in the Kratky plot before aggregation takes place. Continued exposure of the protein to this temperature does not lead to any further unfolding to the BSA molecule, which would be illustrated in the Kratky plot as a continuing increase in the q^2I intensity to high q resulting in no peak in the plot. This correlates well with the literature which indicates that complete unfolding of BSA may not occur until almost 90 ℃ [14]. This data shows that only partial unfolding of the BSA molecule is required in order for aggregation to occur. This partial loss of tertiary structure is sufficient for the hydrophobic areas of protein which contribute to this loss of structure to cause binding

between protein monomers and initiate aggregation. This particulate aggregation behaviour appears concurrent with other amorphous aggregation of BSA within the literature, in terms of only perturbed tertiary structure being necessary for aggregate formation [15].
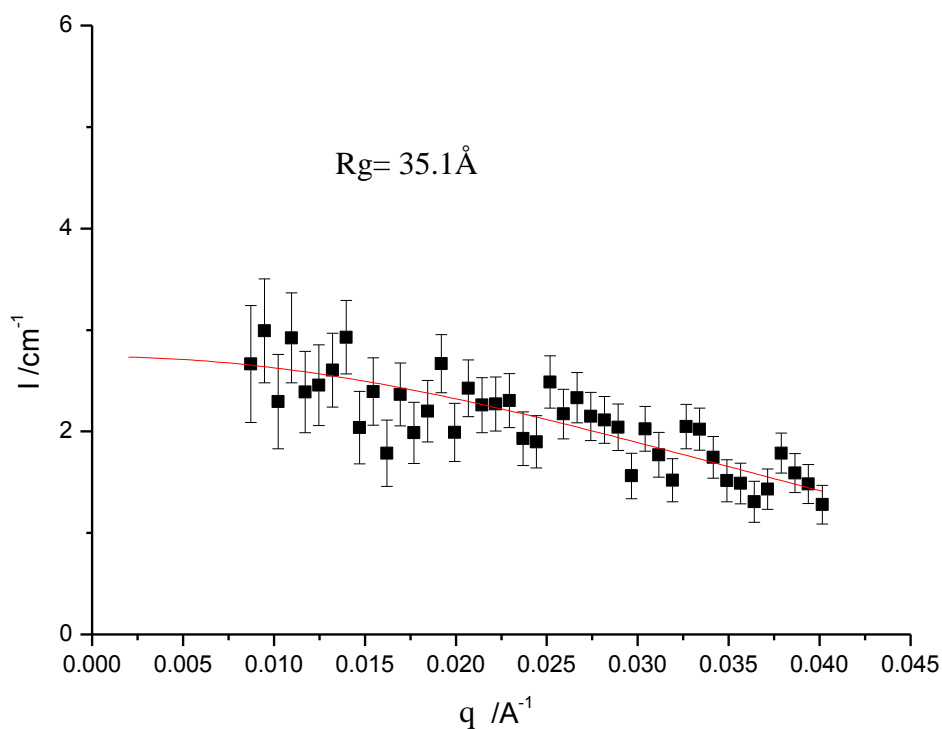
Figure 5.18: Guinier plot of BSA after 7 minutes of heating to 80 ºC

Figure 5.18 illustrates the Guinier plot of the BSA before aggregation during heating to 80 ºC. The plot shows that the radius of gyration of the monomer has increased slightly at 35.13 ±1.86 Å, so this increase can be taken with minimal certainty as it appears within the range of error in the monomer. Similar values are also seen in the additional experiments where the pre aggregation frames showed radius of gyration values of 36.40 ±1.9 Å and 36.20 ±0.74 Å for the monomer at in the 75 ºC and 70 ºC runs respectively (Figure 5.19 and Figure 5.20).
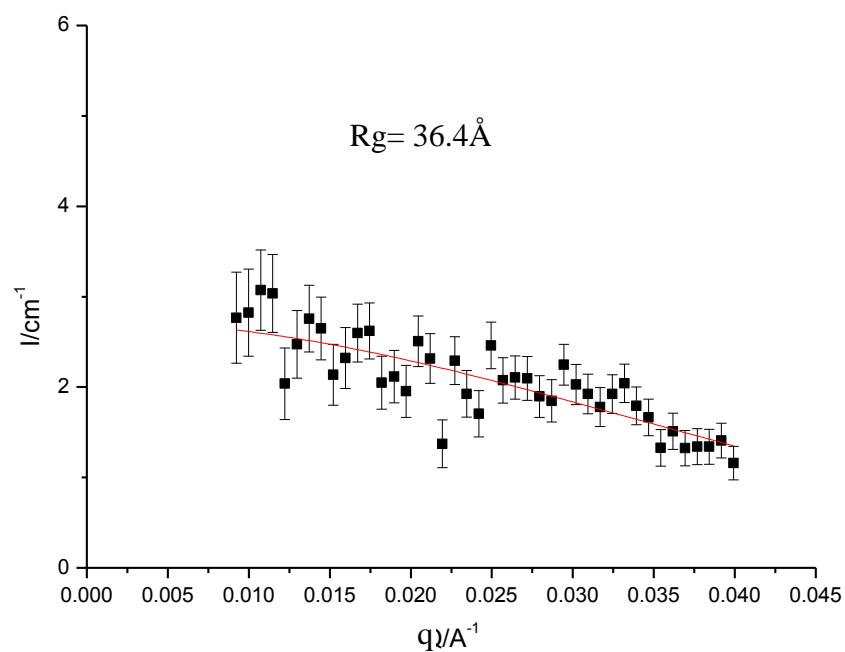
Figure 5.19: Guinier plot of BSA after heating to 75 ºC for 25 minutes.
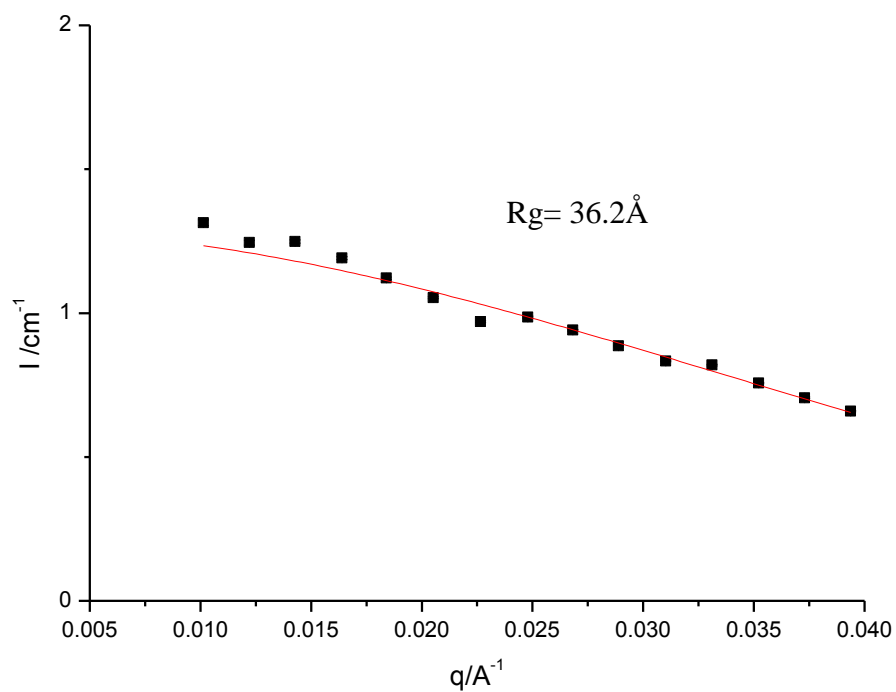
Figure 5.20: Guinier plot of BSA after heating to 70 ºC for 24 minutes. Error bars are the size of the points.

136

These values for the radius of gyration in combination with the Kratky plots illustrate the partially folded swollen nature of the BSA intermediates which are present before aggregation occurs. These can be considered to be the aggregate prone intermediates which form the initial aggregate nuclei.

5.6 Data Analysis – Aggregation

After the increase in heat of the protein sample, protein aggregation can be seen to occur, characterised by an increase in intensity at low Q (Figure 5.21). This is indicative of larger structures becoming present in the illuminated sample volume, in this case; protein aggregates.
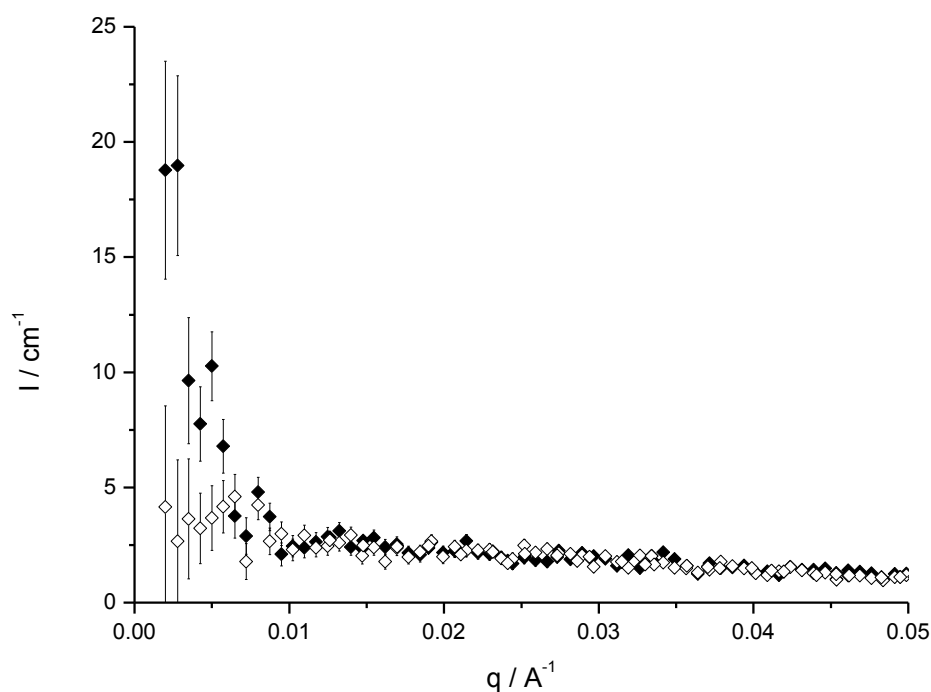
Figure 5.21: Intensity plot of BSA at 80 ºC before (open diamonds), and 10 minutes into the aggregation process (closed diamonds)

The protein sample is held at an isothermal temperature for the remaining duration of the experiment. As this continues, the intensity at high Q is seen to increase with time (Figure 5.22). This is caused by the continual increase in size of the large structures within the sample.
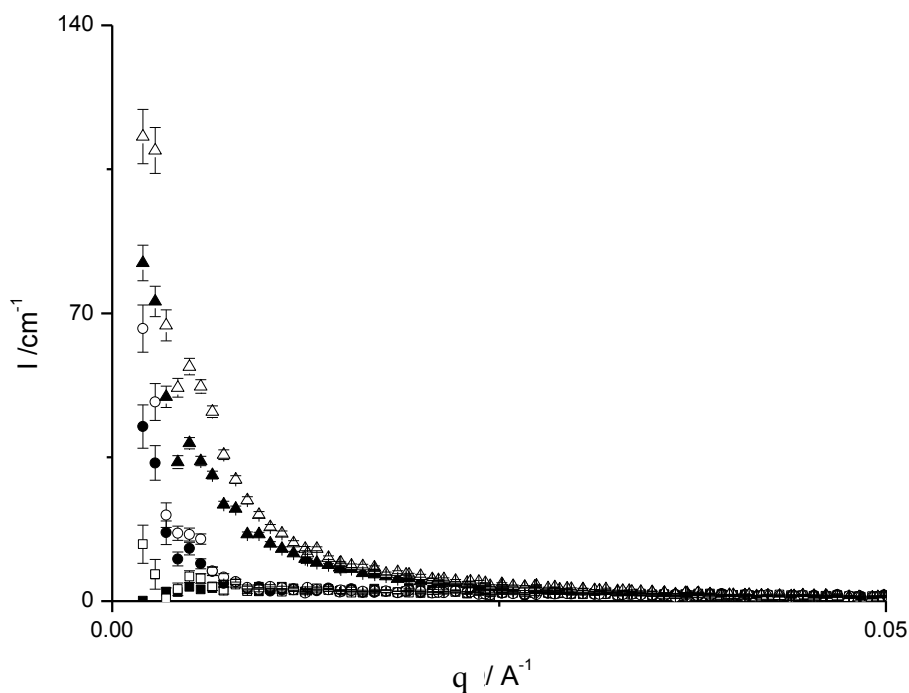
Figure 5.22: Progression of scattering intensity from BSA whilst heated at 80 ºC for 1 min(closed squares), 10 min(open squares), 20 min (closed circles),30 min (open circles) 40 min (closed triagnles), 50 min (open triangles)

From the log intensity plot of the initial BSA monomer (Figure 5.10), we can see how the monomer can be considered as a sphere, although this is an approximation of its perturbed globular form. As aggregation progresses, the structures within the sample maintain a spherical form (Figure 5.23), which is to be expected, as it is shown that aggregates formed at around this pH will form spherical aggregates [16, 17]. As the temperature is maintained, the structures formed no longer adhere to a q^{-4} decay, instead dropping off at a rate of $\sim q^{-2.2}$ indicative of a Gaussian chain. This can be caused by the clustering of particulate aggregates together in order to form these large chains [13].

138

This is a feature only observed in the experiments performed at 80 ℃, in both the 75 ℃ and 70 ℃ experiments, the profile of the log intensity graphs remain similar to that of the native state (Figure 5.10). This suggests that during these runs the protein aggregates maintain their spherical form throughout the experiment and do not cluster into chains. Reasons why this may be the case will be considered further on in this chapter.
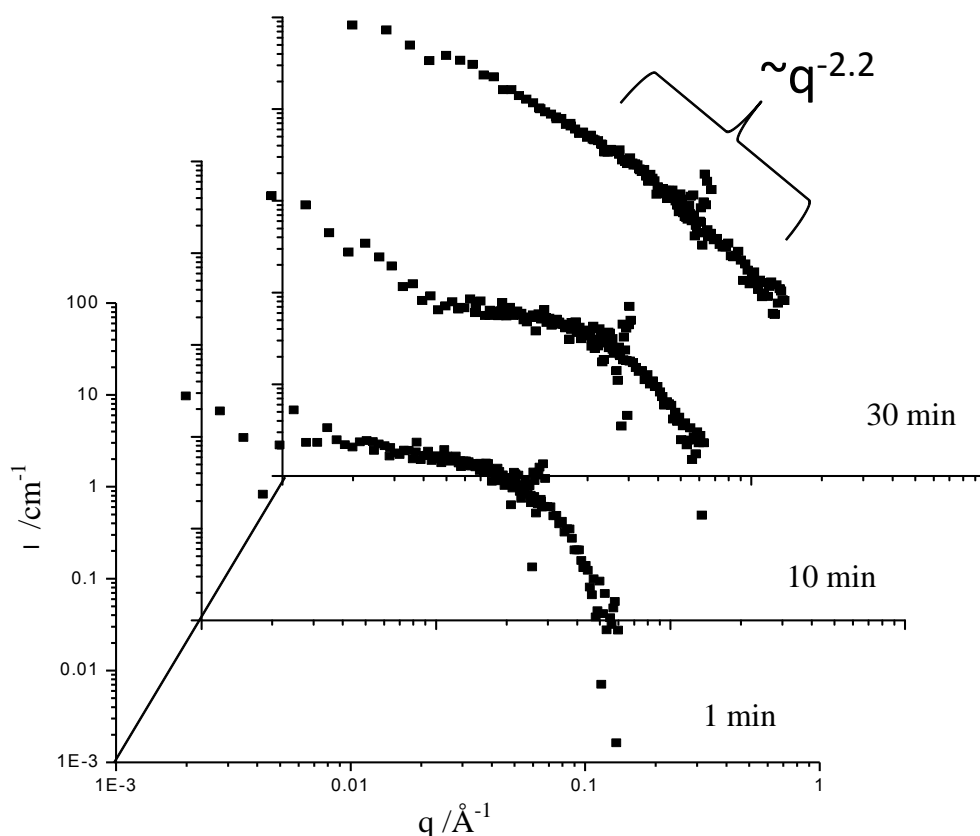
Figure 5.23: Log Log plot of BSA 1 min, 10 mins and 30 mins into aggregation at 80 ℃, ordered through the depth of the graph.

As previously noted, the BSA molecules can be considered to be spherical (Figure 5.10). The scattering from dilute spheres of known scattering length density contrast, scattering volume and radius can be modelled using the Rayleigh equation [18]. For a homogenous sphere with a radius R, the form factor of the spheres, $P(q)$ is:

$$P(q) = \left[\frac{3[\sin(qR) - qR\cos(qR)]}{(qR)^3} \right]^2$$

<div align="center">Equation 5.5: Rayleigh scattering from homogenous spheres.</div>

As this is a dilute system, *S(q)=1*, therefore the equation above describes the scattering from the BSA molecules completely, when considered as spheres.

When Equation 5.5 is fitted to the monomer, a good fit is achieved (Figure 5.24). The Rayleigh slope fits well at low *q*, where the majority of the high quality data lies. At higher q, the low resolution data deviates from the Rayleigh plot, which will be due to the low count time taken over this range, and some smearing effects. The radius reported from the Rayleigh plot of 36.5 Å does appear small in comparison to the radii of gyration achieved from Guinier plots of the same data, it does bear some agreement to other values for the radius of hydration presented in other papers on BSA [19], where the value was given as 33.9 ±2.7 Å.
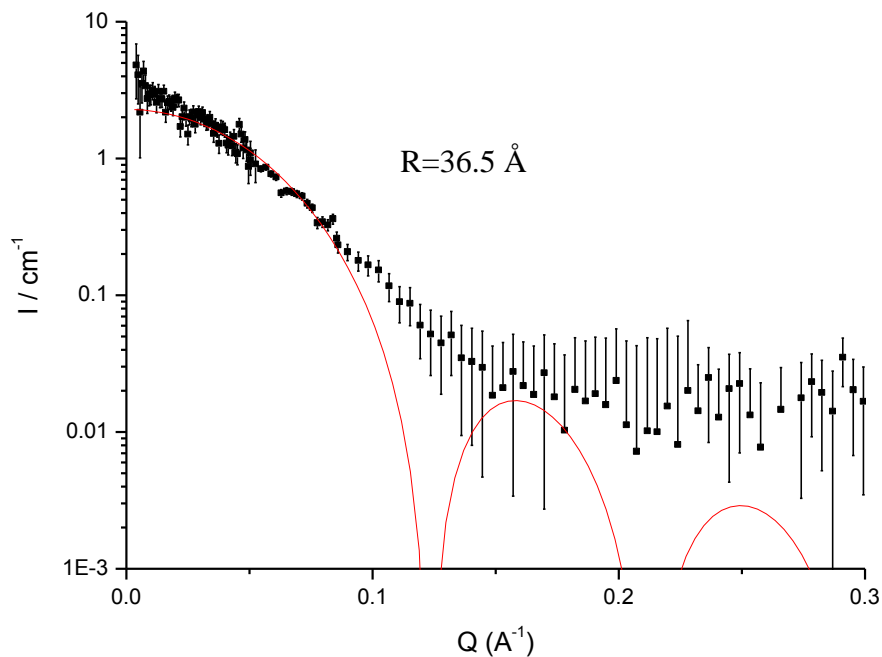
<div align="center">Figure 5.24: Fit of Rayleigh scattering from a homogenous sphere to scattering intensities from BSA monomer.</div>

It has been shown from the scattering plots that both the monomer and the aggregates can be considered as spheres. By creating a model based on the Rayleigh scattering from spheres in Equation 5.5, values can be obtained for the radius of the aggregate at each time point, along with the relative ratio between scattering from spherical monomer, and scattering from spherical aggregates. For form factor amplitudes from monomers $F_m(q)$ of radius R_m and aggregates $F_a(q)$ of radius R_a:

$$F_m(q) = \frac{3[\sin(qR_m) - qR\cos(qR_m)]}{(qR_m)^3}$$

Equation 5.6: Amplitude for Rayleigh scattering from monomer.

$$F_a(q) = \frac{3[\sin(qR_a) - qR\cos(qR_a)]}{(qR_a)^3}$$

Equation 5.7: Amplitude for Rayleigh scattering from aggregate.

Where:

$$I(q) = I_0 v^2 \Delta\rho^2 \left[(V_m F_m(q)) + (V_a F_a(q)) \right]^2$$

Equation 5.8: Scattering equation for dual population of monomer and aggregate spheres.

Where:

$$V_a = 1 - V_m$$

Equation 5.9: Calculation of volume fraction of aggregate.

Where V_m is the volume fraction of the monomer, hence V_a will return the volume fraction of the aggregate. I_0 is a scaling factor, v the volume of the scattering bodies, and $\Delta\rho$ is the difference in scattering length densities before the protein and the solution.

As in Equation 5.5, there is considered to be no particle interactions as the number of particles is the same or decreasing as aggregation occurs. If this is found not to be true, then the difference between the modelled and collected data should yield a curve with resemblance to that of a structure factor, with $S(q)$ tending towards 0 at low q. With the initial volume fraction and monomer radius known, it is possible to model the ratio between the V_a and values by subjecting them to the known limit (that of the initial monomer volume fraction) whilst allowing the aggregate radius to be free to fit to the scattering curve provided the aggregate radius was larger than that of the monomer. This was applied to each scattering plot run at the three experimental temperatures, with the resultant data being a product of fitting this model to over 300 scattering plots.
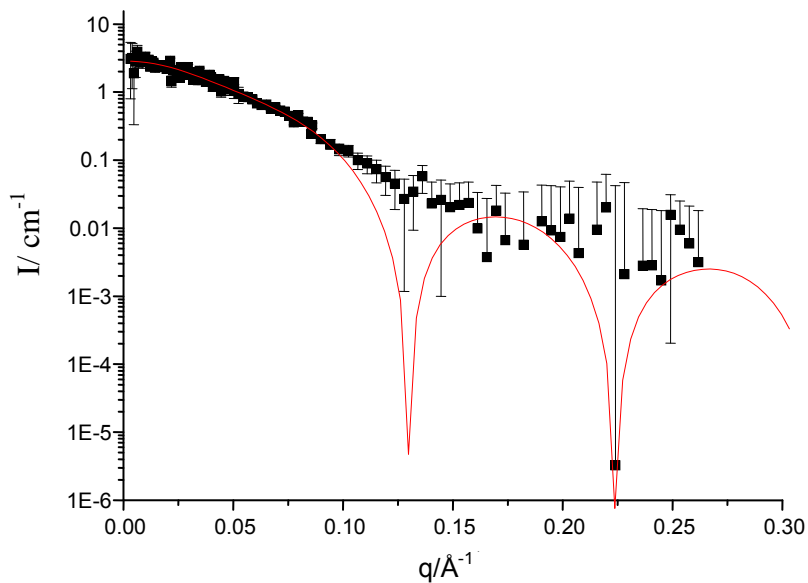
Figure 5.25: Dual population sphere fit to 20 minute frame of BSA at 75 ºC.

Figure 5.25 is an example of the model fit fit to the data, created using Equation 5.8. The model once again shows an excellent fit at the majority of the data, which is present at the low q scale and also with the high q data. There is also small evidence of the data

showing some fit to the minima within the model fit to the plot, albeit with the presence of data points with very high error in these regions.

Complete fitting of the data set for each temperature run yields the values of both aggregate radius and monomer fraction over a time course.
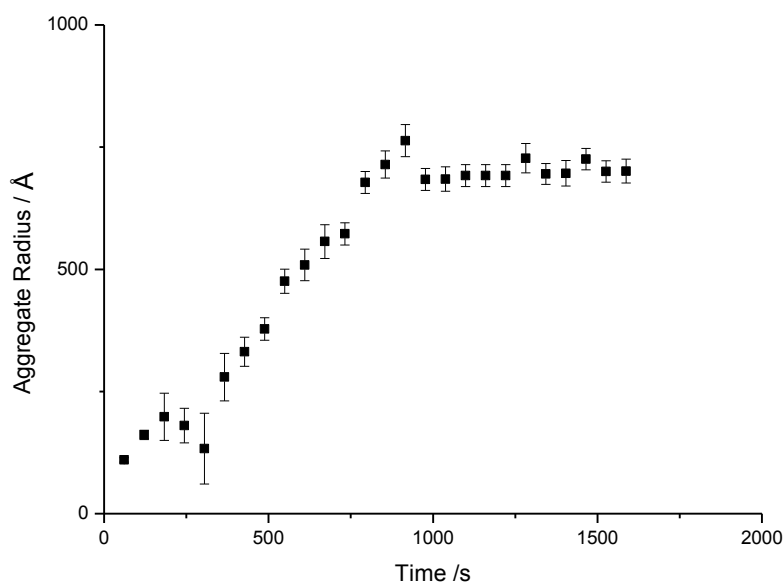
Figure 5.26: Aggregate radius growth of BSA over time when incubated at 80 °C.

Data is normalised so that the first time frame represents the first detection of aggregate. As the initial aggregate size detected is of the order of 90 Å, whilst the monomer only has a radius of ~30 Å, it is possible that aggregates may exist at a smaller size range than this, but they are either a) indistinguishable from the monomer in terms of the model or b) not stable enough to maintain the size population.

Figure 5.26 illustrates the rapid growth in aggregate radius that occurs initially, followed by a much slower growth rate when the protein is around 600 Å in radius. This effect is mirrored in Figure 5.27 where the monomer is depleted rapidly until 1000s into

aggregation. From this stage the depletion of monomer is slowed and lies in good correspondence with the slow in growth rate of radius in Figure 5.26.
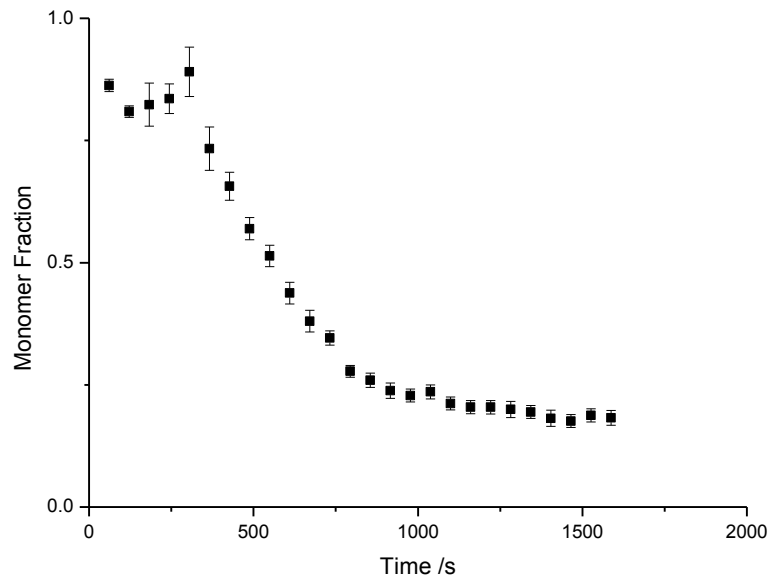
Figure 5.27: Monomer fraction depletion over time of BSA at 80 ℃.

A noticeable discrepancy is visible in Figure 5.27 over the initial time frames where there is some variation in monomer fraction that does not tend towards 0 as time decreases. This may be the result of a low amount of scattering coming from the aggregates at this stage due to their low size and concentration. As a result the model may have difficulties separating this scattering from the monomer, leading to a higher error in the points at the start of the experiment. When compared together (Figure 5.28) at these initial values will also show how some dependence exists between these two parameters within the model; however this only seems to exist in instances of low aggregate scattering.
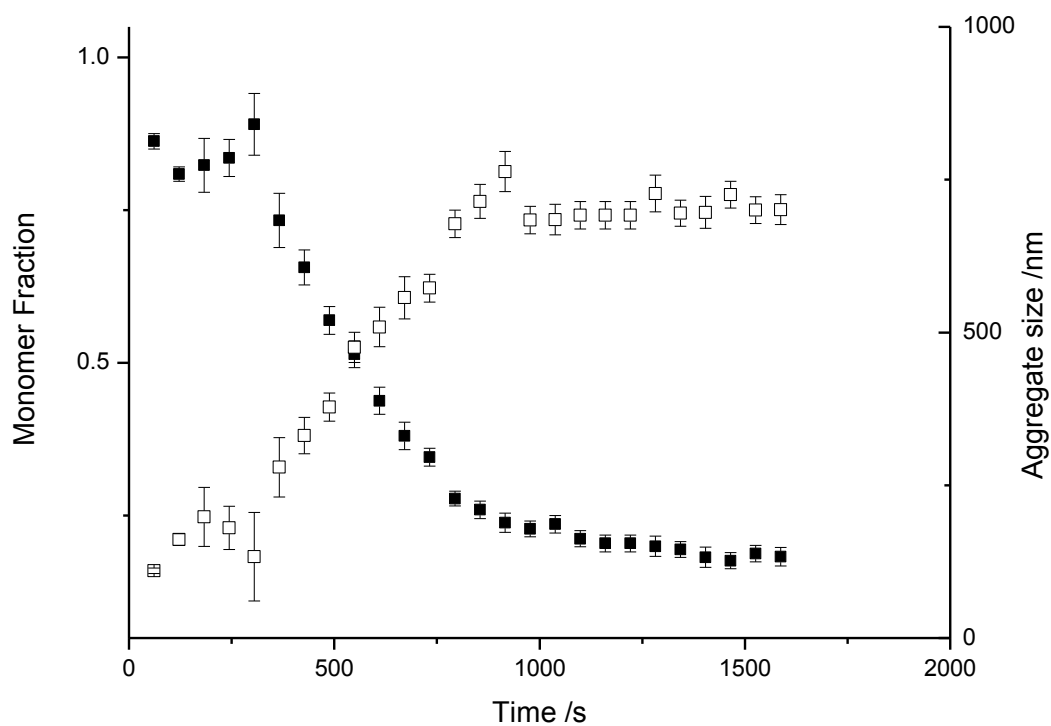
Figure 5.28: Monomer fraction (closed squares) and aggregate size (open squares) progression over time for BSA incubated at 80 ℃.
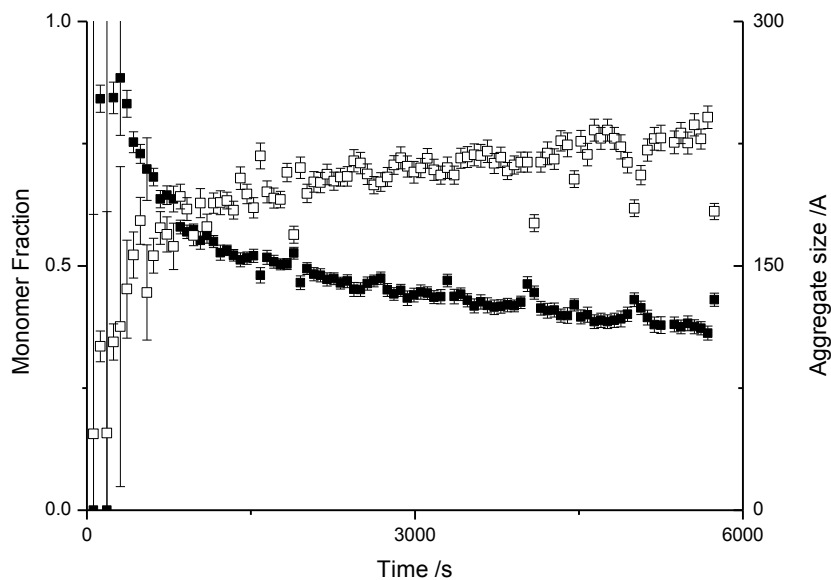
Figure 5.29: Monomer fraction (closed squares) and aggregate size (open squares) progression over time for BSA incubated at 75 ℃.

145

Figure 5.29 and Figure 5.30 illustrate similar characteristics of the aggregation of BSA at lower temperatures (75 ℃ and 70 ℃ respectively). Similar trends are present in these plots to those of 80 ℃; the emergence of monomer fraction appears at circa 0.85, whilst aggregate size emerges at 100 Å. There is an instance of aggregates emerging at 45 Å in Figure 5.29, however the errors related to this data are significantly large. This will be due to both the rapid growth change that may be occurring over the minute acquisition time for these data points and the low percentage of scatter produced from this small concentration of aggregates.
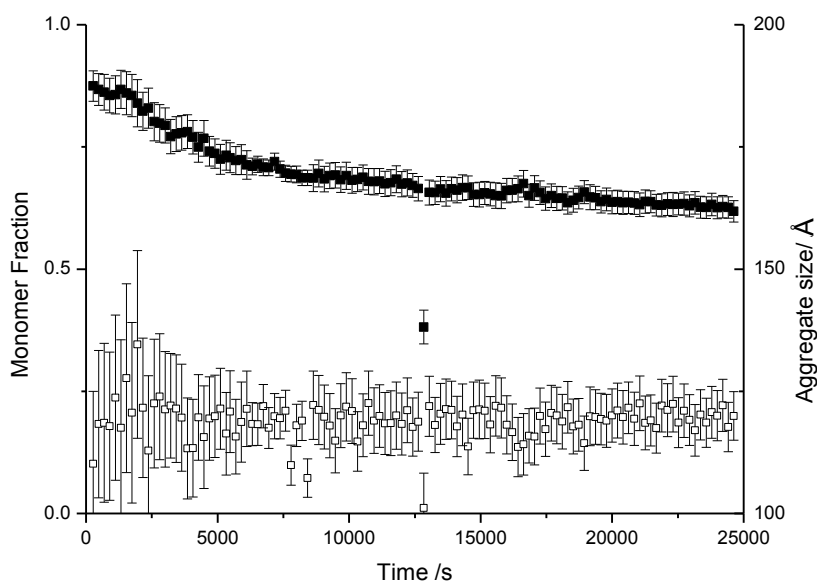
Figure 5.30: Monomer fraction (closed squares) and aggregate size (open squares) progression over time for BSA incubated at 70 ℃.

The rates at which aggregation occurs is temperature dependent as evident in the 3 plots (Figure 5.28, Figure 5.29, Figure 5.30). A high temperature results in a quicker depletion of the monomer fraction (to 0.1 over the course of 1000 seconds: Figure 5.28), over a lower temperature (at 70 ℃ the monomer fraction is reduced to 0.6 after 25000

seconds: Figure 5.30). All plots do show a consistent decline and tend towards zero, indicating that aggregation will continue until the entire monomer population is depleted.

The aggregate radii are seen to form more quickly and result in larger sizes (800 Å at 80 °C over 100 Å at 70 °C). This correlates with the monomer depletion plots, where slow linear monomer depletion in the latter stages is mirrored by slow rate linear growth. This may indicate that the aggregate population is stable, as the changes in rates appear proportional.

Aggregate concentration C_a, assuming all aggregates are of identical spherical size can be calculated by Equation 5.10, from the monomer fraction M_f and number of monomers in each aggregate A_N, where C_0 is the initial monomer concentration of the BSA at 3.01×10^{-5} M.

$$C_a = \frac{1 - M_f \times C_0}{A_N}$$

Equation 5.10: Calculation of aggregate concentration.

The number of monomers in an aggregate is a ratio between the volume of the aggregate and the volume of a BSA molecule, where its radius; R_m is taken as 34.37 Å and aggregate radius R_a is derived from the data.

$$A_N = \frac{R_A{}^3}{R_M{}^3}$$

Equation 5.11: Calculation of number of monomers in an aggregate.
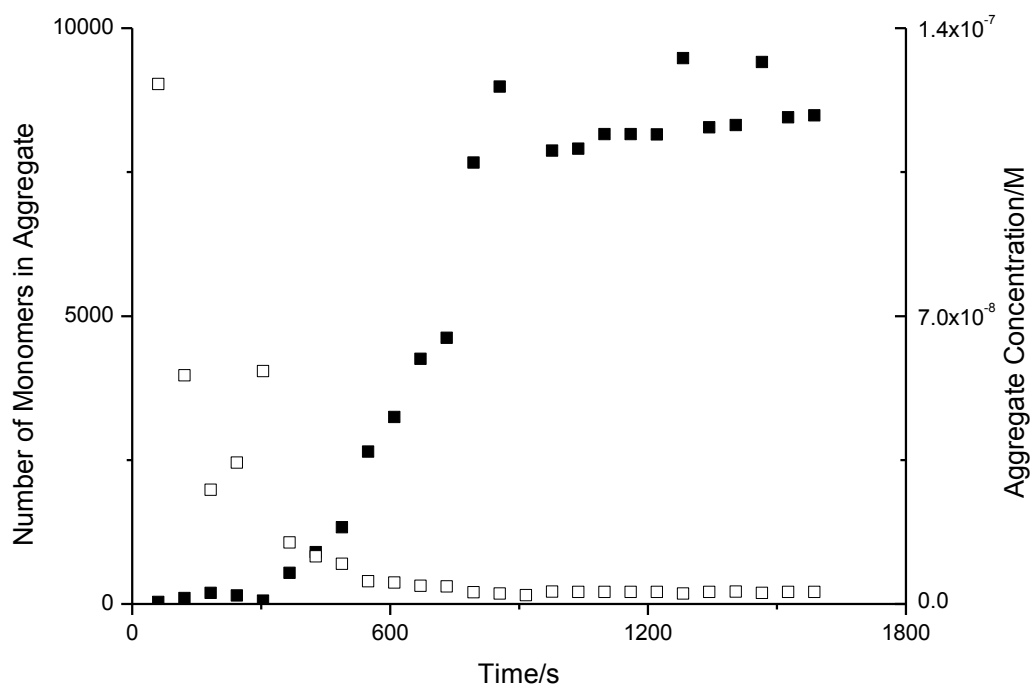
Figure 5.31: Number of monomers in aggregates (closed squares) and aggregate concentration (open squares) for BSA at 80 ℃. Errors are of magnitude replicated in Fig 5.29.
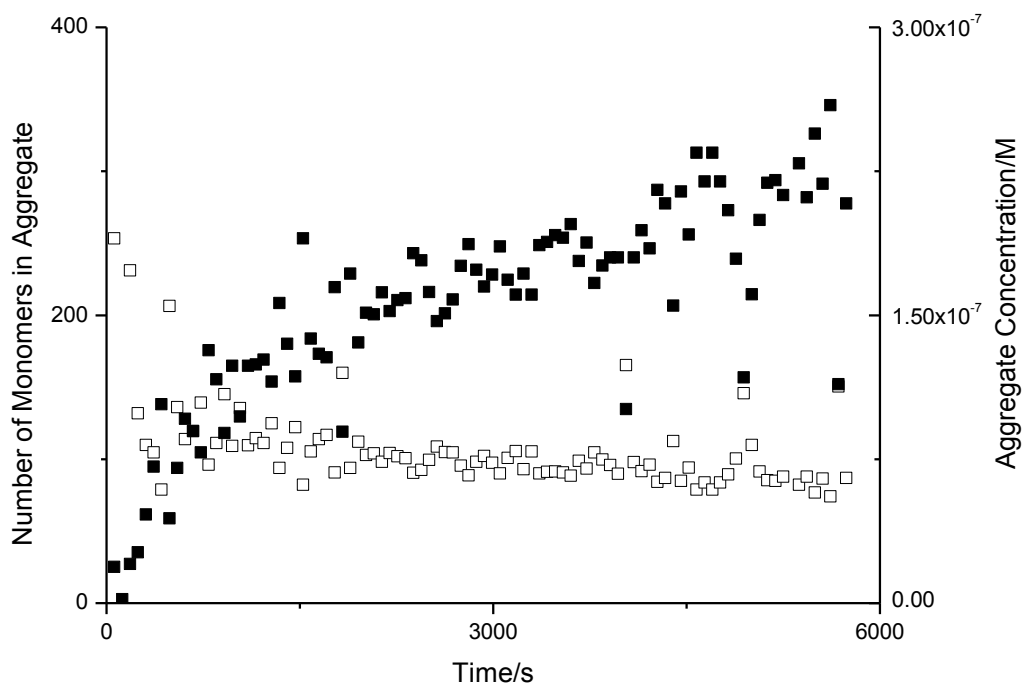
Figure 5.32: Number of monomers in aggregates (closed squares) and aggregate concentration (open squares) for BSA at 75 ℃. Errors are of magnitude replicated in Fig 5.30.

Figure 5.31 illustrates the aggregation behaviour of BSA at high temperature. Initially the aggregate concentration starts high ($\sim$1.4$\times$10^{-7} M). This is the first detectable aggregate concentration, therefore these aggregates are considered to have formed at sizes and concentrations which the model is unable to separate from the scattering produced from the monomers. The closed squares in Figure 5.31 indicate that the initial aggregates consist of 30 monomers before rapid growth in size, with an "S" shaped growth where the slower phase of aggregate growth is occurring at around 8000 monomers in size and the aggregate concentration remaining stable at $\sim$ 2.9$\times$10^{-9} M.

Figure 5.32 illustrates a similar trend to that seen in at 80 ℃, with the aggregate concentration decreasing whilst the size of aggregates increases. Both plots are less extreme than those seen at 80 ℃, where the resultant aggregates contain fewer monomers.
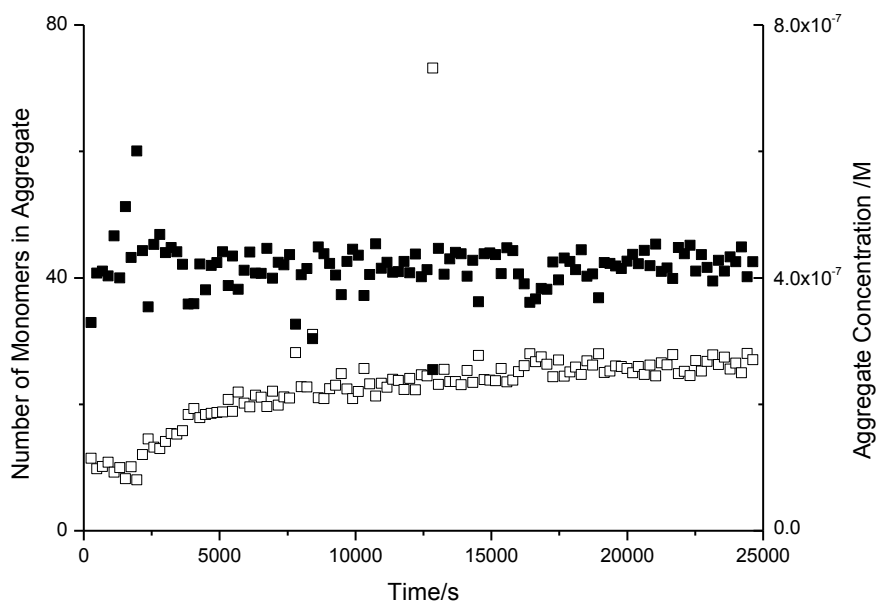
Figure 5.33: Number of monomers in aggregates (closed squares) and aggregate concentration (open squares) for BSA at 70 ℃. Errors are of magnitude replicated in Fig 5.31.

Figure 5.33 illustrates the aggregation behaviour of BSA at 70 ℃. Here it can be seen that the number of monomers in an aggregate are present at a consistent size, representative of around 42 monomers in each aggregate. The concentration of aggregates shows an initial decrease (which may be within experimental error) before a rapid growth with an exponentially decreasing rate. The concentration of the aggregates maintains a steady increase through the latter stages of the experiment which it appears would continue past the final timeframe of the experiment.

At the high temperatures, the decrease in aggregate concentration indicates that small aggregates may be combining together to form larger aggregates, which certainly appears possible as the monomer content in each aggregate at minimum doubles over each time frame. This appears a more robust argument for the aggregate concentration decrease over the unstable aggregates reverting to monomeric form, as at a monomer count of 30, the aggregate would be expected to be stable as it is much larger than other examples of stable aggregate [20, 21]. A similar mechanism can be applied to the aggregation behaviour at 75 ℃, as Figure 5.32 exhibits the same attributes as Figure 5.31.

This aggregation behaviour can be described in terms of the unfolding of the BSA molecule. Thermal unfolding of BSA begins at around 60 ℃ and continues until 90 ℃ [12], however the shift seen with the behaviour in 90 % D_2O is shifted down by ~5 ℃ (Figure 5.16). At 80 ℃ the BSA monomer is structurally perturbed and a large fraction of the molecule has exposed hydrophobic regions which are highly "reactive" (capable of self association and). In order for protein molecules to associate, hydrophobic regions on the each colliding protein must interact. The results indicate that these "reactive" monomers interact with each other in a manner which maintains or increases the overall "reactive" surface area percentage of the small aggregate. As the concentration of free

monomer is solution is still large (Figure 5.28), it is the increased reactivity of the small aggregates which drives their self association until they decrease in concentration. These small aggregates have a much higher "reactivity" than the perturbed monomer, therefore collisions between small aggregates is the driving force in aggregate growth, characterised by the large aggregate growth change whilst there is little monomer depletion. Although this appears novel for particulate aggregates, the association of oligomers during a protein nucleation "burst phase" has been modelled in the formation of fibrils [22, 23].

The resultant aggregates formed from the clustering of small aggregates are less "reactive" than the small aggregates. A separate explanation would be that the aggregates are at such a low concentration in relation to the concentration of monomer that they fail to interact. However, at an aggregate concentration of 6×10^{-8} M, there is a markedly different rate of aggregate growth at 80 ℃ and 75 ℃ (Figure 5.31, Figure 5.32). At 80 ℃ at a concentration of 6×10^{-8} M aggregates, proposed small aggregate clustering is occurring, whilst at 75 ℃, it is the proposed monomer addition. This indicates that the reactivity of the respective aggregates must differ. The emergence of the larger, less "reactive" aggregates results in monomer addition becomes the prevailing growth mechanism, which would be more thermodynamically favourable than new aggregate nuclei forming [24], where aggregate concentration is seen to be stable, and aggregate growth slow, due to the small impact on radius each subsequent monomer provides to the aggregate (Figure 5.31), and the low reactivity of the two particles (when compared to the small aggregates). A similar method is true for the experiment at 75 ℃, where the extent of monomer unfolding is less than that at 80 ℃, but still sufficient enough to increase the "reactivity" of the small aggregates, leading to aggregate self association as before.

At 70 ℃ however (Figure 5.33) the extent of protein monomer unfolding is low. As a result, structurally perturbed monomers will still associate, however the "reactive" hydrophobic surface are of the aggregate may not increase. Therefore as the aggregate gets larger, the percentage of "reactive" surface area decreases. With growth this would decrease to such an extent that the aggregate becomes essentially non-reactive, as the chance of a "reactive" monomer colliding with the "reactive" surface area on the aggregate becomes miniscule. Therefore self association of "reactive" monomers is the sole method of aggregation when aggregates reach sufficient size, which results in both the size of aggregates remaining constant throughout the experiment, and the aggregate fraction increasing, with the rate relative to the monomer concentration. Both of these features can be seen in the plot in Figure 5.33.

5.7 Data Analysis – Final Aggregate Structure

Samples pre-aggregate at temperatures between 70 ℃ to 78 ℃ in 2 ℃ increments were studied using SANS in order to determine the final overall structure of the aggregates. Additionally any trend between incubation temperature and structural features was investigated.
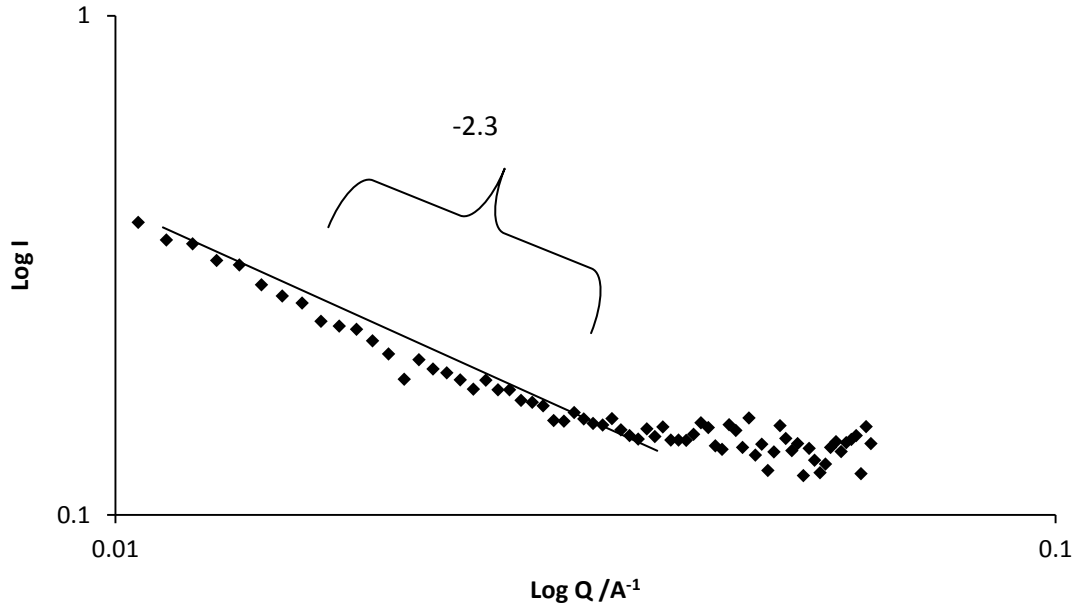
Figure 5.34: A log scattering plot of BSA aggregates formed at 78 ºC. Error bars are the size of the points.

Figure 5.34 illustrates the analysis of the scattering plots in order to obtain a fractal parameter for the aggregates. The slope of -2.3 indicates a mass fractal with irregular structure, possibly resembling that of a Gaussian chain, which would ideally give a value of -2. This represents the joined spherical aggregates clustered together in the form of the strand or aggregate network. Similar findings have been seen with the use of SEM [16].

Figure 5.35 indicates the fractal parameters of the scans run on the 10 separate aggregate samples. It can be seen that there is little variation, or trend with temperature, in addition the repeat samples show good reliability in the samples. It can be assumed that there is no significant change in structure to the overall clustered aggregate, despite the effect temperature has on the size of the spherical aggregates [16]
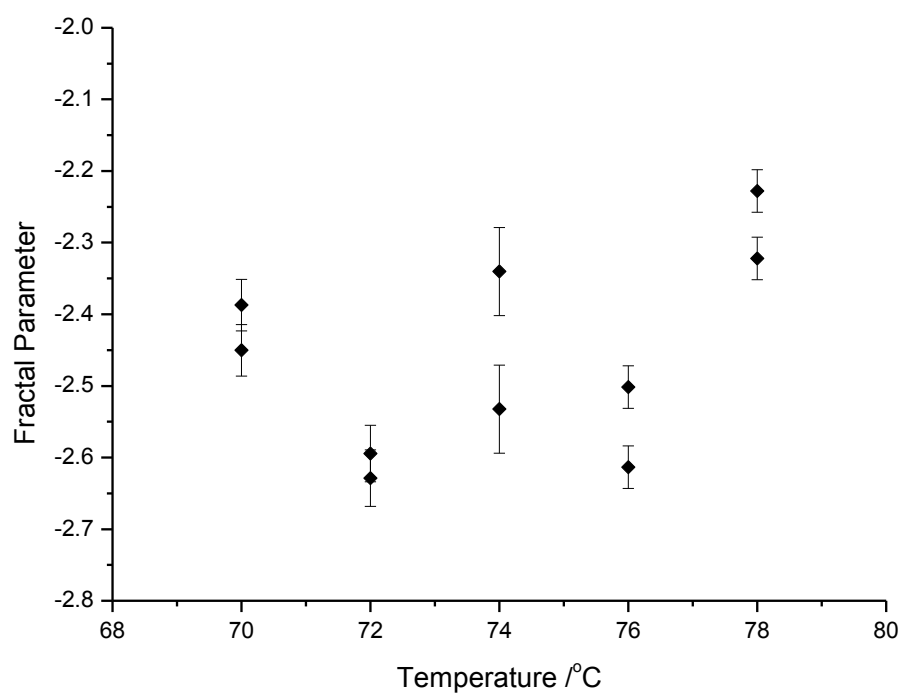
Figure 5.35: Fractal parameters for BSA aggregated at varied temperatures. Each temperature value contains a repeat.

5.8 Conclusion

As this work complements work undertaken in following sections, the conclusions from this work are present within the conclusion section (Chapter 8), to prevent repetitive statements and allow cross-chapter discussion.

5.9 References

1.   Dewhurst, C., *Graphical Reduction and Analysis SANS Program for Matlab™.* 2003: Grenoble.

2.   Lindner, P. and T. Zemb, *Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter (North-Holland Delta Series).* 2002: North Holland.

3.   R., et al., *Effect of Urea on Bovine Serum Albumin in Aqueous and Reverse Micelle Environments Investigated by Small Angle X-Ray Scattering, Fluorescence and Circular Dichroism.* Brazilian Journal of Physics, 2004. **34**: p. 58-63.

4.   Yohannes, G., et al., *Thermal aggregation of bovine serum albumin studied by asymmetrical flow field-flow fractionation.* Analytica Chimica Acta, 2010. **675**(2): p. 191-198.

5.   Shweitzer, B., D. Zanette, and R. Itri, *Bovine serum albumin (BSA) plays a role in the size of SDS micelle-like aggregates at the saturation binding: the ionic strength effect.* Journal of Colloid and Interface Science, 2004. **277**(2): p. 285-291.

6.   *The RSCB Protein Data Bank.* Available from: http://www.rcsb.org/pdb. Accessed Jan 2012

7.   Svergun, D.I., et al., *Protein hydration in solution: Experimental observation by x-ray and neutron scattering.* Proceedings of the National Academy of Sciences, 1998. **95**(5): p. 2267-2272.

8.   Svergun, D., C. Barberato, and M.H.J. Koch, *{{\it CRYSOL} {--} a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates}.* Journal of Applied Crystallography, 1995. **28**(6): p. 768-773.

9.   Koutsioubas, A., et al., *Slow and remanent electric polarization of adsorbed BSA layer evidenced by neutron reflection.* Soft Matter, 2012. **8**(9): p. 2638.

10. Jacrot, B., *The study of biological structures by neutron scattering from solution.* Reports on Progress in Physics, 2001. **39**(10): p. 911.

11. Ptitsyn, O.B., *Molten Globule and Protein Folding*, in *Advances in Protein Chemistry Volume 47.* 1995, Elsevier. p. 83-229.

12. Michnik, A., *Thermal stability of bovine serum albumin DSC study.* Journal of Thermal Analysis and Calorimetry, 2003. **71**(2): p. 509-519.

13. Glatter, O. and O. Kratky, *Small angle X-ray scattering.* Vol. 102. 1982: Academic press London.

14. Michnik, A., et al., *Comparative DSC study of human and bovine serum albumin.* Journal of Thermal Analysis and Calorimetry, 2006. **84**(1): p. 113-117.

15. Militello, V., V. Vetri, and M. Leone, *Conformational changes involved in thermal aggregation processes of bovine serum albumin.* Biophysical Chemistry, 2003. **105**(1): p. 133-141.

16. Krebs, M.R.H., G.L. Devlin, and A.M. Donald, *Protein particulates: another generic form of aggregation?* Biophysical Journal, 2007. **92**: p. 1336-1342.

17. Krebs, M.R.H., K.R. Domike, and A.M. Donald, *Protein aggregation: more than just fibrils.* Biochemistry Society Transactions, 2009. **37**(4): p. 682-686.

18. Pedersen, J.S., *Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter (North-Holland Delta Series)*, ed. P. Lindner and T. Zemb. 2002: North Holland.

19.   González Flecha, L. and V. Levi, *Determination of the molecular size of BSA by fluorescence anisotropy.* Biochem. Mol. Biol. Educ., 2003. **31**(5): p. 319-322.
20.   Kim, K.S., J.C. Neu, and G.F. Oster, *Many-body forces between membrane inclusions: A new pattern-formation mechanism.* EPL (Europhysics Letters), 2007: p. 99.
21.   Matsumoto, K., et al., *Distinct Interaction of Versican/PG-M with Hyaluronan and Link Protein.* Journal of Biological Chemistry, 2003. **278**(42): p. 41205-41212.
22.   Li, M.S., et al., *Probing the mechanisms of fibril formation using lattice models.* The Journal of Chemical Physics, 2008. **129**(17): p. 175101.
23.   Wu, C. and J.-E. Shea, *Coarse-grained models for protein aggregation.* Current Opinion in structural biology, 2011. **21**(2): p. 209-220.
24.   Kopito, R.R., *Aggresomes, inclusion bodies and protein aggregation.* Trends in cell biology, 2000. **10**: p. 523-530.

Chapter 6 : Aggregation Kinetics

6.1 Introduction

This chapter details experiments performed using ultra violet light scattering spectroscopy to elucidate the particulate aggregation kinetics on three protein structures; bovine serum albumin, lactoglobulin, and myoglobin. The change in aggregation rate across a temperature range is measured to extract the kinetics of each system, and the effect changes to the protein solution have on the aggregate growth kinetics is also studied.

6.2 Experimental Overview
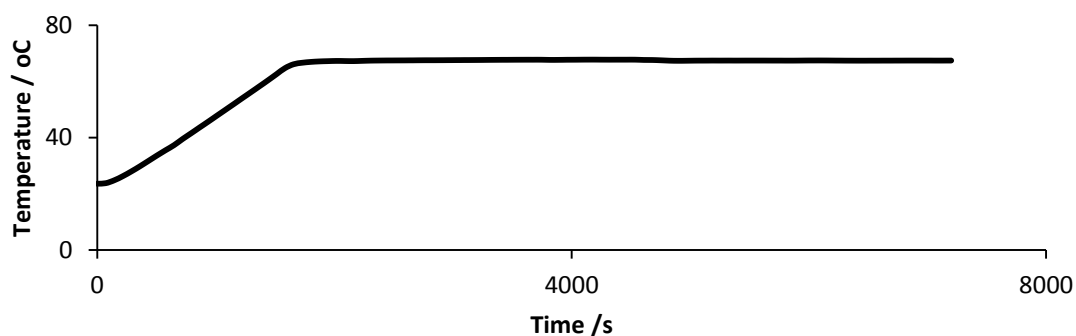
Figure 6.1: Temperature profile recorded for a heat ramp up to an isothermal temperature of 67.5 ℃

Protein in buffer solution at a concentration of 2 mg/mL was placed into a temperature controlled UV spectrophotometer, initially at 25 ℃. The sample was then heated up and held at the desired temperature (Figure 6.1). During both the heating rate and the isothermal hold, absorbance spectra were taken of the protein sample across a range of wavelengths. At the exact time a spectrum is taken a temperature probe registered the temperature within the protein sample to provide a data set which contains a spectra and temperature value for each time point in the aggregation process. This data collection is continued until the aggregation is complete, at which point the data logging process is

stopped. Further details on this procedure can be found within the methods and materials section.

6.3 Data Analysis

This section details the analysis process undertaken on each aggregation experiment to produce the data discussed further in this chapter. One experiment is used as an example throughout to show the process and comparable features of each plot.

In order to ascertain a suitable temperature range over which to perform the aggregation experiments, an initial experiment on a protein in solution was performed whereby the temperature was increased at a constant rate of 0.02 ℃/min from 25 ℃ to 95 ℃. The absorbance of the protein was measured at a single wavelength; 320 nm. This experiment would yield information on at which temperature aggregation begins to be detectable using this experimental setup. An example is shown for myoglobin (Figure 6.2).
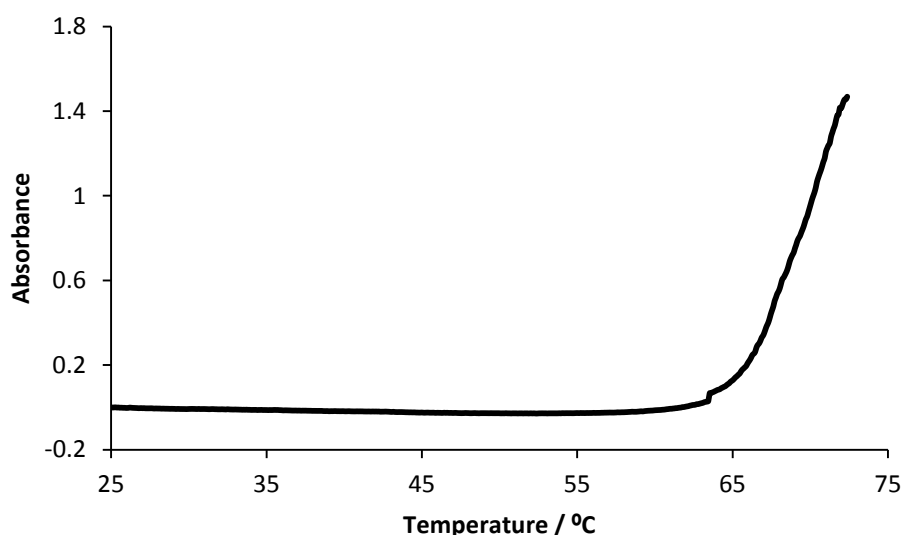
Figure 6.2: Temperature profile of β -lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8. heated at 0.02 ℃/min

At 63 ℃ there is a detectable change in the absorbance. Beyond this value aggregation is prevalent, indicated by the continued increase in absorbance. This appears to agree well with ranges stated in other publications, however some attention must be paid to the role the solution effects have on the aggregation temperature [1, 2]. The value obtained here was then used as a starting reference for studies using this sample.
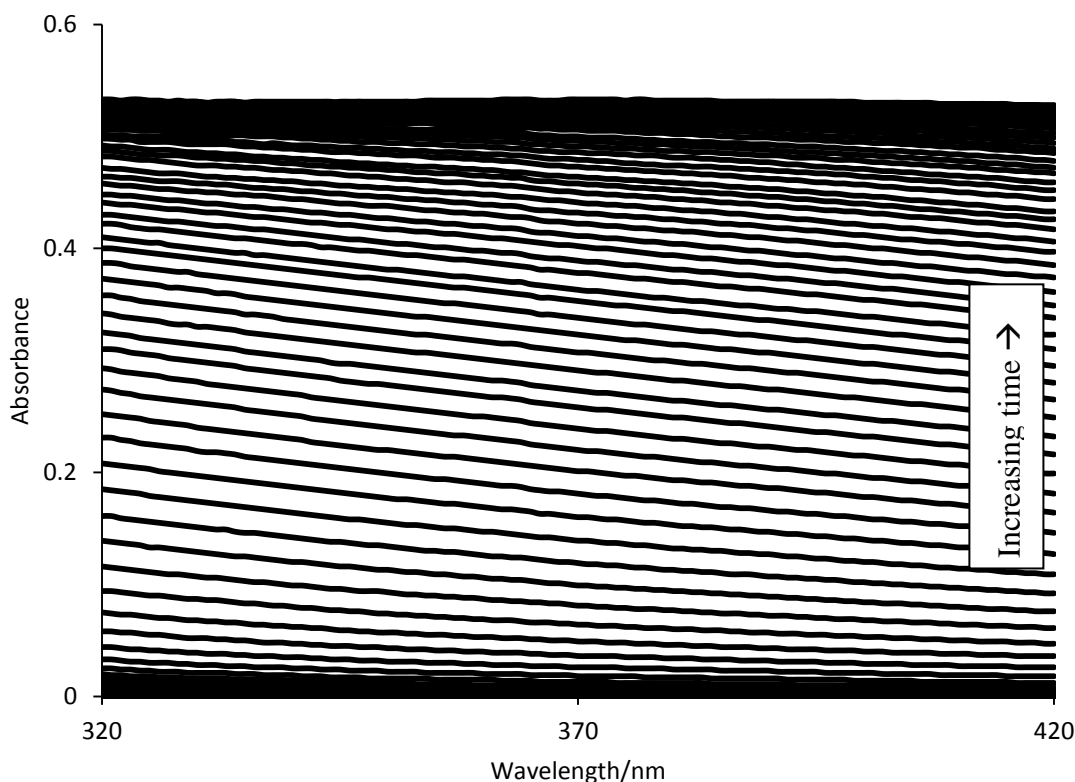
Figure 6.3: Absorbance spectra of lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8 heated to 68.1 ℃. Time duration is depicted by increasing absorbance.

Figure 6.3 illustrates spectra recorded for lactoglobulin whilst heating to 68.1 ℃. As the the protein is exposed to the high temperature over time, the absorbance increases. It can be seen that at the lower wavelengths (early aggregation), the increase in absorbance is larger than at the higher wavelengths (late aggregation). This wavelength dependent scattering will be used to extract information on the particle sizes within the protein sample. It can be seen initially that there is very little change in absorbance as

the protein solution is heated up until aggregation begins; characterised by the increase in absorbance across the spectrum. This increase is due to an increase in scattering; either due to more particles scattering the light, larger particles scattering the light, or a combination of the two. Towards the top of the scale the change in absorbance spectra abates; this indicates no change in scattering is occurring so the particles are no longer growing in size or number. Finally the drop in absorbance is representative of the aggregates coming out of solution and settling towards the bottom of the cell outside of the beam area, indicating an apparent drop in aggregate concentration. This stage is known as clustering of which there is no known model, therefore data collection stops once the absorbance peak is reached. This therefore indicates an end to the aggregation process.
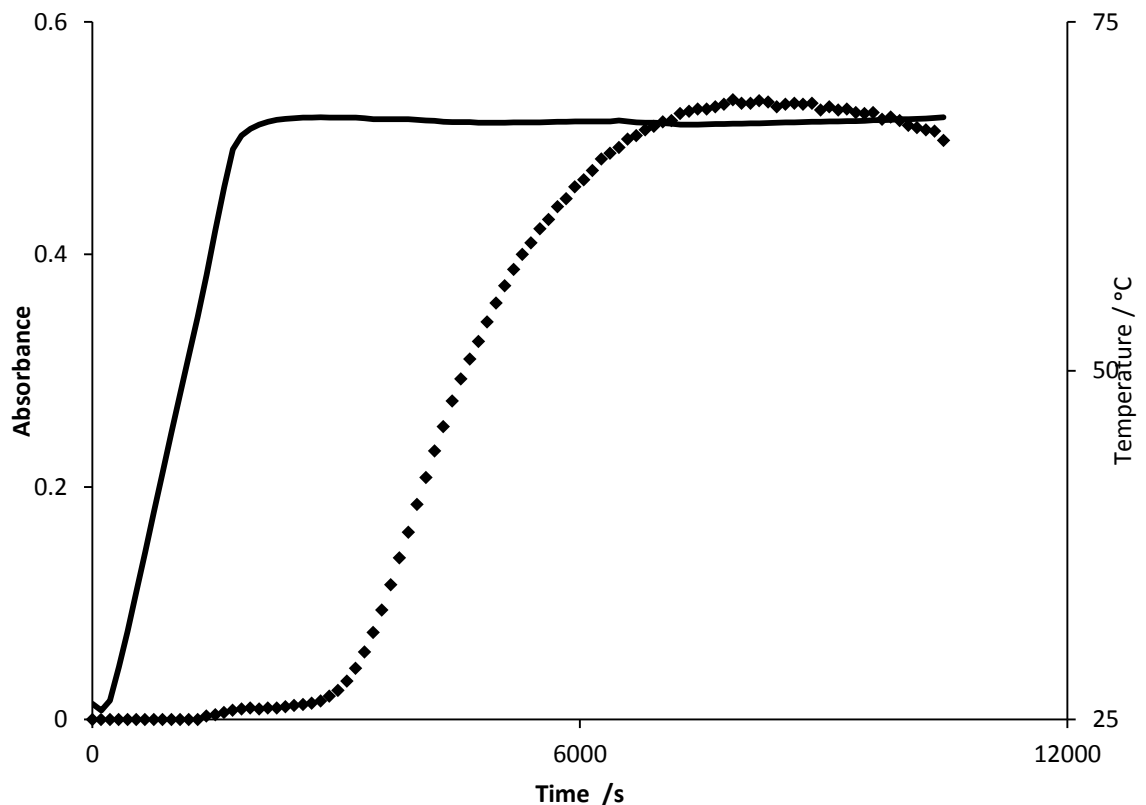
Figure 6.4:Aborbance at 320 nm over time of lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8. Temperature is indicated by the line.

Figure 6.4 illustrates the same experiment as Figure 6.3, shown from the absorbance at 320 nm with time values associated with them. The initial flat period at the start of the graph relates to the heating up of the protein in solution at a constant heating rate from 25 ℃. The start of the absorbance increase corresponds well with the temperature reaching its maximum value. Whilst the temperature is held isothermally, aggregation proceeds, as indicated by the increasing absorbance, until it reaches its maximum. At this point the aggregation process is complete. After the maximum absorbance peak, the decrease in absorbance can be interpreted as the aggregates decreasing in size or number, or this could also be caused by the particles clustering, causing localised concentration fluctuations, which would be seen as a decrease in absorbance [3].

In order to analyse the spectra collected from the aggregation experiments a correction needs to be applied to the raw data. The spectrophotometer does not take absorbance measurements at each wavelength at the same time, therefore each spectra needs correcting to the absorbance value of the sample at the time the first spectrum value was taken. In order to perform this correction, the time period for the spectrophotometer to take a single scan was recorded across the applicable wavelength range.

Table 6.1:Time between first data point collection and selected wavelength collection point.

Wavelength /nm	Time from first collection point /s
420	0
400	2.66
380	5.32
360	7.98
340	10.64
320	13.3

In order to correct the data so that all absorbance values in a spectrum correspond to the same time point, values from Table 6.1 were used to correct the raw spectrum data.
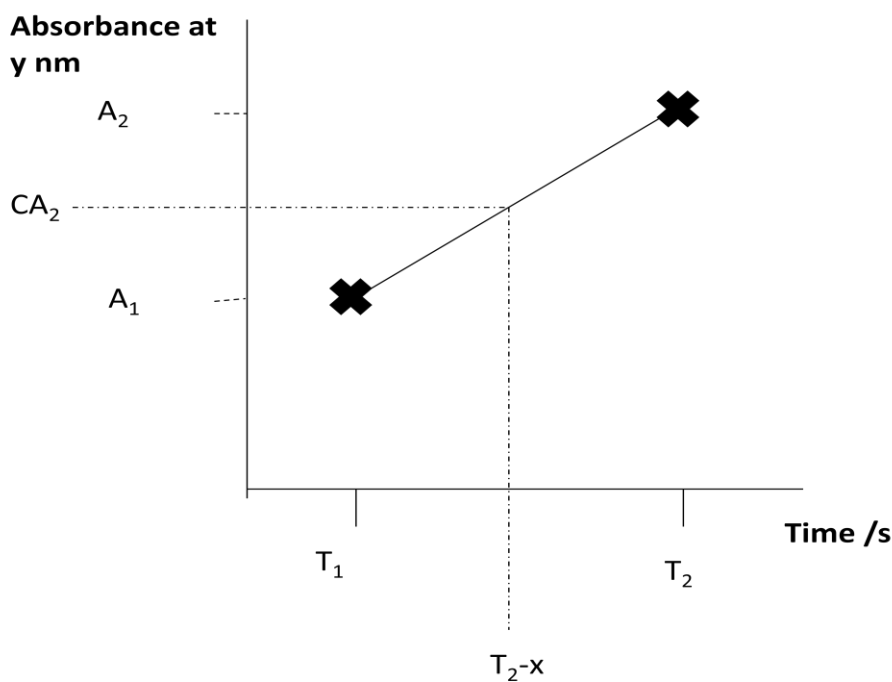
Figure 6.5: Calculation of corrected absorbance values

$$CA_2 = \frac{(A_2 - A_1)(T_2 - x - T_1)}{(T_2 - T_1)} + A_1$$

Equation 6.1: Equation for calculation of corrected absorbance.

Figure 6.5 and Equation 6.1 illustrate the calculation for the corrected absorbance value of the second spectra CA_2 at y nm. Where x is the time correction value taken from Table 6.1, A_1 and A_2 the absorbance values for the first and second spectra absorbance value at y respectively and T_1 and T_2 the time values for the first and second absorbance values at y. This is performed for all absorbance values across each wavelength of each spectra. As can be seen from Figure 6.6 which illustrates the full extent of the corrections, there is no difference between raw and corrected data at the large wavelengths as this is where the absorbance scan begins, however the discrepancy becomes larger as the wavelengths get shorter. It is worth noting that this correction method assumes a linear change in absorbance between each successive spectra,

therefore it is important to have a large number of spectra in an aggregation experiment

in order to maintain the accuracy in the shape of the absorbance curve.
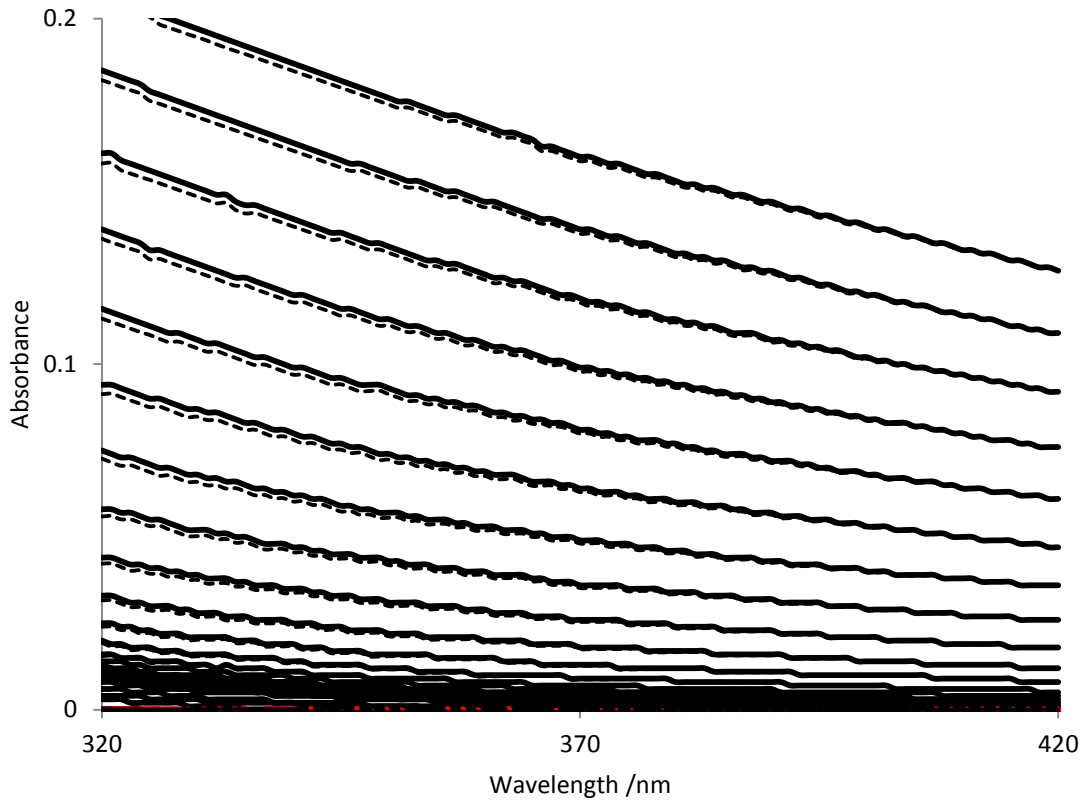
Figure 6.6: Corrected absorbance spectrum of lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8
(dotted) compared with raw data values (black).

The scattering exponent must be determined from these corrected spectra in order to

determine the particle size. The wavelength dependent scattering follows a power law

where:

$$A(\lambda) = a\lambda^{\beta}$$

Equation 6.2 Equation for wavelength dependent absorbance.

The absorbance A at wavelength λ is given by Equation 6.2 where α is a constant and β

is the scattering exponent. This exponent can be calculated from the fit of Equation 6.2

to each spectra or by obtaining a gradient from a plot of log A versus log λ, see Equation

6.3.

$$\log(A(\lambda)) = \log(a) + \beta \log(\lambda)$$

Equation 6.3: Scattering exponent calculation in the terms of y=mx+c.

This scattering exponent is calculated for each spectra obtained within the experiment to give a scattering exponent for each time value calculated. An example of such can be seen in Figure 6.7.
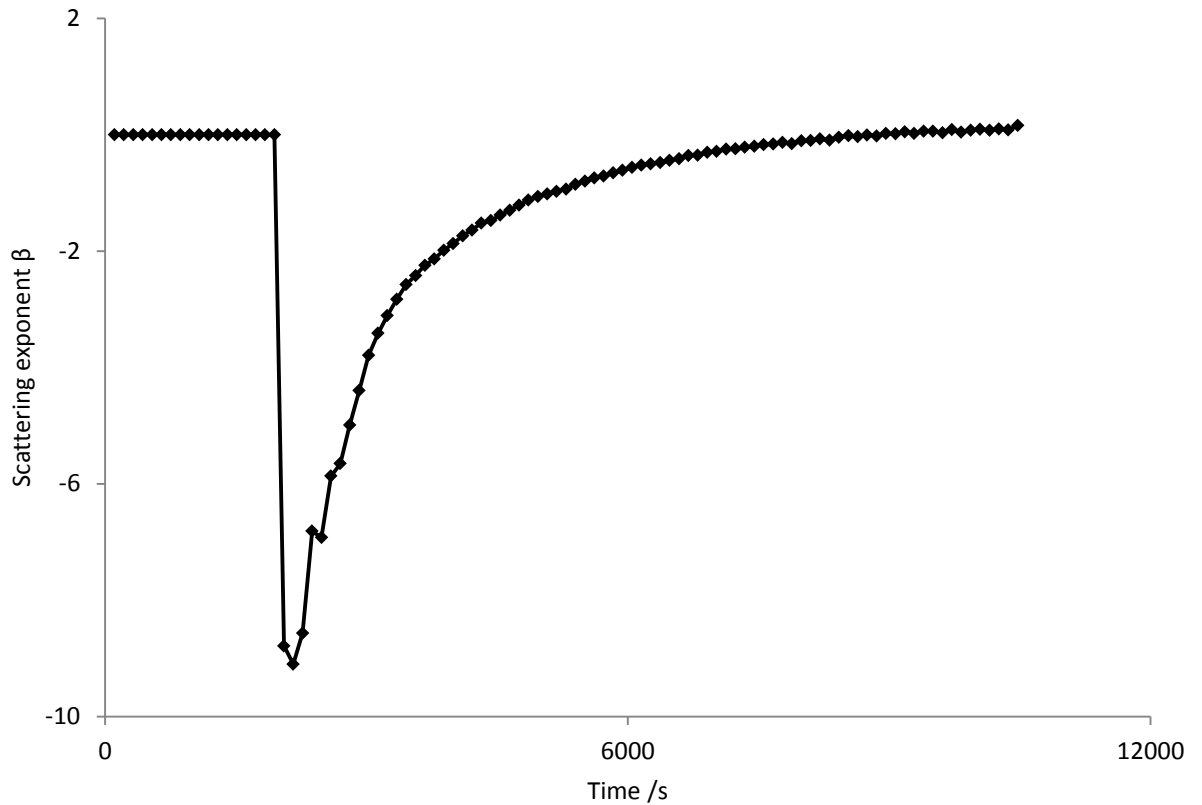
Figure 6.7: Scattering exponent lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8. at 67.1 ℃

Figure 6.7 shows how the scattering exponent has a meaningful value when the absorbance in the solution begins to rise. Before the drop in the scattering exponent, there is no absorbance or scattering, and so the exponent is zero. The scattering exponent begins to rise in conjunction with the rise in absorbance seen in Figure 6.4 and will even continue to rise despite the absorbance falling off after reaching a peak. This may indicate that there are larger structures present, despite the absorbance tailing off, indicating a decrease in concentration. This gives some support to the idea of clustering

occurring after the absorbance has reached its peak and particulate aggregate formation and growth has occurred.

The theory section (Chapter 3.4) details how the scattering exponent can be converted in to a spherical protein diameter using the refractive indices of the protein and the medium. This procedure produces a sixth order polynomial function which allows the conversion of scattering exponent to diameter.
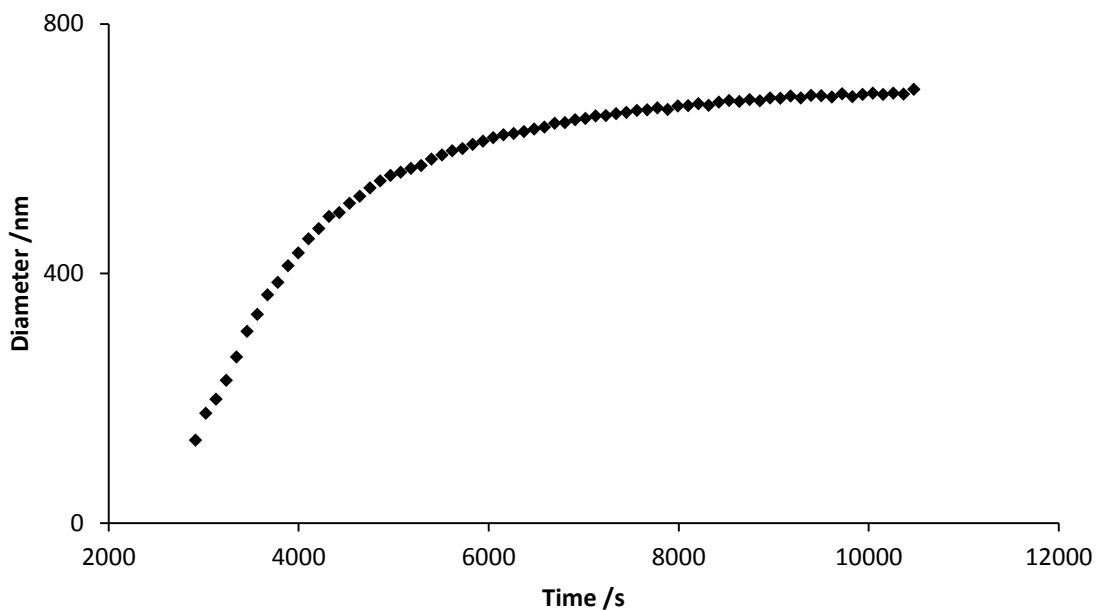
Figure 6.8: Diameter versus time plot of lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8. at 67.1 °C. Error bars are the size of the points.

As can be seen in Figure 6.8, the plot shows very similar resemblance to that of Figure 6.7, except the values of the y axis now represent the diameter values of the protein aggregate. It is also worth noting that before the diameter increase there are some missing points from the plot. This is a common feature during this part of the plot, where the scattering exponent value converts to a negative diameter value which is clearly nonsensical. This highlights a limitation within the experiment where the smallest sizes detected are around 50-100 nm based on the protein system and can be attributed to the constraints of using UV light to probe such small length scales and also

the rapid dynamic growth of the system at these time points at length scales below 50 nm.

The growth curve in Figure 6.8 grows exponentially before levelling out at a maximum diameter value. In order to determine the number of aggregates formed within the sample, this maximum diameter value must be accurately determined. This is performed by fitting an exponential function to the diameter growth data (Equation 6.4). Where the diameter; d at time; t is determined from the maximum diameter d_m, a fitting parameter d_0, and a time constant τ.

$$d(t) = d_m - d_0 e^{-t\tau}$$

Equation 6.4: Fitting function to determine final diameter value.
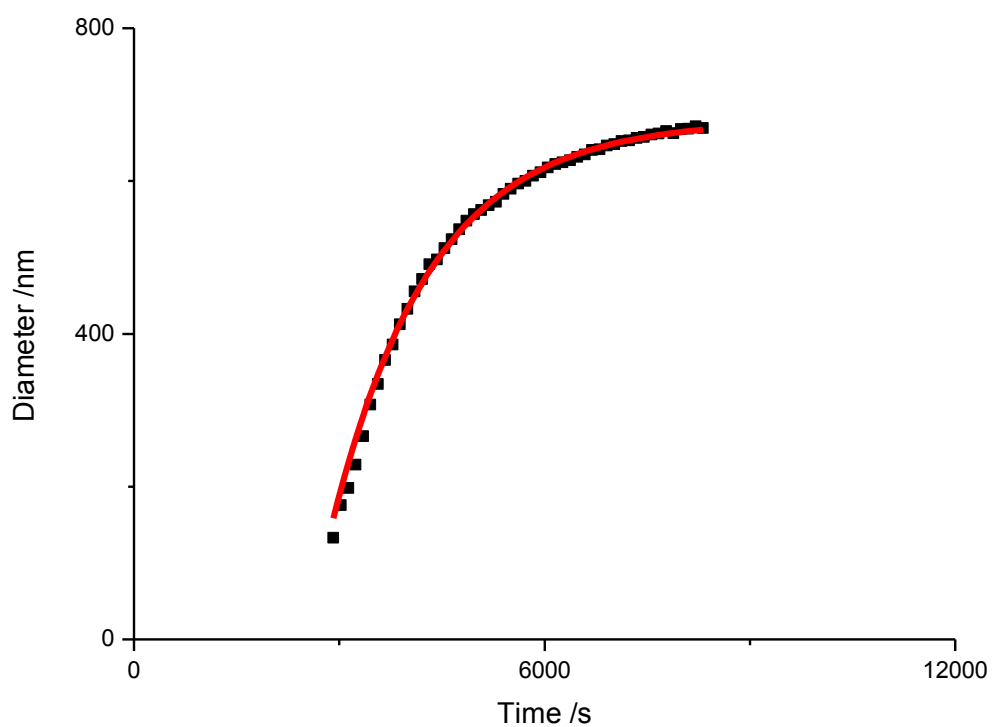
Figure 6.9 Fit of Equation 6.4 to diameter values of in 100 mM sodium phosphate, 100 mM NaCl pH 5.8 at 67.1 ℃

When the fit of the function from Equation 6.4 is applied to the data in Figure 6.8 it can be seen that the equation fits very well to the data (Figure 6.9). To ensure a successful fit, the data towards the end of the incline was used over that at the start of the aggregation process as it leads to a more accurate determination of the final diameter than that data at the start of the run, as determined by R^2 values.

It is taken that the protein aggregates have reached their maximum size due to total depletion of the monomer, therefore all the protein is present within the aggregate. It is therefore possible to calculate the number of protein monomers that make up each aggregate A_{nm}, assuming each aggregate is identical in size (this is a fair assumption as formation of particulate monomers have been seen to have a uniform size distribution [4] (Equation 6.5)).

$$A_{nm} = \frac{\frac{4}{3}\pi \left(\frac{d_m}{2}\right)^2}{V_m}$$

Equation 6.5: Calculation of aggregate number for maximum particulate diameter.

Where V_m is the volume of the protein molecule, calculated using the web program VADAR [5], which calculates the volume of a protein molecule from its crystallographic structure taken from the PDB [6].

The maximum aggregate number A_{nm} can then be used to calculate the concentration of aggregates C_a present within the sample (Equation 6.6), where C_{m0} is the known initial concentration of the monomer:

$$C_a = \frac{C_{m0}}{A_{nm}}$$

Equation 6.6: Calculation of aggregate concentration

With the value for the concentration of aggregates calculated, it is possible to calculate the value for the concentration of monomer present at each time value during the aggregation process. This is can be calculated from each diameter throughout the time course of the experiment. An equation similar to that of Equation 6.5 is used to calculated the aggregation number at each time point, $A_n(t)$, from each diameter d (Equation 6.7).

$$A_n(t) = \frac{\frac{4}{3}\pi\left(\frac{d(t)}{2}\right)^2}{V_m}$$

Equation 6.7: Calculation of number of monomers in each protein aggregate at each time point.
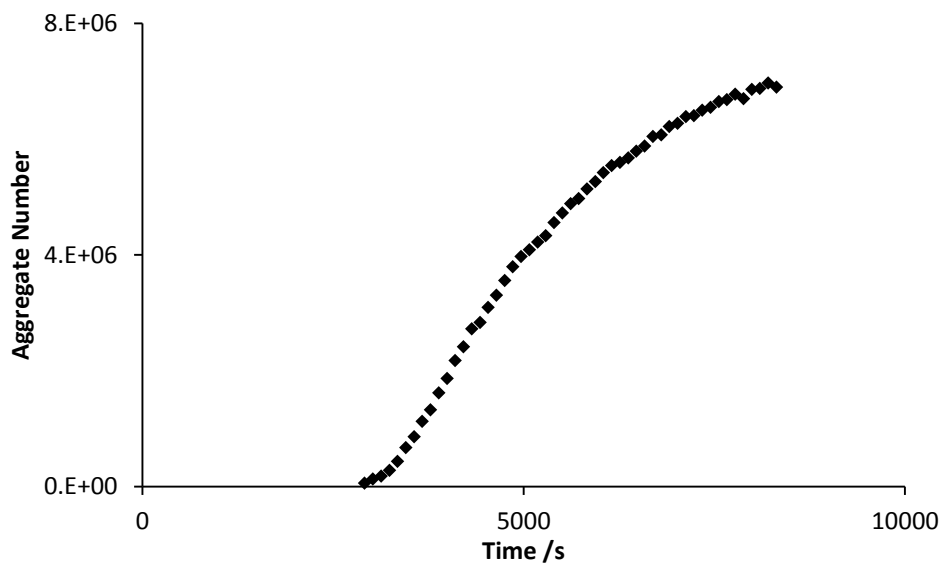
Figure 6.10: Aggregate number versus time for lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8 at 67.1 ℃. Error bars are the size of the points.

Figure 6.10 shows the aggregate number produced for each diameter from Equation 6.7. The plot range has been edited to include only viable data, where data from negative diameters has been omitted, and also data beyond the peak; where clustering is thought to be occurring, has also been removed. Barring omissions and the y-axis scale Figure 6.10 looks very similar to Figure 6.8, as it is a scaling factor which has been applied to the diameter plot.

The values generated in Equation 6.7 can be used to calculate the concentration of monomer at each respective time point; $C_m(t)$.

$$C_m(t) = C_{m0} - (C_A \times A_N(t))$$

Equation 6.8 is the final step in converting the diameter values into monomer content in solution. This allows for a plot of monomer concentration versus time to be produced.
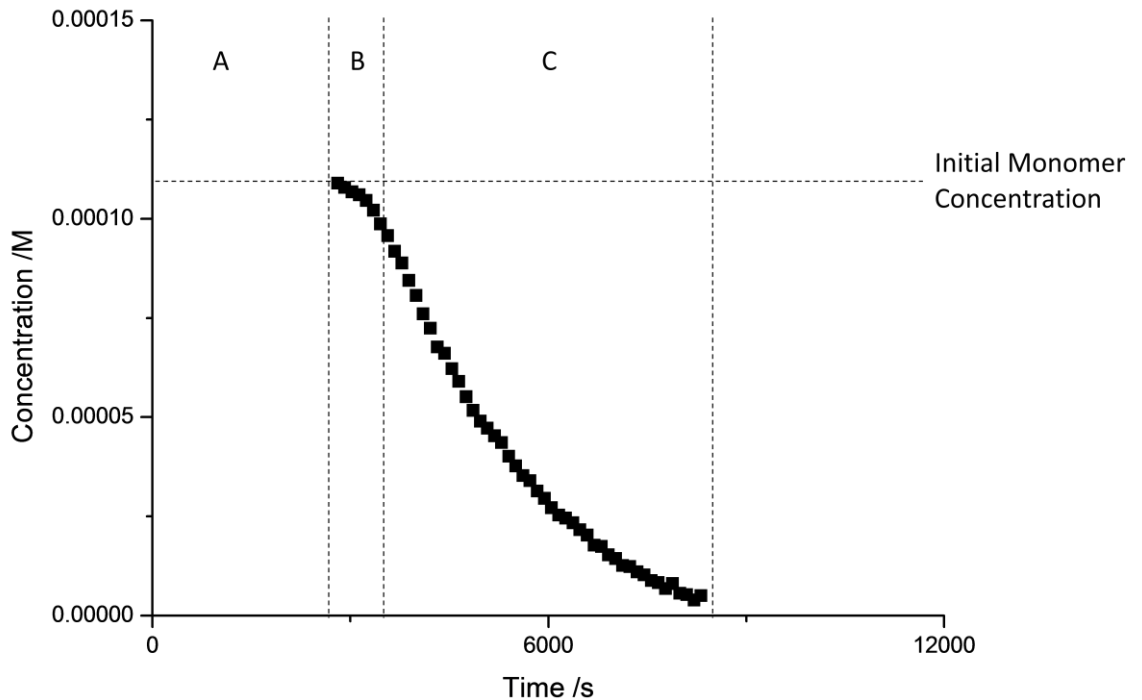
Figure 6.11: Concentration of monomer over time for lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8 at 67.1 °C. Sections A,B, and C indicate different regions within the curve. Error bars are the size of the points.

Figure 6.11 shows the concentration values produced from the example data using Equation 6.8. Section A of the data indicates the period of the experiment before aggregation. The plot does not show the actual value of the monomer concentration, in this case 1.09×10^{-4} M. This is because there is no input value for the diameter, as the size of the monomer is too small to be detected using this technique. Section B relates to

the formation of the protein aggregate population, where the ever increasing rate is only limited by the surface area of aggregates; which is increasing. Section C indicates the growth phase of the particulates, where the rate of monomer depletion is decreasing over time. Here the rate limiting factor is the concentration of available monomer, as this decreases; so does the rate.

The process of aggregation is usually considered to be a second order reaction, as the rate depends on both the concentration of the monomer and the concentration of the aggregates. However, in this aggregation process the aggregate concentration is static once established (as verified by monodisperse particulate aggregate sizes under SEM [4]). With this, the particulate aggregation process is a pseudo first order reaction and therefore can be modelled as such with Section C in Figure 6.11 being fit to a first order rate model for the depletion of monomer. This can be seen in Equation 6.9, where k_{ob1} is the first order rate constant.

$$C_m = C_{m0} e^{(-k_{ob1}t)}$$

Equation 6.9: First order rate equation for depletion of monomer.

Figure 6.12 illustrates the fit of Equation 6.9 to the growth section of the example experiment, where a good fit is achieved. The first order rate constant; k_{ob1} (s^{-1}), incorporates all factors responsible for the rate of the reaction, including the concentration of aggregates.
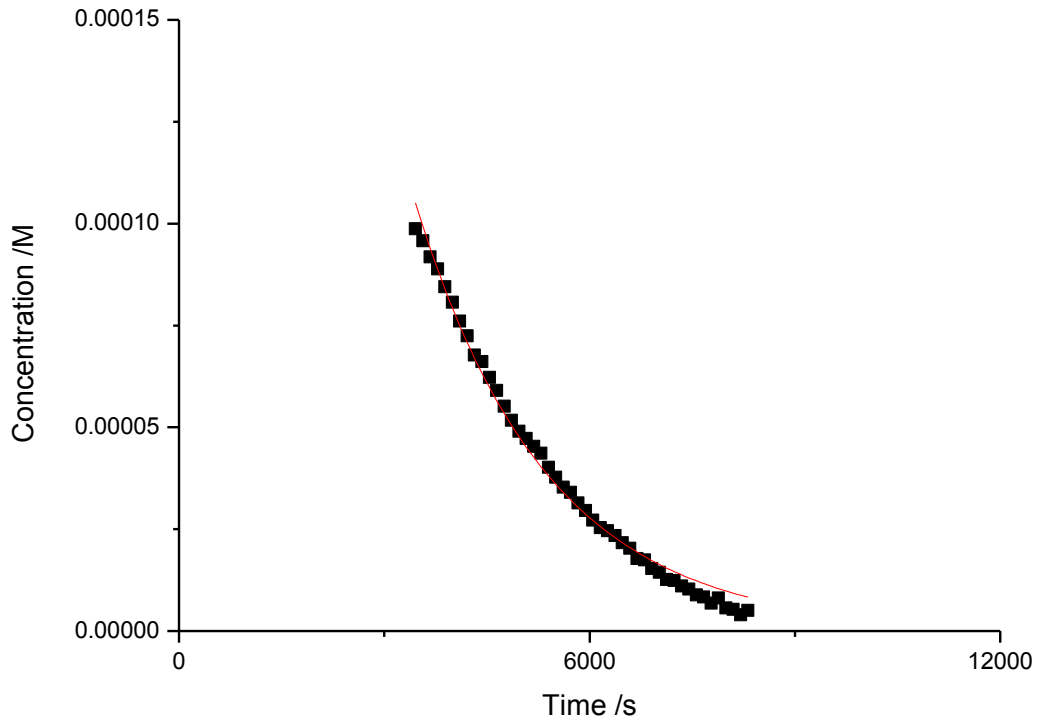
Figure 6.12: Fit of Equation 6.9 (red line) to growth phase of lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8 at 67.1 ℃. Error bars are the size of the points.

$$k_{ob2} = \frac{k_{ob1}}{C_A}$$

Equation 6.10: Equation for second order reaction rate calculated from pseudo first order rate.

Equation 6.10 illustrates how the observed second rate constant k_{ob2} (dm^3 mol^{-1} s^{-1}) can be calculated from the observed first order rate and the aggregate concentration ; C_A (mol dm^{-3}). This second order rate relates the change in monomer concentration to the aggregate concentration and monomer concentration by Equation 6.11.

$$\frac{dC_m}{dt} = k_{ob2}C_M C_A$$

Equation 6.11: Second order rate equation

It is at this point that the model for particulate protein aggregation is presented, as seen in the following equations.

$$a.\ N \rightleftarrows I$$

$$where: \quad N \overset{k1}{\rightarrow} I \quad N \overset{k2}{\leftarrow} I$$

$$b.\ I + I \overset{k3}{\rightarrow} A_2$$

$$c.\ I + A_m \overset{k4}{\rightarrow} A_{m+1}$$

$$d.\ A_m + A_n \overset{k5}{\rightarrow} A_{m+n}$$
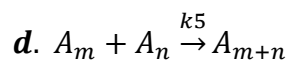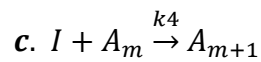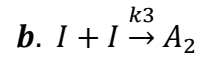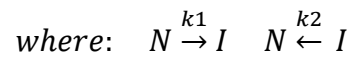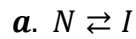
Equation 6.12: Particulate protein aggregation pathway [3].

Equation 6.12 a illustrates the structural perturbation of the protein that occurs before aggregation. *N* represents the native state of the protein molecule which is in equilibrium with an aggregation prone conformation of the protein; an intermediate state; *I*. Both the forward and reverse reactions have rate constants associated with this initial process. Equation 6.12 b represents the nucleation of particulate aggregates where two aggregation prone intermediates combine to form the smallest possible aggregate; A_2. This is governed by a separate rate constant; k_3. Aggregate growth by intermediate monomer addition is shown in Equation 6.12 c where single intermediates add on to the aggregate to produce a larger aggregate. Again this has its own rate constant associated with it; k_4. Once the monomer is depleted, additional aggregation can be seen to take place in the form of aggregate clustering highlighted in Equation 6.12 d. This leads to the clustered spherical morphology of aggregates seen under SEM [4] This was confirmed using ESEM to view aggregates produced from the UVLSS process (Figure 6.13, which illustrates spherical cluster-like nature of the final aggregates.
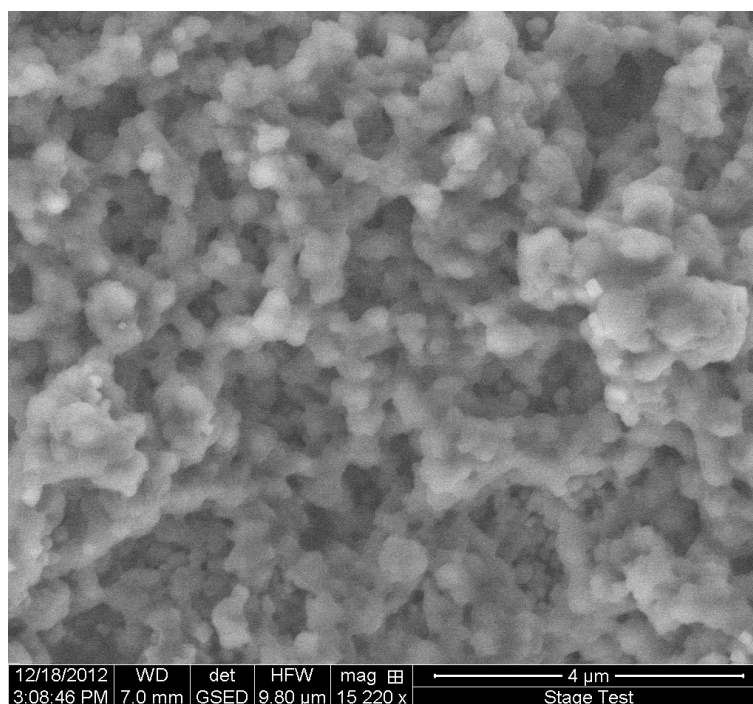
Figure 6.13: ESEM image of BSA aggregates produced at 76 ºC at 2.84Torr.

Step b is considered to be a rapid process [7] which occurs prior to step c dominating the aggregation. This formation of aggregates takes place only over one short time span, as can be seen by particulate aggregates all being of similar size [4]. Therefore this process is not occurring over the growth period, though it does contribute by producing the aggregate concentration, which after step b remains constant. Step c can be modelled using the current data ranges collected here as the rate of monomer consumption can be followed. Step d is observed within this experiment (Figure 6.4), but this data is hard to model due to the irregular shapes and sizes formed from this process. This data range is therefore omitted from the analysis. Therefore over the selected data range, there are 3 rate constants determining the overall rate of the process; k_1, k_2, and k_3.

In order to model the reaction rates occurring in this process, two rate limiting cases were considered; unfolding limited aggregation and association limited aggregation.

173

Unfolding limited aggregation would require step c to be of a much larger magnitude than step a, hence:

$$k_4 \gg k_1 \ and \ k_2$$

Using Equation 6.13 and Equation 6.12 it can be seen that the formation of any intermediate is converted into particulate aggregates, and does not return to *N*. Any monomer present within the system can be assumed to be in the native state as any intermediate will be instantly consumed. Therefore the model is reduced to a first order rate equation:

$$\frac{dC_m}{dt} = k_1 C_M$$

Equation 6.14: Rate equation for unfolding limited aggregation.

Using Equation 6.14, the rate of monomer consumption would relate to the rate of native monomer unfolding to intermediate.

Association limited aggregate would produce the reverse of Equation 6.13, where step a is more rapid than step c, therefore:

$$k_1 and \ k_2 \gg k_4$$

Equation 6.15: Rates for association limited aggregation.

Equation 6.15 illustrates that the monomer will be present as both native and intermediate states. Therefore in association limited aggregation an equilibrium exists between the native and the intermediate state:

$$k_1 C_N = k_2 C_I$$

Equation 6.16: Association limited aggregation equilibrium.

The equilibrium constant; K, can be described as a ratio between the rates of both the forward and reverse reactions (Equation 6.17).

$$K = \frac{k_1}{k_2}$$

Equation 6.17: Equilibrium constant equation.

The total monomer within this system is the sum of both the concentrations of native and intermediate states (Equation 6.18).

$$C_M = C_N + C_I$$

Equation 6.18: Calculation of total monomer.

The previous three equations (Equation 6.16, Equation 6.17, Equation 6.18) can be combined to give the concentration of monomer in terms of the concentration of intermediate, and the equilibrium constant.

$$C_M = (1 + K)\frac{C_I}{K}$$

Equation 6.19: Calculation of monomer concentration from intermediate concentration and equilibrium constant.

In the association limited aggregation kinetics the resultant model is a second order rate equation, which is dependent on the concentration of both aggregate, and intermediates (Equation 6.20).

$$\frac{dC_m}{dt} = k_4 C_I C_A$$

Equation 6.20: Second order rate equation for association limited aggregation.

However, the concentration of the intermediate cannot be easily determined as the experimental results only produce information on the monomer as a whole (i.e native and intermediate). Therefore Equation 6.19 can be substituted into Equation 6.20 in

175

order to put the second order rate equation in terms of both monomer and aggregate concentrations (Equation 6.21).

$$\frac{dC_m}{dt} = \frac{k_4 C_M C_A K}{(1 + K)}$$

Equation 6.21: Second order rate equation for association limited aggregation in terms of aggregate and monomer concentration.

By comparing Equation 6.11 and Equation 6.21 it is possible to relate the observed second order rate constant to the equilibrium constant and the association rate constant.

$$K = \frac{k_{ob2}}{(k_4 - k_{ob2})}$$

Equation 6.22: Determination of equilibrium constant from observed second order rate constant.

In Equation 6.22 it is assumed that the association rate constant k_4 is independent of temperature. This is a reasonable assumption given that the intermediate monomers are so highly reactive due to the large areas of unfavourable hydrophobic patches present on the surface of both the monomer and the aggregate [8]. This would result in the activation energy between monomer and aggregate being negligible. The relationship between the equilibrium constant and the absolute temperature; T, can be described by Equation 6.23.

$$\Delta G = -RT \ln K$$

Equation 6.23: Calculation of standard free energy change.

Where R is the gas constant and ΔG is the standard free energy change. This can then be utilised to determine the best fit value for k_4 across a data range. The initial value for k_4 was chosen as a value above the largest value of k_{ob2} so that the logarithmic term produced a real value and then iterated to find the largest R^2 value for k_4. From Figure 6.14 it can be seen that the R^2 fit value increases as k_4 increases, but the rate of increase

decreases exponentially. In order to counter this, two straight lines were fitted to the extremes of the plot, and their intercept used to determine a *k4* value at which there is minimal increase in R^2 value with increasing k_4.
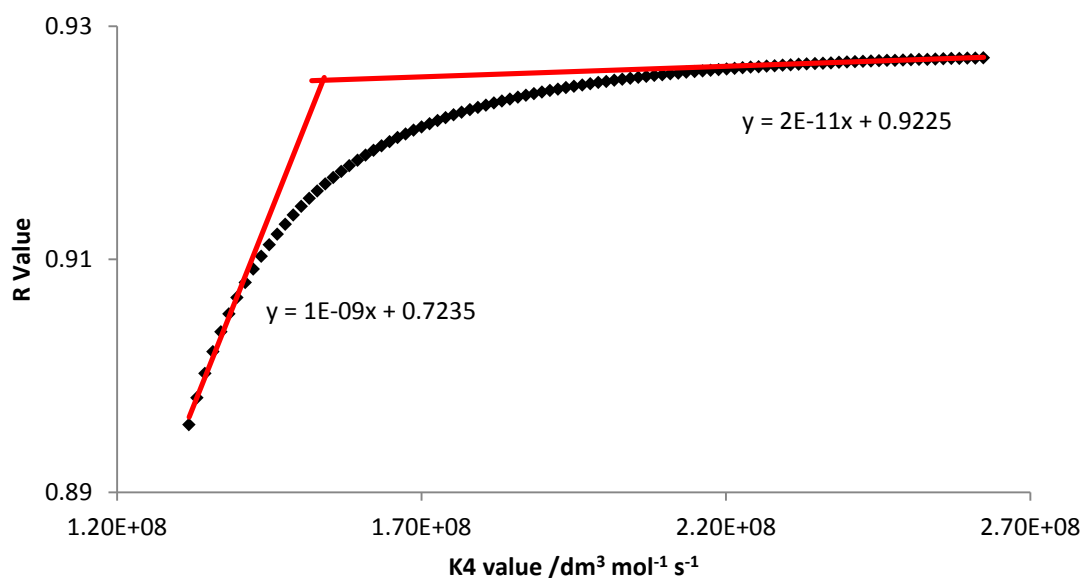
Figure 6.14: Determiniation of K4 value for lactoglobulin in 100 mM sodium phosphate, 100 mM NaCl pH 5.8.

The k_4 value obtained represents the frequency factor as the activation energy is assumed to be zero and therefore can be compared to literature values.

The temperature sensitive portion of Equation 6.22; *K* can be described in terms of Equation 6.24:

$$\ln(K) = \frac{\Delta S}{R} - \frac{\Delta H}{RT}$$

Equation 6.24: Calculation of equilibruim constant in terms of entropy and enthalpy.

Where *ΔS* is the entropy change for the process (J mol^{-1} K^{-1}) and *ΔH* is the enthalpy change for the process (J mol^{-1}). The entropy value indicates the extent of order or disorder created by the aggregating protein on itself and the surrounding water molecules. The enthalpy values indicate the energy release or uptake from bonds

formed and broken. Fitting this equation to the temperature range of data yields values for both ΔS and ΔH, which can be compared to literature values.

Finally it is possible to extract the fraction of protein present in the perturbed intermediate state; X, based on a rearrangement of Equation 6.19:

$$X = \frac{C_I}{C_M} = \frac{K}{(1 + K)}$$

Equation 6.25: Calculation of fraction of monomer in perturbed state.

Equation 6.25 allows for the fraction of perturbed monomer to be plotted at each temperature value, and would yield results that can be comparable to the literature values for melting temperatures of the protein. These values will be compared later in the chapter.

6.4 Behaviour Across Protein Systems

Krebs et al. [4] show how the formation of particulate aggregates is a generic form of aggregation that occurs across a number of protein systems. In order to confirm similarities in the kinetics of particulate formation, three proteins from the paper were studied; β-lactoglobulin, BSA and horse heart myoglobin.
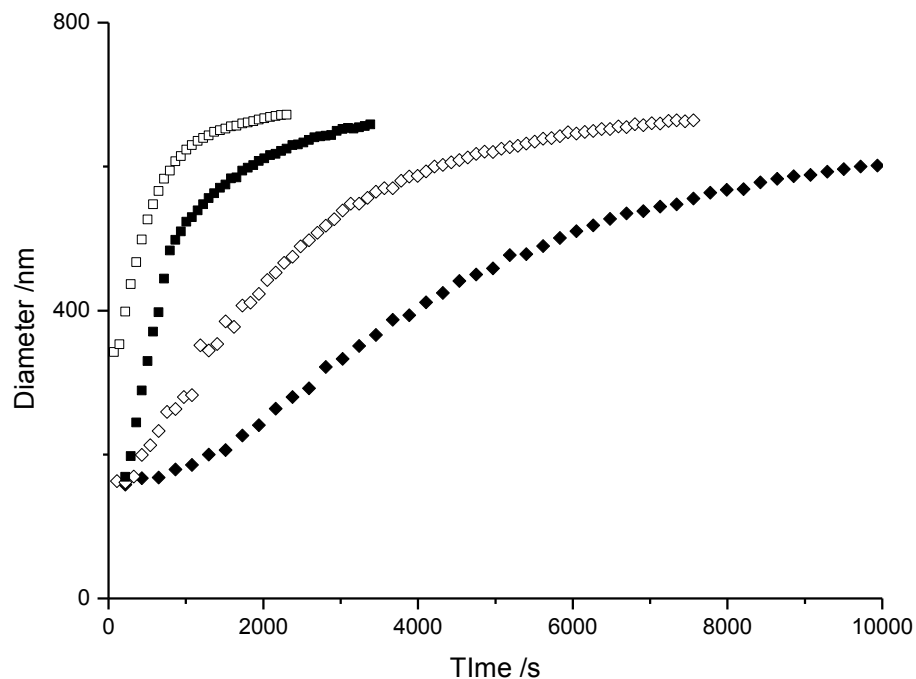
Figure 6.15: Aggregate diameter against time plot for lactoglobulin at 72.0 ℃ (open squares) 70.2 ℃ (closed squares) 68.5 ℃ (open diamonds) and 66.9 ℃ (closed diamonds). Error bars are the size of the points.

Diameter and concentration plots over time are the simplest plots that give some insight into the aggregation process occurring with these three separate proteins. Selected temperature aggregation runs for each protein are presented to illustrate the comparisons across the protein systems, and are representative of the whole data series. All data has been normalised to the time aggregation is first detected.

179
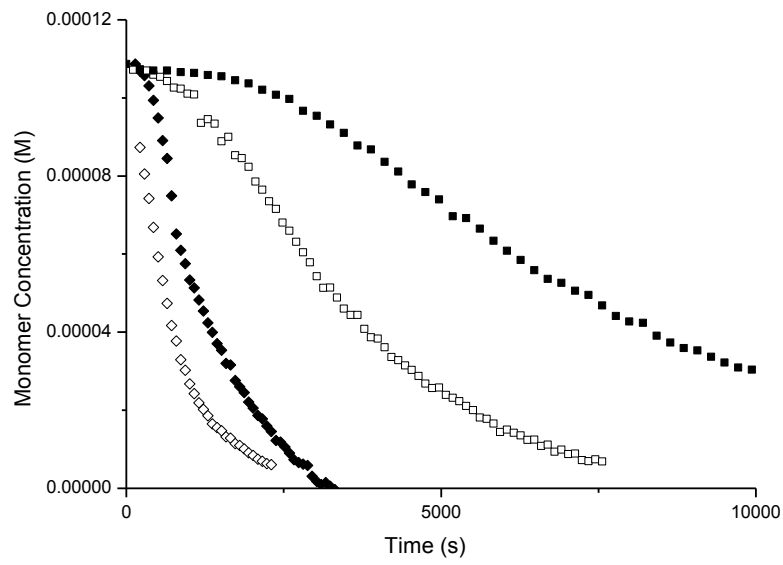
Figure 6.16: Concentration against time plot for lactoglobulin at 72.0 °C (open diamonds) 70.2 °C (closed diamonds) 68.5 °C (open squares) and 66.9 °C (closed squares). Error bars are the size of the points.
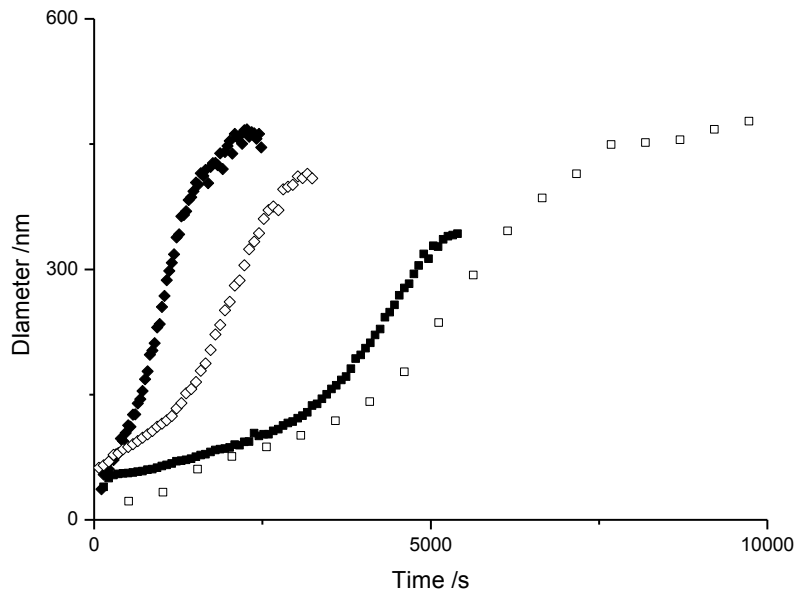
Figure 6.17: Diameter versus time plot for BSA at 72.6 °C (closed diamonds) 70.2 °C (open diamonds) 67.5 °C (closed squares) 65.5 °C (open squares). Error bars are the size of the points.
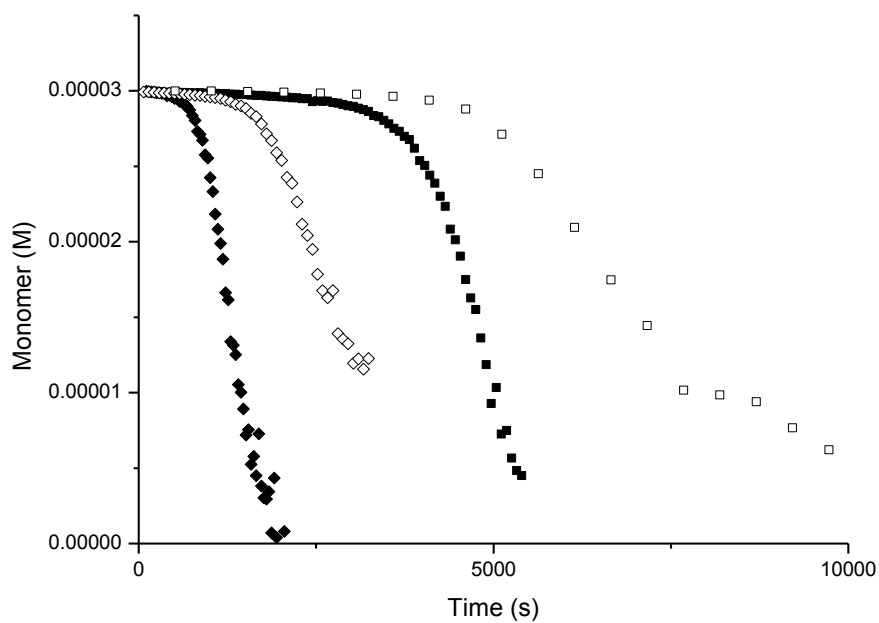
Figure 6.18: Concentration against time plot for BSA at 72.6 ℃ (closed diamonds) 70.2 ℃ (open diamonds) 67.5 ℃ (closed squares) 65.5 ℃ (open squares). Error bars are the size of the points.
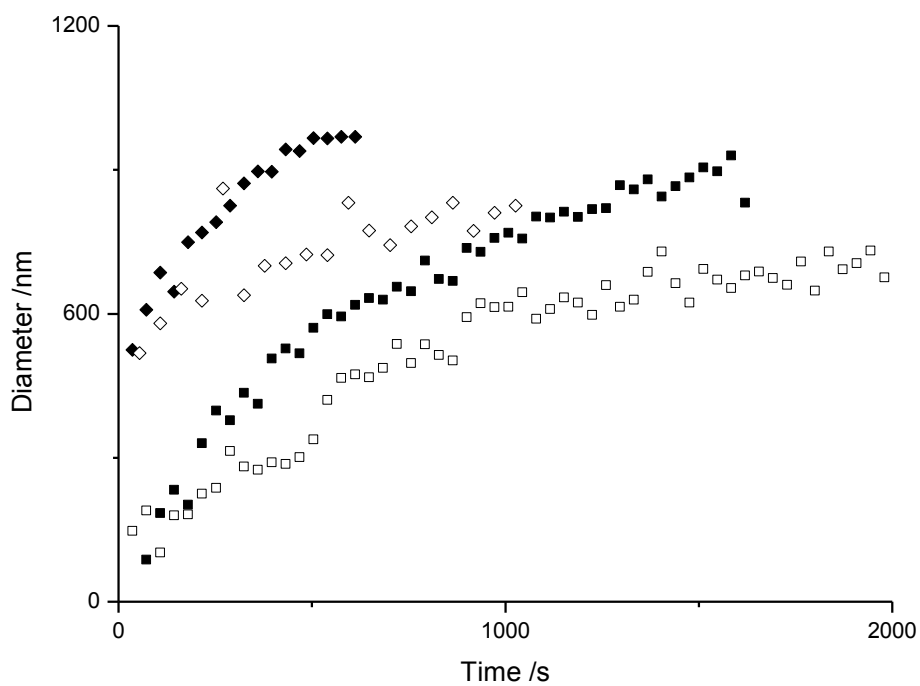
Figure 6.19: Diameter of myoglobin aggregates against time at 67.3 ℃ (closed diamonds) 64.7 ℃ (open diamonds) 63.0 ℃ (closed squares) 62.2 ℃ (open squares). Error bars are the size of the points.

181
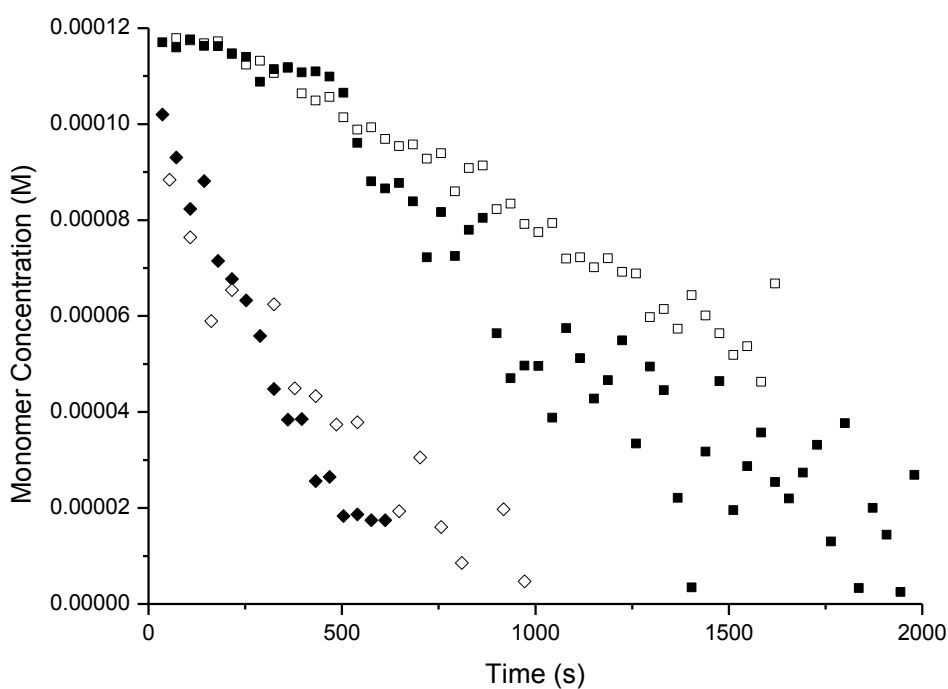
Figure 6.20: Concentration of myoglobin monomer over time at 67.3 ℃ (black) 64.7 ℃ (red) 63.0 ℃ (green) 62.2 ℃ (blue). Error bars are the size of the points.

Figure 6.15 illustrates the growth of the β lactoglobulin aggregates. The smallest aggregate size which appears to be detected is around the size of 160 nm, however at the higher temperature of 72 ℃ this value is larger, which is probably due to the rapid nature of the aggregation at this temperature. The higher temperatures exhibit characteristics of the growth phase of the particulate aggregates as the rate of growth decreases over time. At the lower temperature of 66.9 ℃ initially there is some indication that the rate of aggregate growth is increasing; at this stage the limiting factor is not the monomer concentration (as in the growth in higher temperature plots seen within Figure 6.15), but the surface area of the aggregates. As more monomer combines to the aggregate the surface area increases and during the linear section of the plot a transition occurs in the limiting factor of the process, and the exponentially decreasing growth rate is observed. This effect is more prominent in the aggregate diameter plots of

BSA (Figure 6.17), where across there is some evidence of this transition across a range of temperatures. On the other hand, myoglobin (Figure 6.19) shows only the sign of monomer concentration being the rate limiting factor in aggregation, across all temperatures.

The monomer concentration profiles of the proteins (Figure 6.16, Figure 6.18, Figure 6.20) show the rate of monomer depletion increases with increasing temperatures. Concentration plots tend to be comparable with their counterpart diameter plot in terms of change in initial rate. All plots have an exponential curve and are tending towards a zero concentration of monomer. As this data is limited to the growth phase of the aggregates and data beyond points is considered to be part of the aggregate clustering process; the data indicates that clustering only commences when there is very little remaining monomer in solution. There appears to be an exception to this in Figure 6.18 (open diamonds), but in general this appears to be generic across all the evaluated ranges.

It is worth noting that the myoglobin plots display some deviation between points (Figure 6.19 and Figure 6.20). This is due to the wavelengths required to probe the structure of the myoglobin (as discussed in the methods section). As a result of the low scatter produced at these wavelengths, there is a significant size of deviation in the data; therefore a large number of data points were collected in order to gain an accurate representation of the myoglobin system.

At each temperature, the aggregate diameter plots converge upon a final diameter size once the entirety of the monomer is consumed. These final diameters, fitted using Equation 6.4 were plotted against temperature to see if there was any correlation.
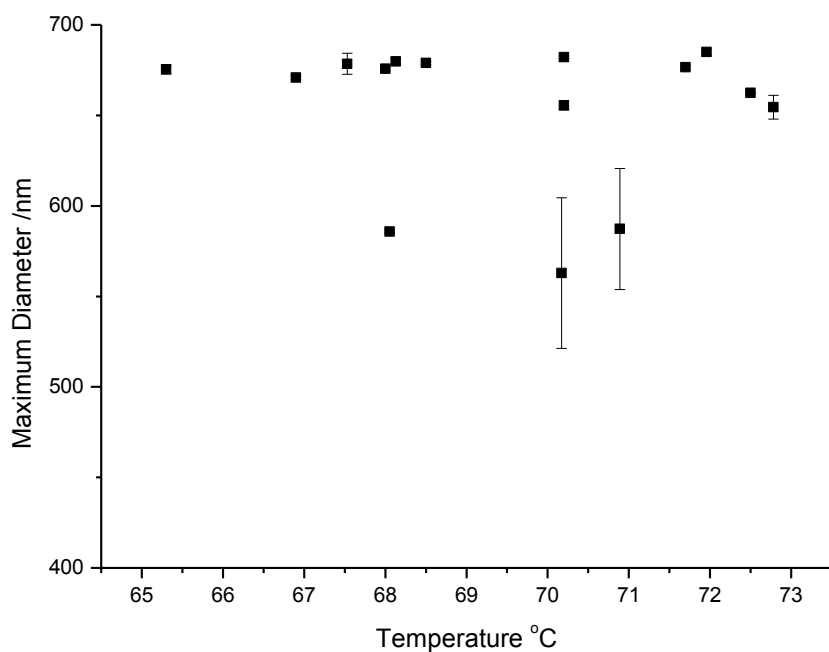
Figure 6.21: Final aggregate diameter of lactoglobulin aggregates
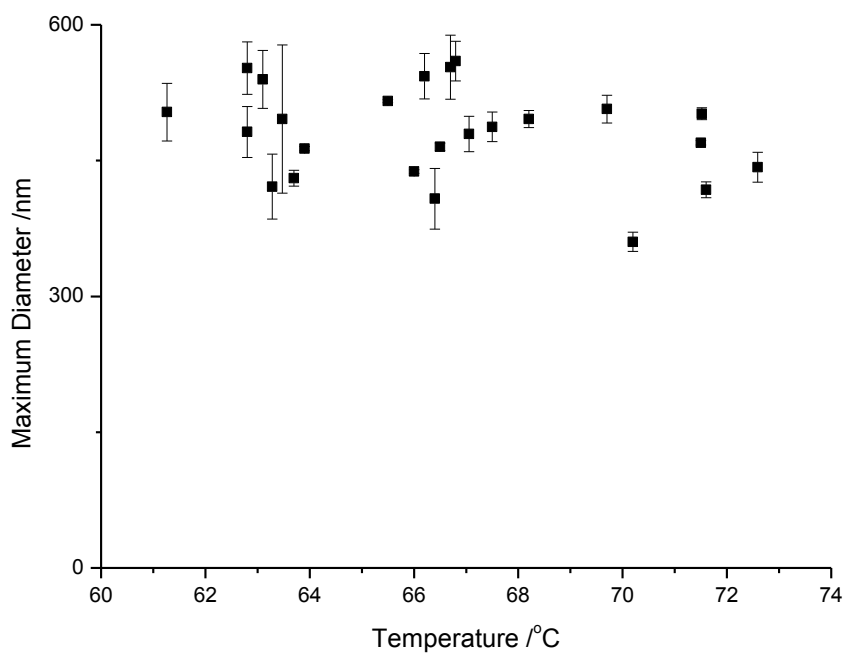
Figure 6.22: Final aggregate diameter of bovine serum albumin aggregates.

184
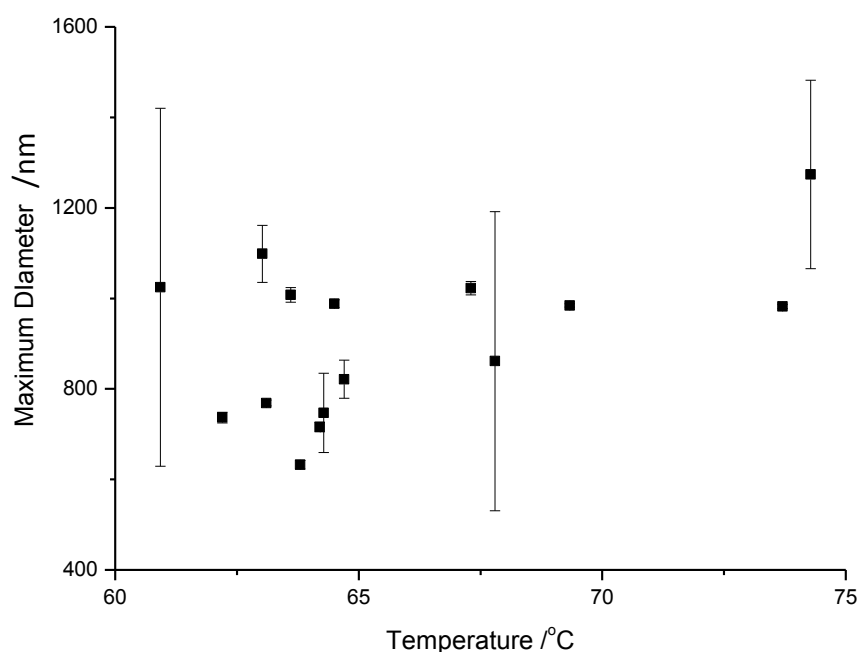
Figure 6.23: Final aggregate diameter of myoglobin aggregates.

As the final aggregate diameter plots show (Figure 6.21, Figure 6.22, Figure 6.23) there appears to be no correlation between temperature and maximum diameter of the aggregate. These results appear to disagree with the findings from Krebs *et al.* [4] which show that at higher temperatures, smaller aggregates are formed as there are more nucleation sites over which the remaining intermediate monomer is spread. The results appear to show that the nucleation process is independent of temperature, and as aggregate concentration and final diameter size are proportional; temperature has no effect on either (Figure 6.24). Considering the model for particulate aggregation (Equation 6.12), this indicates that the initial nuclei are of similar sizes which could indicate that the extent of unfolding within the protein molecules are similar across this temperature range and that further unfolding does not occur.

The maximum diameters of each protein vary; β-lactoglobulin has an average final diameter of around 670 nm, whilst BSA is 500 nm and myoglobin 900 nm. Here there

appears to be a trend between the size of the native protein monomer and its final aggregate diameter, as myoglobin is the smallest of the three, whilst BSA is the largest [6].
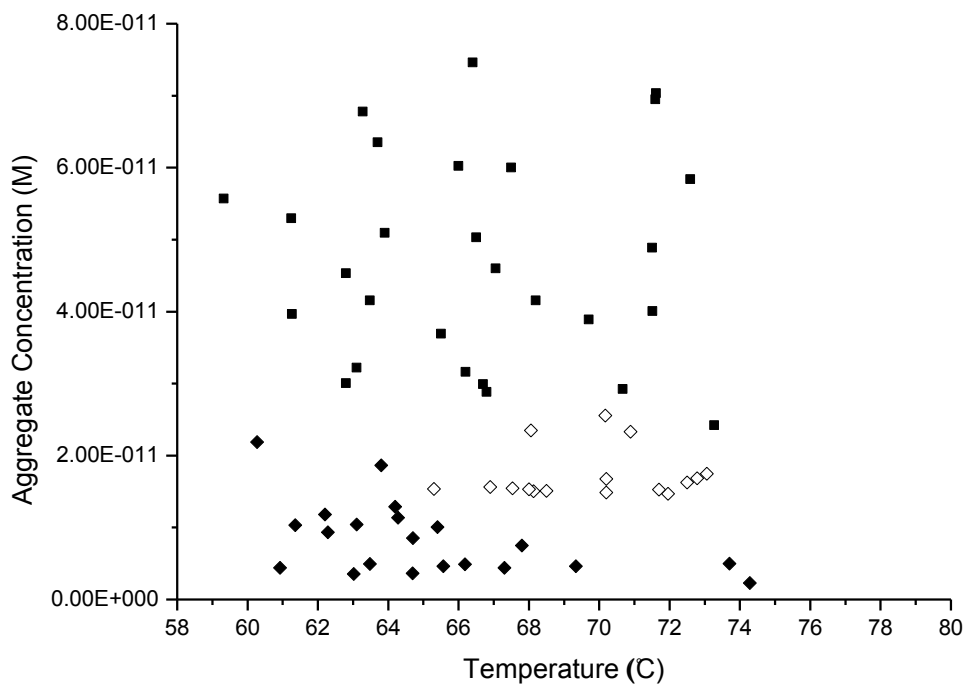
Figure 6.24: Aggregate concentration against temperature for lactoglobulin (open diamonds) BSA (full squares) and myoglobin (full diamonds). Error bars are omitted for clarity, but still indicate no trend.

In order to assess the viability of both the unfolding and association limited aggregation models, Arrhenius plots were applied to the extracted rate constants for each data range for the three proteins. The Arrhenius equation (Equation 6.26) relates the temperature dependence of the rate constant; K, by a frequency factor, or collision rate; A.

$$K = Ae^{-\frac{E_a}{RT}}$$

Equation 6.26: Arrhenius equation.

Arrhenius plots of ln$K$ against $1/T$ would yield a gradient of $-E_a/R$ and intercept ln(A), which would give the frequency factors for aggregation in the three systems.
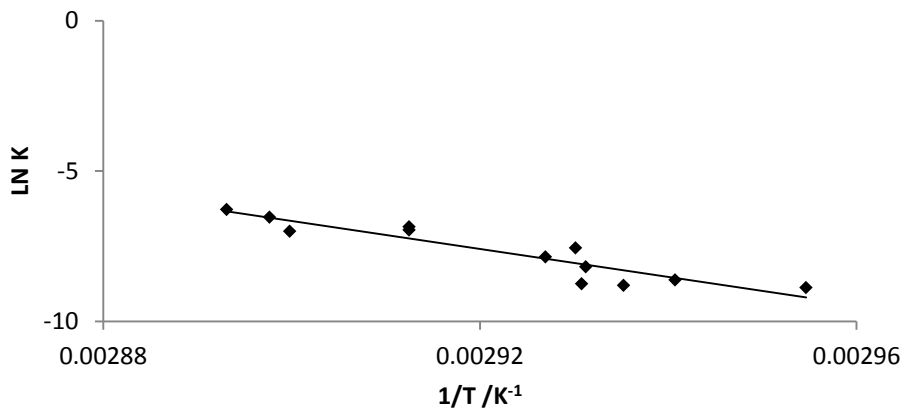
Figure 6.25: Arrhenius plot for the unfolding limited aggregation model for lactoglobulin.
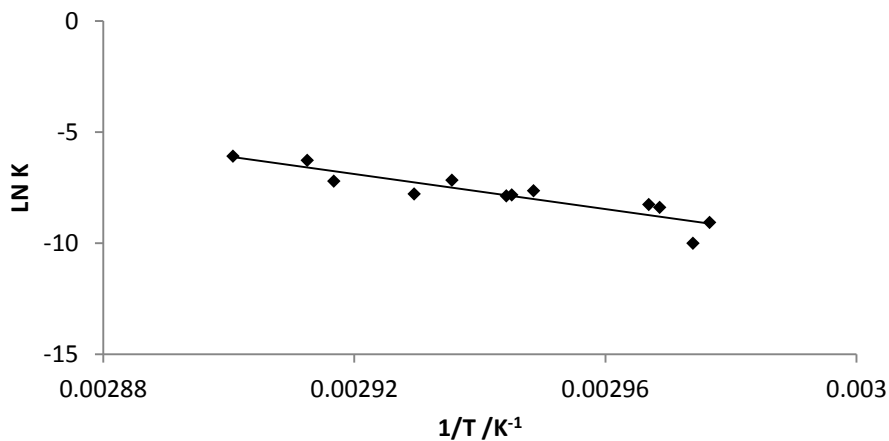
Figure 6.26: Arrhenis plot for the unfolding limited aggregation model for BSA.
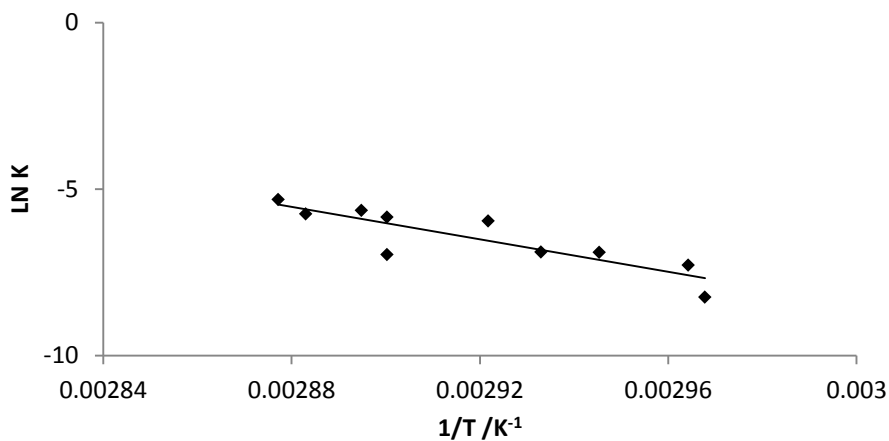
Figure 6.27: Arrhenius plot for unfolding limited aggregation model in myoglobin

The above figures (Figure 6.25, Figure 6.26, Figure 6.27) illustrate the Arrhenius fit to

the unfolding limited aggregation model. All three lines show a reasonable fit to a

187

straight line, in which K is proportional to the absolute temperature. Respective R^2 values are 0.857, 0.831, and 0.752 indicating a reasonable level of fit across the 3 proteins. However, the values produced for the frequency factor from these plots are: 4.38×10^{55} s^{-1}, 1.97×10^{47} s^{-1}, and 1.37×10^{28} s^{-1}. These values are significantly larger than any realistic values seen elsewhere in the literature, where first order small molecule reactions are at most in the region of 10^{17} s^{-1} [9].
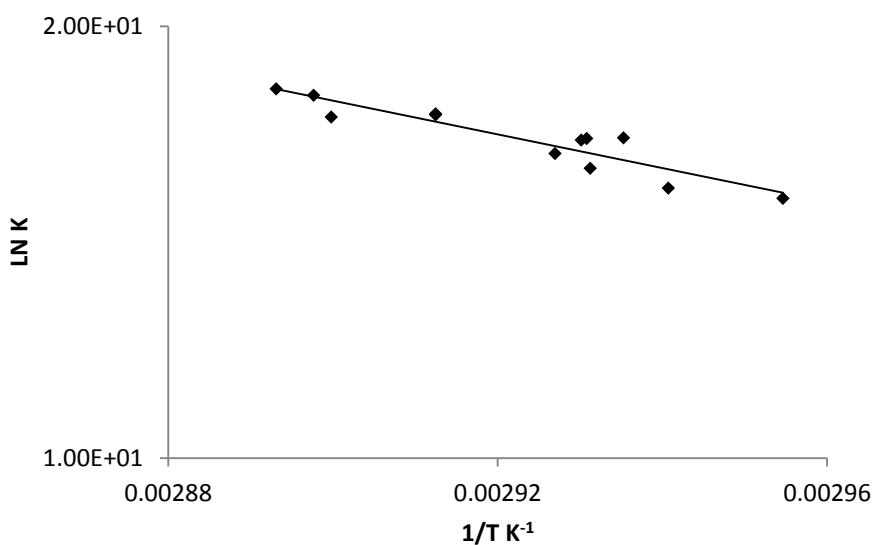
Figure 6.28: Arrhenius plot for the association limited aggregation model for lactoglobulin.
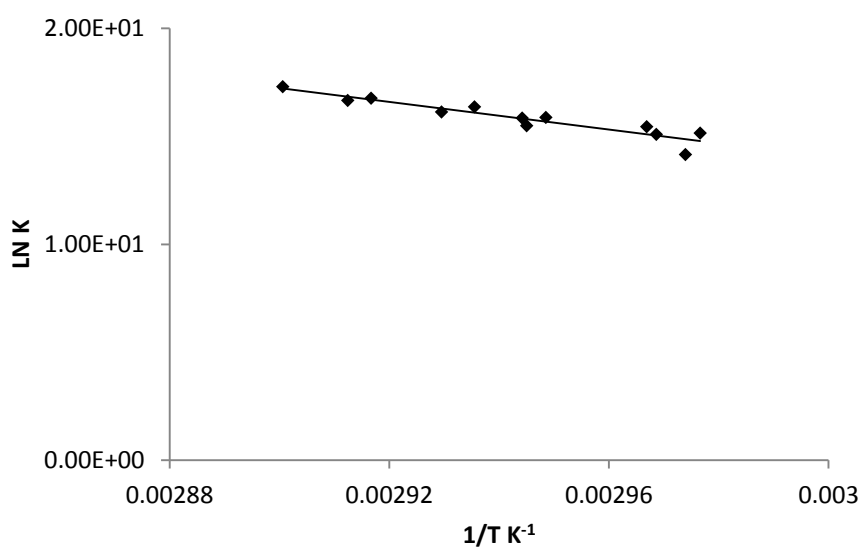
Figure 6.29: Arrhenius plot for the association limited aggregation model for BSA

188
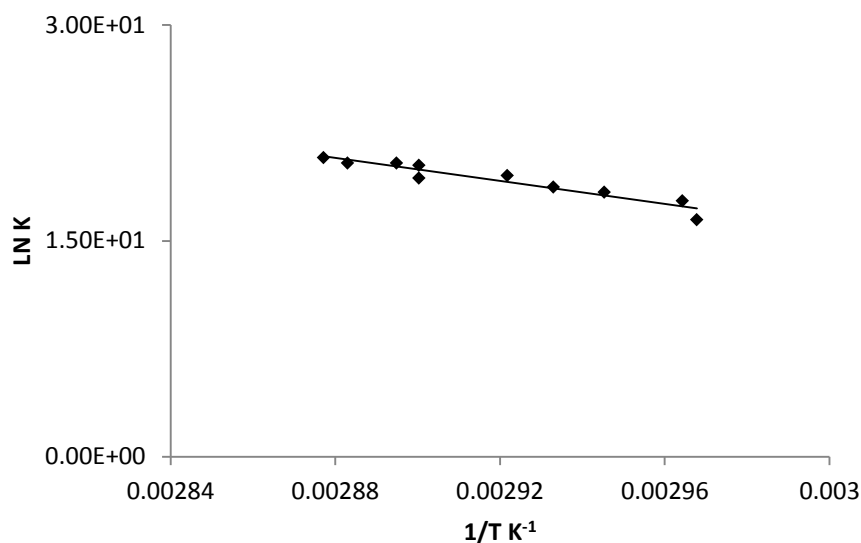
Figure 6.30: Arrhenius plot for the association limited aggregation model for myoglobin

Arrhenius plots for the association limited aggregation of the proteins are illustrated above (Figure 6.28, Figure 6.29, Figure 6.30). There appears to be a moderately increased fit to the straight line trend over the unfolding limited aggregation, with R^2 values reported at; 0.867, 0.873 and 0.9059 respectively. However the reported frequency factors of the second rate reaction are also unreasonably high: 1.31×10^{57} s^{-1}, 4.9×10^{47} s^{-1}, and 9.3×10^{58} s^{-1}. This illustrates the Arrhenius plot may not be the most valuable way of modelling this data in order to present meaningful values.

The frequency factor for association; K_4 can be obtained by varying its value to obtain the best fit for the equilibrium between native and intermediate (Figure 6.14, Equation 6.22).

Table 6.2:Reported K4 values for the three proteins.

Protein	K_4 Value /dm^{3}mol^{-1}s^{-1}
B-Lactoglobulin	7.27×10^8
BSA	5.60×10^7
Myoglobin	1.07×10^9

Table 6.2 illustrates the change in K_4 values for each protein. These values appear to give a much more sensible value for the collision rate for each protein. There also appears to be the trend that the smaller the protein, the more frequent the collisions.

The fitted value of K_4 allows the fraction of intermediate present at each temperature to be elucidated (Figure 6.31, Figure 6.32, Figure 6.33). These plots give an indication of the equilibrium shift at each temperature.
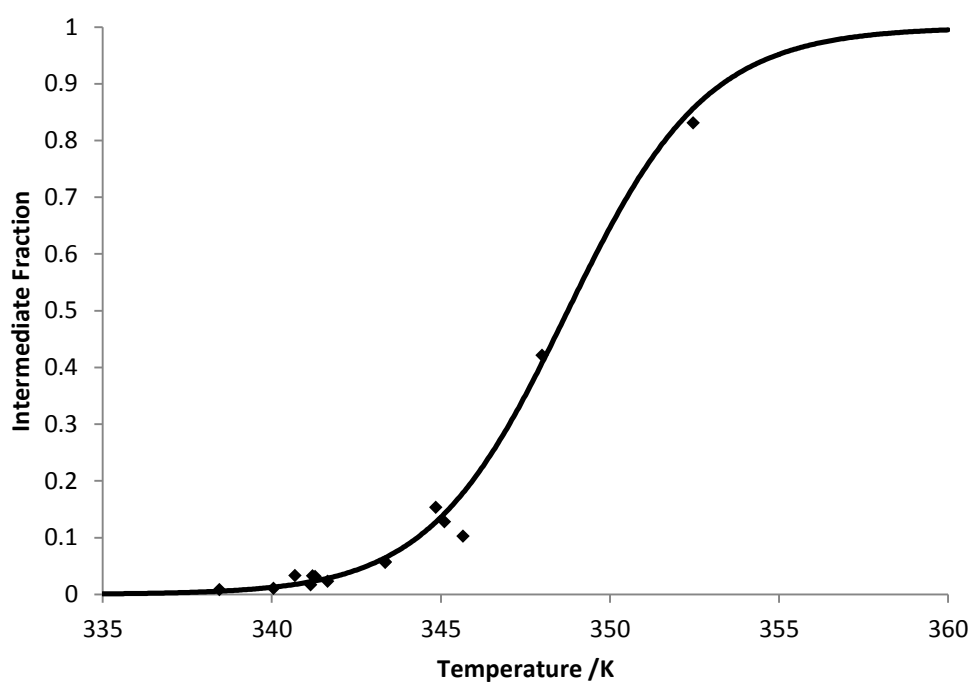
Figure 6.31: Intermediate fraction against temperature for lactoglobulin.
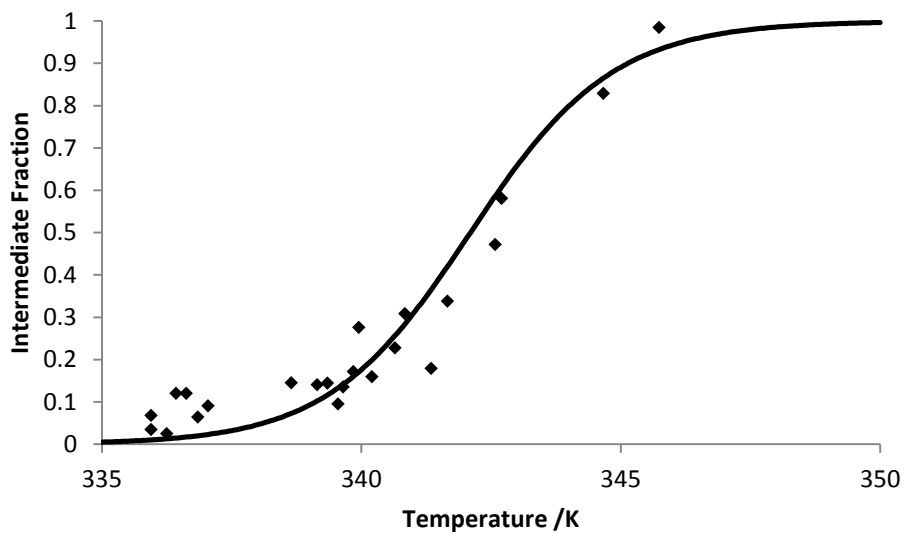
Figure 6.32: Intermediate fraction against temperature for BSA.
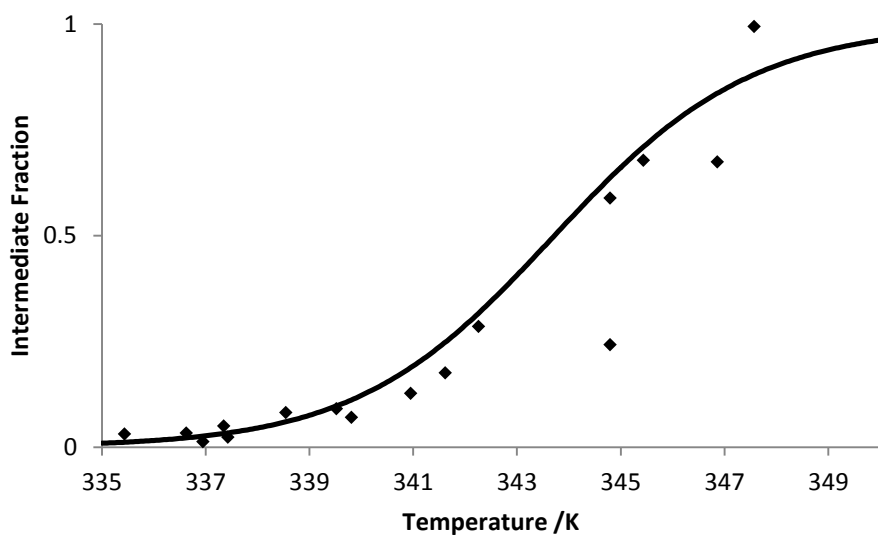
Figure 6.33: Intermediate fraction against temperature for myoglobin aggregates.

The lines indicated on these plots are lines of best fit produced from the values for ΔH and ΔS of the unfolding process. The associated values are presented in Table 6.3.

Table 6.3: Calculation of Gibbs free energy data.

	β-Lactoglobulin	BSA	Myoglobin
ΔH / kJmol^{-1}	492 ±28	709 ±55	514 ±50
ΔS / Jmol^{-1}K^{-1}	1411 ±79	2075 ±162	1496 ±145
T at $\Delta G = 0$ /K	348.8	342.1	343.7

Table 6.3 shows the extracted information on the melting temperatures and stabilities and can be compared to literature values for the unfolding of each protein.

The melting temperature of β-lactoglobulin produced from the unfolding model agrees well for quoted values of the unfolding temperature of the protein of 348.8 K [10]. Additionally the quoted value of 492 kJmol^{-1} shows some agreement with the reported value here, and the discrepancy may be explained by the difference in pH value between the two experiments; where the value reported here is obtained in more stable conditions [11].

The value presented here for the melting temperature of myoglobin shows good agreement with that seen in the literature [12]. Additionally the ΔH values show agreement with trends seen in studies of oxidated forms of myoglobin and microcalorimetric measurements performed on the sperm whale variant of the protein [13, 14].

The melting temperature for BSA of 342.1 K is marginally higher than the reported values from DSC studies [15, 16], but following the trend of increased stability at pH closer to the protein isoelectric point, this value shows good agreement. Additionally the ΔH values for unfolding bear good agreement, with reported values here being

marginally larger. This again can be attributed to the increased stability at this studied pH.

Similarly to the ΔH values the entropy values are high for proteins of a larger size (Table 6.3). These positive entropy values indicate that during the unfolding the protein is tending towards disorder. This can be explained by the interaction of both polar and non-polar residues with the water molecules, where there are more non-polar residues imposing a disorder on the surrounding molecules than polar residues creating order. This appears acceptable as the majority of hydrophobic residues are buried in the protein fold, therefore there are more to be exposed when the protein unfolds. Additionally, the trend between size and entropy change can be attributed to more hydrophobic residues in the larger proteins disordering more water molecules than that of the small proteins.

6.5 Effect of Tween on Aggregation

The UVLSS technique can be used to assess the effect of additives on the aggregation rate of proteins. Tween 20 is an additive often used within protein buffers to increase the yield of native protein produced [17]. Tween 20 has also been shown to reduce the aggregation, however the mechanism for this is not clear [18, 19]. In these instances 0.01 % Tween 20 w/v was included in the aggregation buffer and comparisons made to elucidate the effect and mechanism of Tween 20 on particulate aggregation.
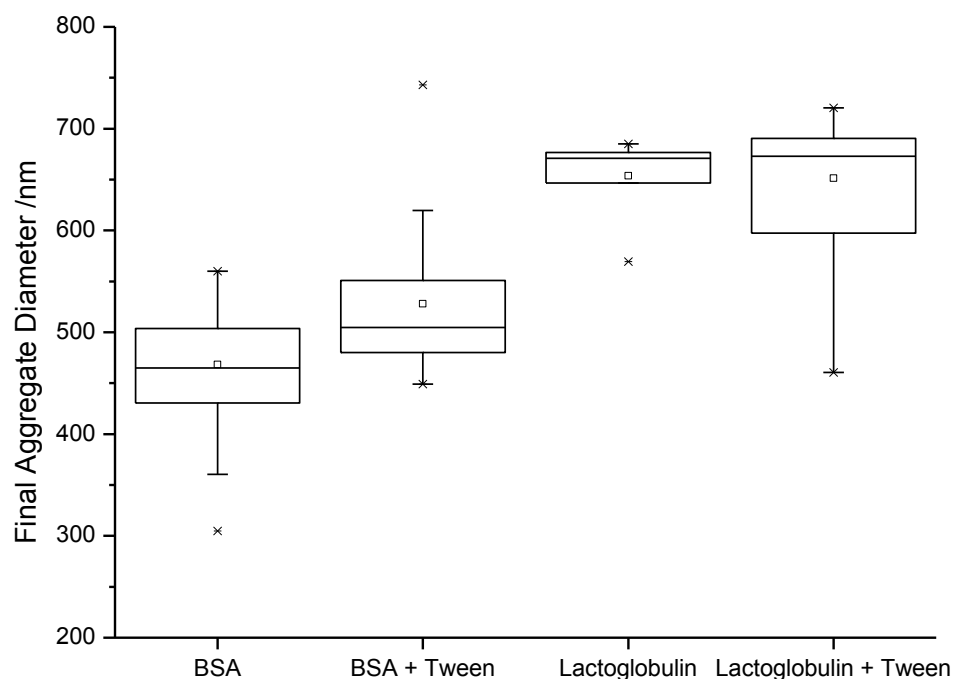
Figure 6.34: Comparison of final diameters of aggregates in BSA and lactoglobulin both with and without Tween 20.

Figure 6.34 compares the range of diameters produced in two protein systems with and without Tween. As has been previously shown, the final size of the aggregate does not appear to be determined by the temperature, therefore the diameters and their distributions can be represented by box plots. Figure 6.34 illustrates that both BSA and lactoglobulin aggregate sizes are not affected by the presence of Tween 20. There appears to be a slightly larger distribution of sizes across the Tween data sets, but this distribution difference is mainly insignificant. It can be inferred from these plots that the interaction of Tween with these aggregates introduces no new levels or porosity into the aggregates, or indeed change the concentration of aggregates formed significantly. Therefore it seems Tween has little relevance in the nucleation stages of the aggregate formation.

Table 6.4: K4 Comparison between protein without additives and protein with Tween 20.

Protein	Native K_4/ s^{-1}	K4 +0.01 % Tween / s^{-1}
B-Lactoglobulin	7.27×10^8	1.24×10^8
Bovine Serum Albumin	5.6×10^8	2.95×10^8

As can be seen in Table 6.4, the effect of Tween has some impact on the collision rate constant K_4. Although this may be within error, this could be accounted for due to the interference from the Tween in some collisions, which may reduce the rate slightly. Tween has been shown to bind hydrophobic regions within the molten globule of a protein [18]. This indicates that the Tween 20-protein interactions compete with the protein-protein interactions to reduce the collision rate, as is shown by the decreased K_4 values in Table 6.4.
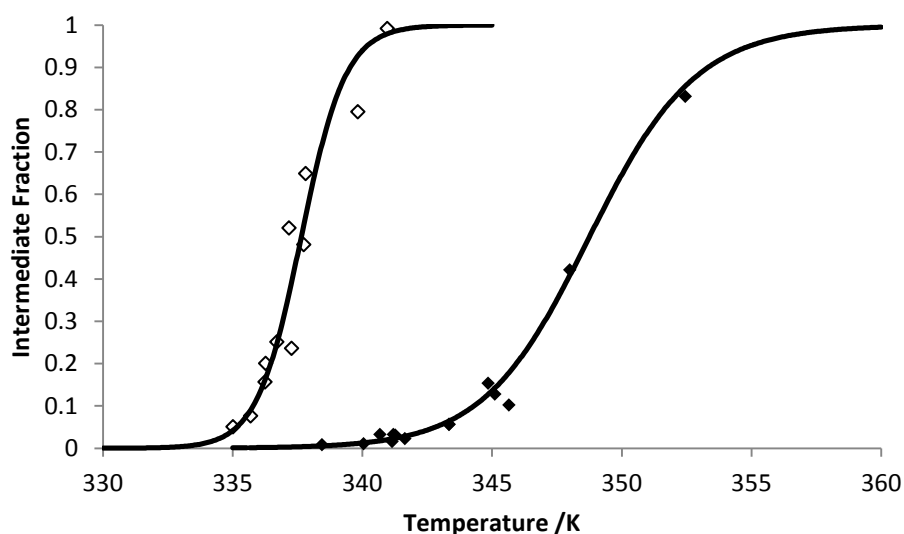
Figure 6.35: Intermediate fraction against temperature of B-lactoglobulin without additive (closed diamonds) and with 0.01 %Tween (open diamonds)

Figure 6.35 illustrates the effect that Tween has on the equilibrium between native and aggregation prone intermediate. It is clear that β-Lactoglobulin with Tween (open diamonds) causes a shift in the temperatures at which the intermediate is formed. The

shapes of the transition curves indicate that the shift in initial formation is only of the size of around 3 degrees. However for full transition of the entire sample from native to unfolded is of the order of 10 degrees. This illustrates that whilst the protein is natively folded, Tween 20 plays a small role in enhancing the unfolding of β-lactoglobulin, however as the transition occurs Tween appears to further enhance the unfolding of the protein. This increase in effect correlates with an increasing amount of hydrophobic structure available on the protein through which Tween may facilitate the unfolding [18].
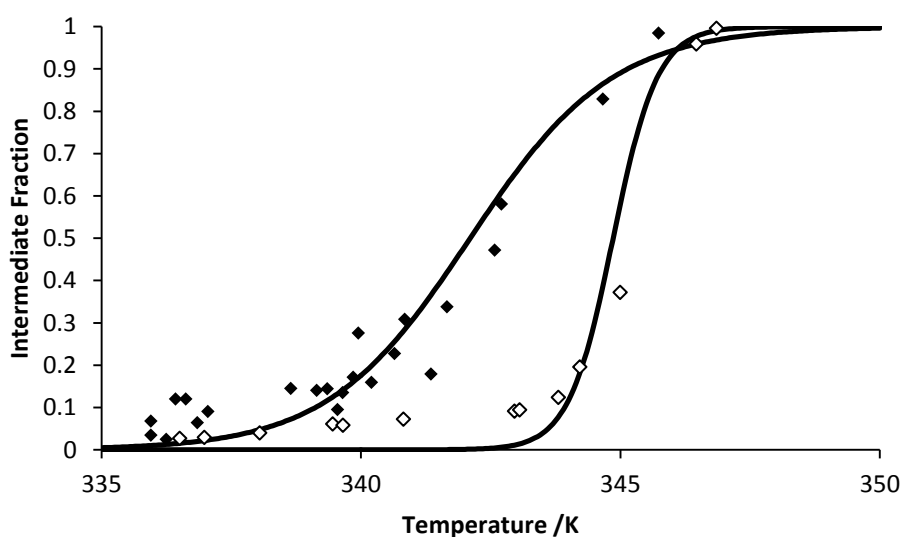
Figure 6.36: Intermediate fraction against temperature of BSA without additive (closed diamonds) and with 0.01 %Tween (open diamonds)

Figure 6.36 compares the effect of 0.01 % Tween on BSA. Despite the initial formation of the intermediate forming at the same temperature ~336 K, there appears to be little increase in this across the temperature scale until ~345 K, where the full sample of protein is present as intermediate fraction, which is at a similar temperature as that of the protein without additive.

As in Figure 6.35, the exposure of hydrophobic regions within the protein does appear to facilitate the binding of Tween to stimulate the unfolding, however it can be seen that BSA in Tween is infact more stable at the lower temperatures as illustrated by the curve.

$$\Delta G = \Delta H - T\Delta S$$

Equation 6.27:Calculation of Gibbs free energy

Using Equation 6.27 the Gibbs free energy of the protein at room temperature can be calculated. This illustrates the stability of the proteins under conditions with and without Tween 20. Table 6.5 illustrates how at 25 ℃ proteins in 0.01 % Tween clearly have a higher stability that the proteins without. The increase is seen as 2 to 3 fold larger in the ΔG values, and indicates why Tween 20 is a competent stabilising additive of the native state.

Table 6.5: Comparative ΔG values of proteins with and without Tween at 25 ℃.

Protein	Native ΔG at 25 ℃ /kJ	ΔG in 0.01 % Tween /kJ
BSA	91.5 ±7.14	348.3 ±32.7
Lactoglobulin	78.7 ±4.4	147.2 ±14.2

6.6 The Effect of Sodium Chloride Concentration

Sodium chloride is a common additive in protein stabilising solutions [20]. Using the UVLSS technique it is possible to investigate the effect that varying concentrations of sodium chloride may have on the protein aggregation process. For this study β-lactoglobulin was chosen as it is a well characterised system in protein aggregation, and from previous work within this chapter it can be seen that the formation of β-lactoglobulin aggregates can be coherently studied by this technique.
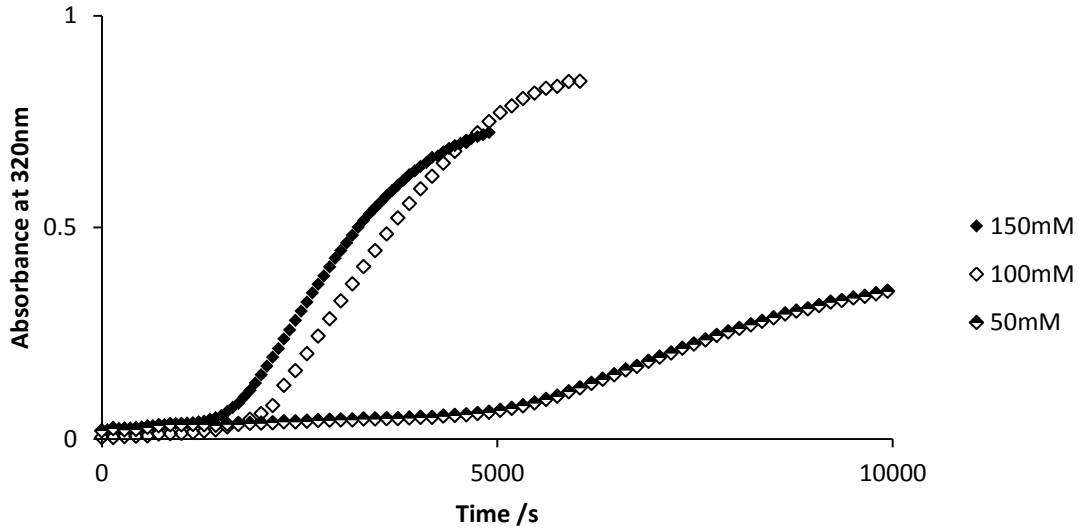
Figure 6.37: Absorbance at 320 nm of β-lactoglobulin heated to 70.8ºC at pH 5.8 sodium phosphate buffer including 150,100 mM and 50 mM sodium chloride concentrations.

Figure 6.37 offers a primitive method for examining the effect of sodium chloride concentration on the aggregation of β-lactoglobulin. All three samples were recorded to have an average temperature of 70.8 ºC during the isothermal stage. At the highest of salt concentrations, the absorbance can be seen to increase at an earlier time than that of a lower salt concentration. This indicates that aggregation is occurring earlier in the more concentrated salt environment, which may suggest the high salt concentration contributes to a destabilising effect to the native protein conformations.

Figure 6.38 illustrates the range of final aggregate diameters produced by each experiment at the three salt concentrations. It can be seen that there is no substantial deviation in the final diameter of aggregates across the concentration range, similar to the results seen for the additive Tween (Figure 6.34). The sole trend within the box plots of Figure 6.38 is that the range increases in line with the salt concentration. This may be attributed to the speed at which the experiments take place at the higher salt concentration, therefore the degree of error in determining the final concentration may be larger.
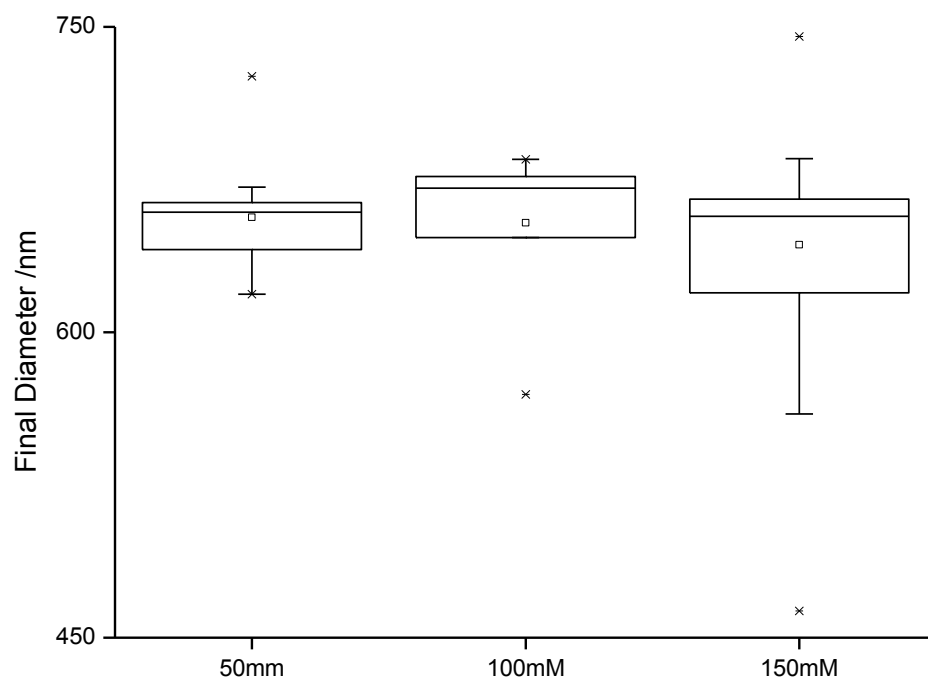
Figure 6.38: Comparison of final aggregate diameters for b-lactoglobulin aggregates in 50 mM, 100 mM and 150 mM salt conditions.
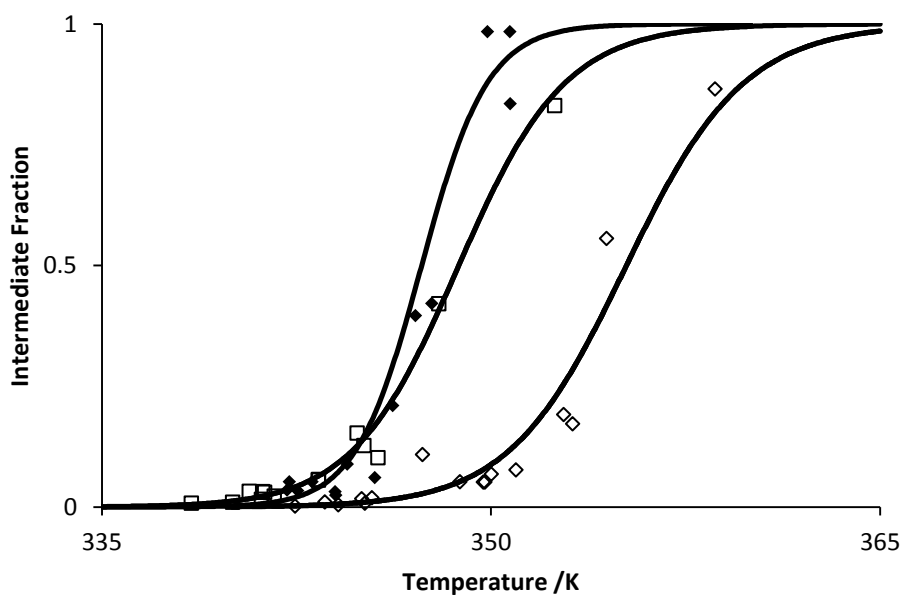
Figure 6.39: Intermediate fraction versus temperature for β-lactoglobulin at 50 mM (open diamonds), 100 mM (open squares) and 150 mM (closed squares) NaCl concentrations. Lines of best fit are applied to each.

Figure 6.39 shows the effect the concentration of NaCl has on the aggregation behaviour of β-lactoglobulin. Clearly higher concentrations have a destabilising effect on the native monomer, as the transition appears to be more rapid and at a lower temperature.

Table 6.6: Values for lactoglobulin aggregation thermodynamics values.

	Lactoglobulin with NaCl concentration		
	50 mM	**100 mM**	**150 mM**
ΔH /kJ	462 ±49	492 ±28	788 ±72
ΔS / kJK^{-1}	1.3 ±0.14	1.4 ± 0.14	2.3 ±0.21
Temperature at ΔG=0 /K	355.2	348.8	347.3
ΔG at 25 ºC	81 ±8.6	78.7 ±4.4	123 ±11.3

Both values for $\mathit{\Delta H}$ and $\mathit{\Delta G}$ at 25 ºC indicate that salt plays a stabilising role in the protein structure (Table 6.6), where at higher salt concentrations, a much larger amount of energy is required to unfold the protein than at lower salt concentrations. However, this further results in the protein at higher salt concentration, resulting in an 8 ºC change in unfolding temperature over the 100 mM NaCl range. This appears to show a similar trend to that of the Tween 20 experiments (Table 6.5). The effect of salt on aggregation is also examined in the following section, therefore the overall effect of salt on aggregation will be discussed in the concluding chapter. However it is worth noting that an increase in salt concentration from 50 mM to 150 mM decreases the K_4 value from 2.9×10^9 s^{-1} to 4.4×10^8 s^{-1} and therefore clearly plays a role in decreasing the number of successful collisions during the aggregation process.

6.7 References

1.   Smeller, L., P. Rubens, and K. Heremans, *Pressure Effect on the Temperature-Induced Unfolding and Tendency To Aggregate of Myoglobin†.* Biochemistry, 1999. **38**(12): p. 3816-3820.

2.   Meersman, F., L. Smeller, and K. Heremans, *Comparative Fourier transform infrared spectroscopy study of cold-, pressure-, and heat-induced unfolding and aggregation of myoglobin.* Biophysical Journal, 2002. **82**(5): p. 2635-2644.

3.   Murphy, R.M. and C.C. Lee, *Laser Light Scattering as an indispensible tool for probing protein aggregation*, in *Misbehaving Proteins - Springer ebook.* 2006.

4.   Krebs, M.R.H., G.L. Devlin, and A.M. Donald, *Protein particulates: another generic form of aggregation?* Biophysical Journal, 2007. **92**: p. 1336-1342.

5.   Willard, L., et al., *VADAR: a web server for quantitative evaluation of protein structure quality.* Nucleic Acids Research, 2003. **31**(13): p. 3316-3319.

6.   Berman, H., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.

7.   Roberts, C., *Non-native protein aggregation kinetics.* Biotechnology and bioengineering, 2007. **98**(5): p. 927-938.

8.   Chi, E.Y., et al., *Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation.* Pharmaceutical Research, 2003. **20**(9): p. 1325-1336.

9.   Atkins, P., *Physical Chemistry, 6th.* 1998, Oxford University Press: Oxford.

10.   De Wit, J.N. and G.A.M. Swinkels, *A differential scanning calorimetric study of the thermal denaturation of bovine β-lactoglobulin Thermal behaviour at temperatures up to 100°C.* Biochimica et Biophysica Acta (BBA) - Protein Structure, 1980. **624**(1): p. 40-50.

11.   Gotham, S., P. Fryer, and A. Pritchard, *β-lactoglobulin denaturation and aggregation reactions and fouling deposit formation: a DSC study.* International Journal of Food Science & Technology, 1992. **27**(3): p. 313-327.

12.   Yan, Y.-B., et al., *Two-dimensional infrared correlation spectroscopy study of sequential events in the heat-induced unfolding and aggregation process of myoglobin.* Biophysical Journal, 2003. **85**(3): p. 1959-1967.

13.   Khechinashvili, N., J. Janin, and F. Rodier, *Thermodynamics of the temperature-induced unfolding of globular proteins.* Protein Science, 1995. **4**(7): p. 1315-1324.

14.   Kelly, L. and L. Holladay, *A comparative study of the unfolding thermodynamics of vertebrate metmyoglobins.* Biochemistry, 1990. **29**(21): p. 5062-5069.

15.   Michnik, A., *Thermal stability of bovine serum albumin DSC study.* Journal of Thermal Analysis and Calorimetry, 2003. **71**(2): p. 509-519.

16.   Michnik, A., et al., *Comparative DSC study of human and bovine serum albumin.* Journal of Thermal Analysis and Calorimetry, 2006. **84**(1): p. 113-117.

17.   Bondos, S. and A. Bicknell, *Detection and prevention of protein aggregation before, during, and after purification.* Analytical Biochemistry, 2003. **316**(2): p. 223-231.

18.   Bam, N., et al., *Tween protects recombinant human growth hormone against agitation-induced damage via hydrophobic interactions.* J. Pharm. Sci., 1998. **87**(12): p. 1554-1559.

19.   Yu, Z., K. Johnston, and R. Williams, *Spray freezing into liquid versus spray-freeze drying: Influence of atomization on protein aggregation and biological activity.* European Journal of Pharmaceutical Sciences, 2006. **27**(1): p. 9-18.

20.   Middelberg, A., *Preparative protein refolding.* Trends in biotechnology, 2002. **20**(10): p. 437-443.

Chapter 7 : **Aggregate Formation in Refolding**

7.1 Introduction

The presence of aggregates during protein refolding is a major issue. Aggregate formation results in lower yield of native proteins produced in the refolding process [1], and when present in therapeutic use, the presence of aggregates is capable of producing immunogenicity and loss of therapeutic function of the protein [2].

Whilst many items in the literature focus on monitoring the yield of native protein produced [3, 4], few analyse the formation of aggregates [5]. Work presented here aims to relate the presence of the molten globule state during refolding to the formation of aggregates of the small alpha helical protein cytochrome C. Furthermore, the effect of both sodium chloride concentration, and final denaturant concentration are investigated. This is performed using stopped flow anisotropy, absorbance, and the use of the fluorescent probe; ANS.

7.2 Experimental Overview

A stopped flow cell, capable of measuring absorbance, fluorescence and fluorescence anisotropy was used to measure the refolding of protein that occurred when solutions from two syringes were mixed. One syringe contained unfolded protein in a high concentration of denaturant, the other contained buffer. Upon mixing the denaturant is diluted and the protein refolds. The refolding conditions are altered either by changing the sizes of the syringes that feed the mixing chamber, therefore determining the final urea concentration, or changing the salt concentration present in the diluting buffer. The aforementioned experimental techniques are then used to probe the behaviour of the protein whilst folding over a 100 second time period.

7.3 Folding Characteristics
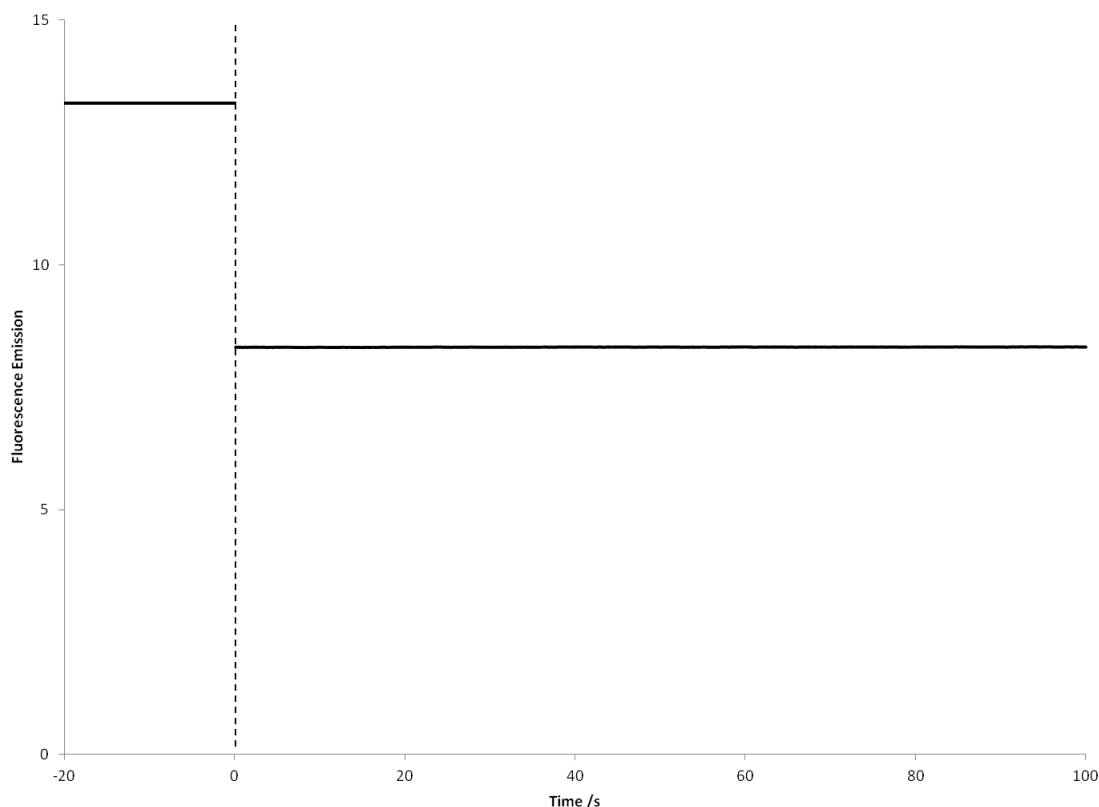
Figure 7.1: The fluorescence emission of ANS present with cytochrome C and 8 M urea during a 1:10 dilution with 100 mM TRIS 0 mM NaCl pH 10.6 refold buffer. The dashed line indicates the initiation of mixing at 0 s. Before 0 s indicates the fluorescence of the ANS sample.

Figure 7.1 illustrates the refolding of Cytochrome C in terms of hydrophobic exposure. ANS is present in one of the syringes with the Cytochrome C, maintained in the unfolded state with 8 M urea. The ANS molecules may bind to the exposed hydrophobic regions present on the protein and produce an increased fluorescence (over the unbound ANS molecules) when excited at 370 nm [6]. The fluorescence value of this sample alone is indicated by the values for time zero in Figure 7.1. However, upon initiating the stopped flow reaction, the fluorescence signal is instantly reduced to a lower fluorescence level.

This finding indicates that the hydrophobic collapse of the protein occurs within the dead time of the instrument, where the presence of a much diluted concentration of urea

204

no longer presents a barrier to folding; the energetically favourable elimination of water and ANS from these regions is rapid. This appears consistent with current understanding where even large multidomain proteins accomplish hydrophobic collapse in the nanosecond timescale [7], and proteins such as cytochrome C begin structural formation of elements of secondary structure even within the deadtime of this experiment [8, 9]. The small fluctuations seen in the ANS fluorescence in Figure 7.1 is assigned to low level aggregation occurring, and will be discussed in detail later in this chapter.
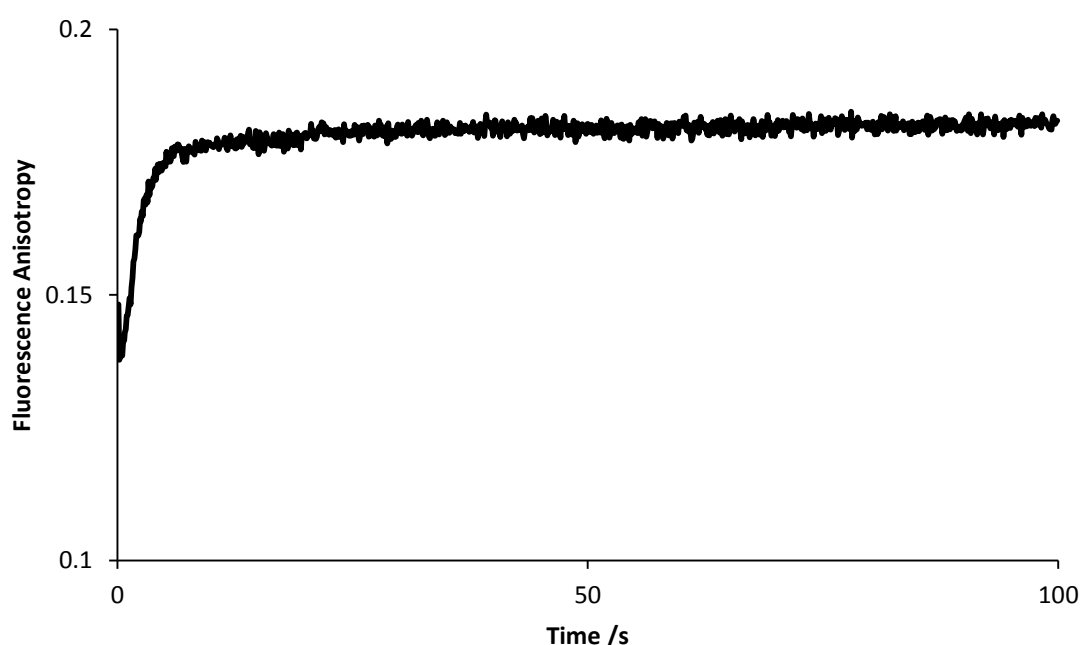
Figure 7.2: Fluorescence Anisotropy over time for Cytochrome C in 8 M Urea refolded at 1:25 ratio at 280 nm.

Figure 7.2 illustrates the corresponding fluorescence anisotropy values collected for the folding experiments in Figure 7.1. Fluorescence anisotropy corresponds to the mobility of the fluorophores present within the protein [10], namely; phenylalanine, tyrosine and tryptophan. At 280 nm the majority of the fluorescence will appear from the tyrosine and tryptophan residues, with tryptophan fluorescence dominating the signal [11].

In contrast to Figure 7.1, Figure 7.2 shows some change in the anisotropy signal during the folding process, indicating that there is some detectable structural change occurring throughout this time period. The change in anisotropy value from a lower to higher value is representative of the fluorophores within the protein becoming less mobile. During the first 10s this change in mobility is drastic, with an additional 20s of small change in mobility before the mobility of the fluorophores remains constant throughout the remainder of the experiment.

From Figure 7.1 and Figure 7.2 it can be seen that the folding of Cytochrome C is not an instant process, and that although the hydrophobic collapse of the protein is clearly in the sub microsecond time span (which is a widely accepted result during protein folding [12] [13]), the structural rearrangement of at least the fluorescent residues occurs over a much greater time span. This may be indicative of the molten globule state of the protein; where the hydrophobic core of the protein structure is formed [14], but there is still some mobility in some outer residues as they adopt their energetically favourable conformations. This mobility may be representative of the unquenched Trp 59 residue which has been suggested to remain loosely packed until the late stages of cytochrome C folding [15].

This residue mobility measured by fluorescence anisotropy during refolding is not unique to cytochrome C, and is evident across a number of protein systems.
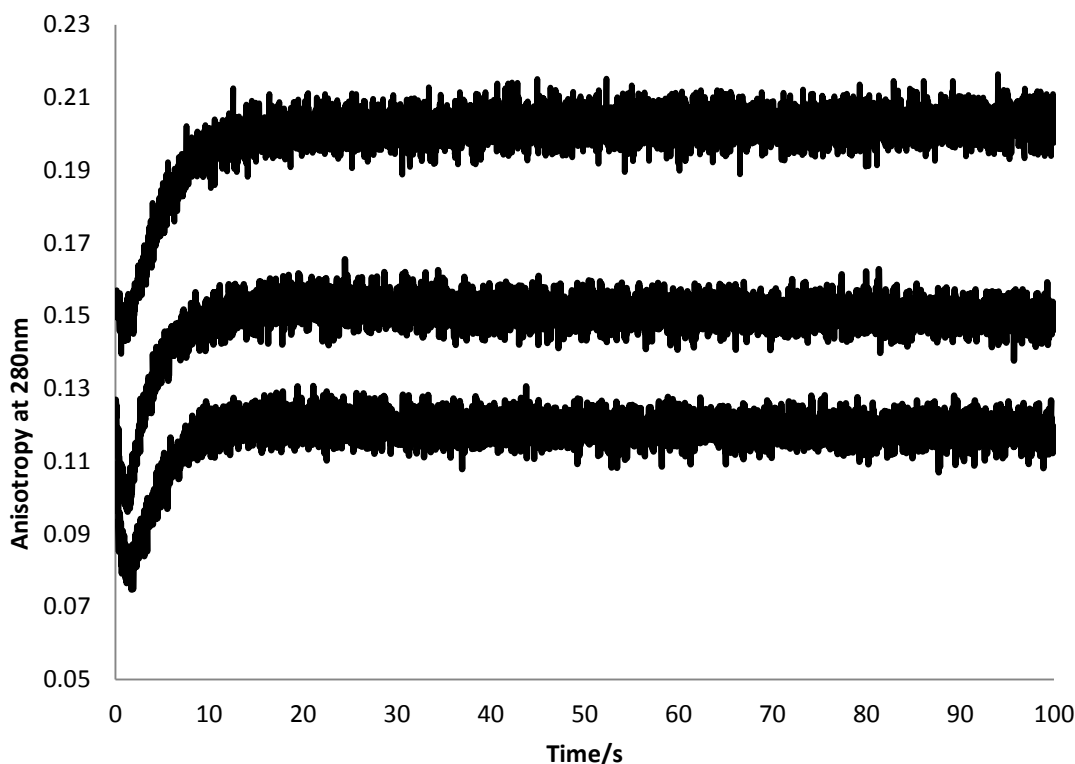
Figure 7.3: Anisotropy of refolding of proteins in 8M urea at a ratio of 1:10 at 280nm. From top to bottom :Ribonuclease A(100mM sodium phosphate buffer at pH 7.8),β-lactoglobulin(100mM TRIS at pH 7.8), α-lactalbumin(100mM TRIS at pH 7.8)

Figure 7.3 shows the anisotropy for refolding of different protein systems at 280 nm. All three figures show some agreement with the findings shown in Figure 7.2; that fluorophore mobility is present over a timescale of tens of seconds. These refolding timescales appear to vary somewhat across the systems, however the general form of the anisotropy plot is the same across proteins. This suggests that this is a ubiquitous stage in protein refolding. It may be worth noting that in both β-lactoglobulin and α-lactalbumin folding, the proteins achieve a slightly over rigid formation, indicated by a minor crest in the anisotropy plots at 20s into the duration of the experiments before relaxing into the stable structure.

It may be further noted that these comparative proteins differ from cytochrome C in that they contain sulphide bridges, which will not be disrupted by the high concentrations of denaturants within the solution, therefore the starting state of the protein before mixing cannot be considered as totally unfolded protein. The presence of disulphide bridges should provide some template for folding and increase the chance of the protein adopting its native structure. This should have minimal bearing on the molten globule state of the proteins however, aside from curtailing the time period before it appears.

Ribonuclease A (Figure 7.3) has been shown capable of undergoing both rapid and slow refolding to produce the native state [16]. Under these experimental conditions, it can be clearly seen that fluorophore mobility does not become stable until ~30s which may relate to the mobility of two tyrosine residues measured in the refolding of ribonuclease A by real time NMR [17].

β-lactoglobulin forms a compact globular structure which is marginally (10 %) larger than the native state within the first 100ms of refolding [18]. The data in **Error! Reference source not found.** suggests that under these conditions, the final structural compaction occurs over a much larger timescale (~30s). This appears concurrent with findings from elsewhere [19], that suggest that once the secondary structure has formed within the rapid time span, the protein maintains a molten globule like structure, which is what **Error! Reference source not found.** characterises.

The molten globule state of α-lactalbumin appears to have a shorter existence (~20s) under these conditions than the other proteins (Figure 7.3). The existence of its molten globule state may have been lengthened by the absence of a metal cofactor which would aid the rate of refolding [20]. Taking this factor into account the molten appearance of

the molten globule state does appear to agree with related work on the refolding of α-lactalbumin where tryptophan fluorescence was seen to be quenched after ~10s [21].

These figures clearly show that fluorophore mobility, or molten globule state, is present well after the hydrophobic collapse and secondary structure formation under these refolding conditions, and appears to be a common feature across proteins.

7.4 Aggregation produced from folding
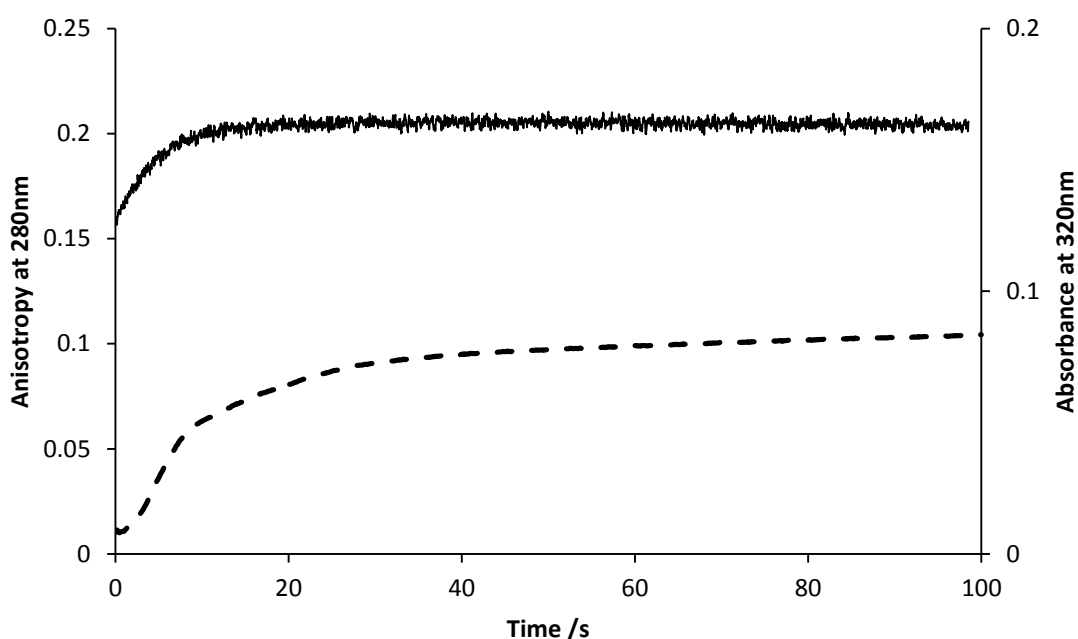
Figure 7.4: Refolding of cytochrome C at a 1:10 ratio followed by fluorescence anisotropy at 280 nm (solid) and absorbance at 320 nm (dotted) line.

The emergence of protein aggregates within the stopped flow sample can be measured by the absorbance of light at 320 nm [22]. This was undertaken alongside anisotropy and ANS probing of the refolding of cytochrome C (Figure 7.4). It can be seen that a small amount of aggregation takes place during the refolding process. At time 0, there is already a presence of aggregates, as can be noted by the small absorbance value at time 0 in Figure 7.4. As the absorbance is baselined to the unfolded sample of cytochrome C, this represents a production of aggregates during the dead time of the instrument, during

the hydrophobic collapse of the protein structure. The absorbance graph then represents an increasing rate of aggregation; either in aggregate number or size, before this growth subdues exponentially.

The high rate of growth appears to concur with the protein being still in a mobile state, however growth still occurs; albeit more slowly when the protein has adopted its rigid structure. Clearly whilst the protein is in the molten globule state it requires some side chain packing in order to stabilise it as might be expected [23], otherwise as Figure 7.4 shows, it has a tendency to aggregate. Furthermore, this increase in aggregation continues whilst the protein is in its rigid form, indicating that the aggregates may be inducing the native protein into an aggregate prone conformation, a behaviour that has been observed in other protein aggregate systems [24].
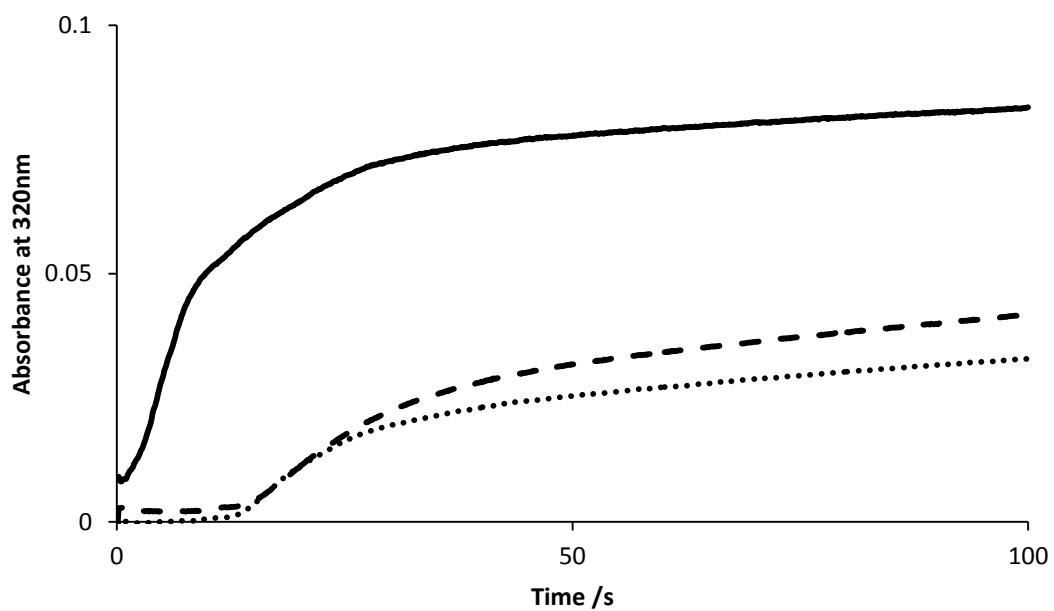

Figure 7.5: Absorbance of cytochrome C at 320 nm during refolding at 1:10 ratio into 100 mM Tris buffer containing 0 mM NaCl (full line) 250 mM NaCl (dashed line) and 500 mM NaCl (dotted line).

One of the most common methods for reducing aggregation during refolding is to use NaCl concentrations [25]. As can be seen from Figure 7.5, increasing salt

concentrations does have this effect on cytochrome C. It can be seen that the initial formation of aggregates is lower by comparison of values at t=0 for refolding taking place at higher NaCl concentrations. This higher concentration appears to prevent the increase in aggregate formation for a substantial period of time, and results in smaller absorbance values seen over the 100s period investigated here. The presence of salt appears to prevent aggregation occurring whilst cytochrome C is in its most mobile state, however growth appears to accelerate whilst the protein is closer to its native state (Figure 7.6). This growth continues whilst the protein appears to hold its rigid structure, again indicating that the native protein is being induced to aggregate by existing aggregates.

Figure 7.6: Absorbance at 320 nm (dashed line), and anisotropy at 280 nm (solid line) of Cytochrome C refolded at 1:10 into Tris containing 250 mM NaCl.

Additionally, the effect of final denaturant concentration on aggregation during refolding was investigated. This was achieved using different ratios of unfolded protein to refolding buffer. The effect of these ratios at a high concentration of NaCl is shown in Figure 7.7.

Where the concentration of urea is high (i.e a 1:5 ratio), it can be seen that there is minimal increase in aggregation, and the rapid growth rate which is seen at any of the 1:10 ratios (Figure 7.5) is not present. Here the high concentration of urea must be preventing protein-protein interactions which would increase aggregation.

Figure 7.7: Absorbance at 320 nm of Cytochrome C refolded at 1:5 (solid line) , 1:10 (dotted line), 1:25 (dashed line).
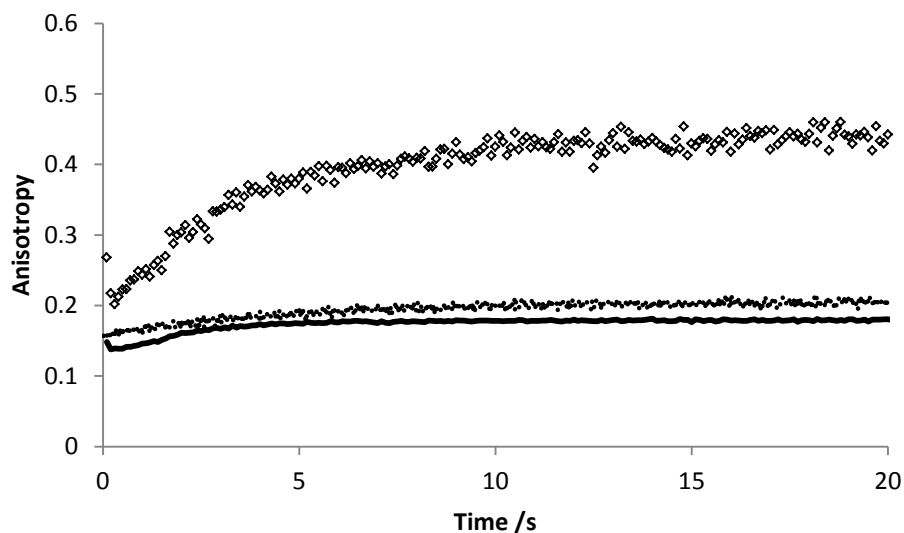
Figure 7.8: Anisotropy at 280 nm of Cytochrome C in 8 M urea unfolded at 1:25 ratio (solid line) 1:10 ratio (dotted line) 1:5 ratio (open dots).

At the low concentrations of urea (i.e 1:25 ratio) the rapid growth section of the aggregates is even steeper than that seen at 1:10 ratio, with minimal lag phase after the hydrophobic collapse before a rapid increase in aggregates begins. It is worth noting from Figure 7.10 that the rapid increase also occurs during the timeframe where the protein is at its most mobile, and again the growth continues whilst the protein appears to be in the rigid native state. Figure 7.7 clearly shows that a higher step dilution of urea concentration results in increased levels of aggregation rather than just the levels of non native protein, and indicates that stepwise urea dilutions should indeed increase the yield from refolding [26, 27].

Fluorescence measurements performed on the refolding products of cytochrome C at the 3 ratios indicate a difference in the quality of folding at each ratio.
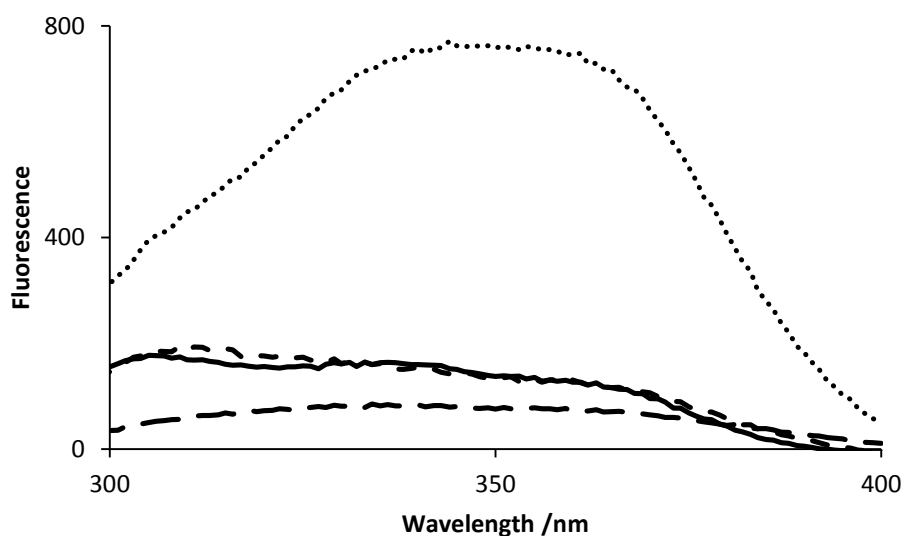
Figure 7.9:Fluorescence of cytochrome C with emission at 280 nm on native cytochrome C (solid line), and products of refolding at 1:5 (dotted line) 1:10 (short dotted line) and 1:25 (long dotted line).

Figure 7.9 illustrates that the refold ratio plays a critical role in the outcome of the fold quality. At a low refold ratio, sufficient denaturant is still present to cause a large increase in fluorescence from excitation at 280 nm from cytochrome C in largely unfolded state, whilst at the high ratio of 1:25, there is a decreased signal, which can be

attributed from the presence of aggregates absorbing light over these wavelengths. (Figure 7.7), whilst at 1:10 the fold can be seen to be very similar to that of the native state.
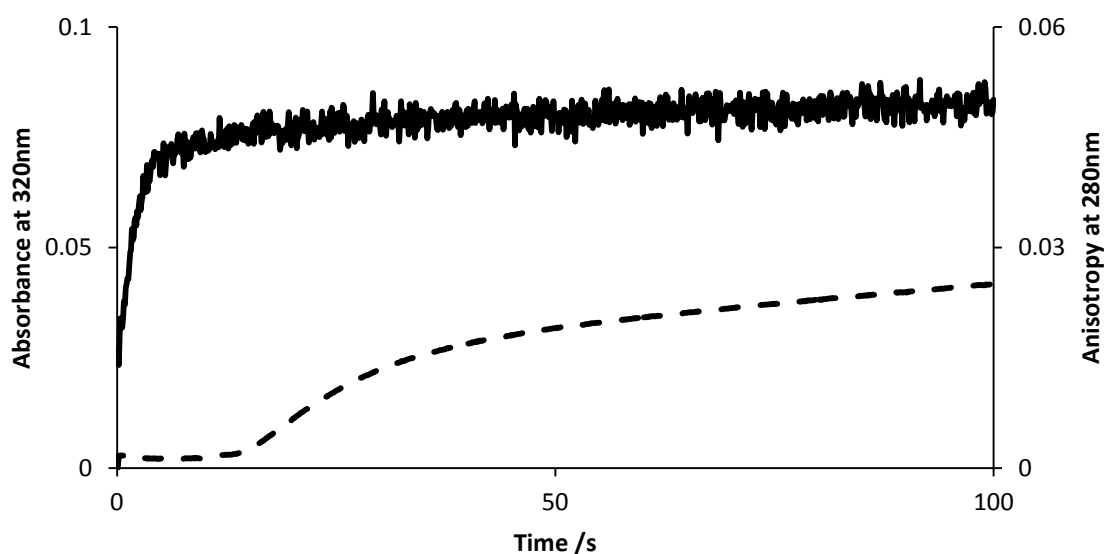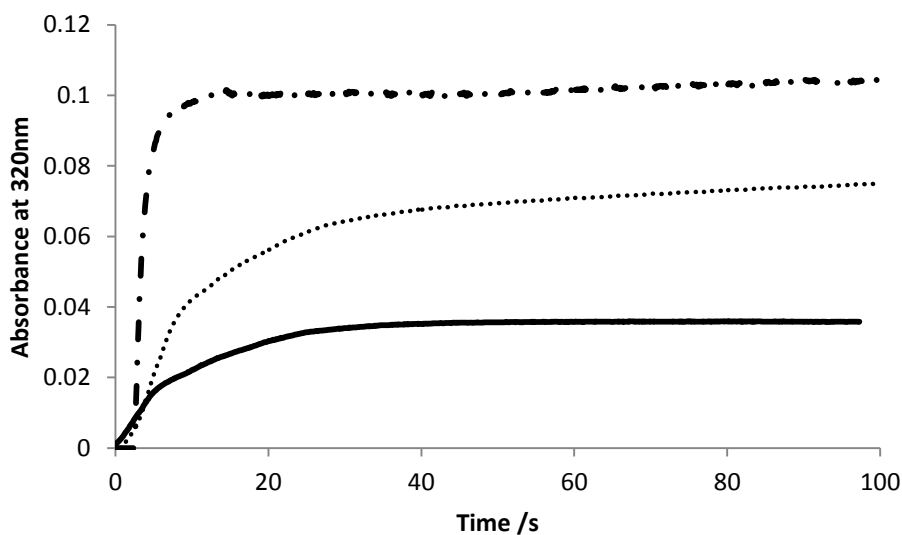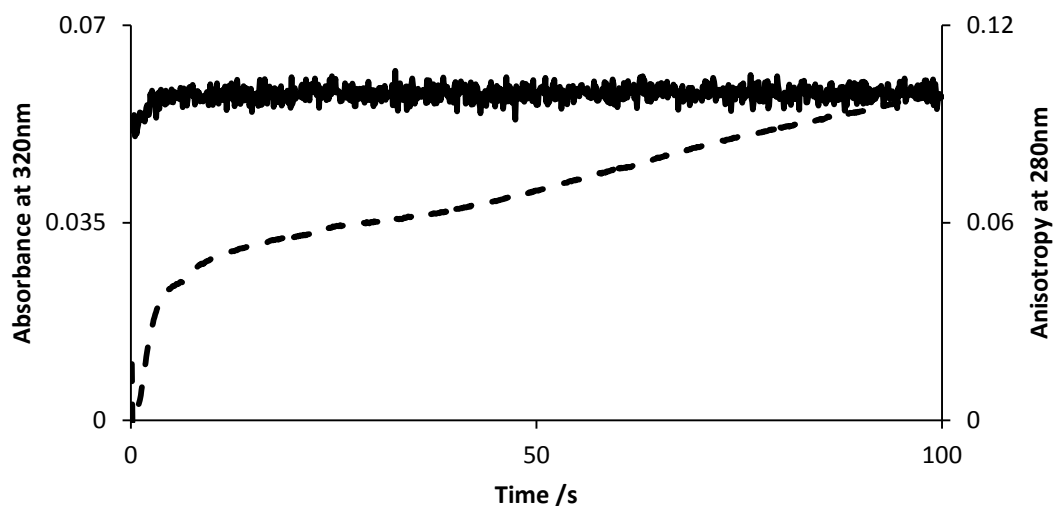

Figure 7.10: Absorbance at 320 nm (dashed line) and anisotropy at 280 nm (solid line) of cytochrome C refolded at 1:25 in 100 mM Tris 500 mM NaCl conentration.

The mobility of the protein in Figure 7.10 achieves a rigid state in a shorter time span than that of the protein when refolded at a 1:10 ratio. This is also the case for 1:5 refolding of cytochrome C. Therefore this may be related to the remaining presence of urea within the solution. At low concentrations, the urea is rapidly eliminated from the structure, allowing the protein to achieve its rigid state rapidly, whilst at much higher concentrations of urea, the dilution may only cause a partial unbinding of urea, allowing it to adopt a second, partially unfolded state. This state however is clearly aggregation prone as can be seen in Figure 7.10. The effect of salt on the aggregation state of the protein will be discussed in the concluding chapter, as its effects have been probed in more than the work seen in this chapter, however the results seen here may be relate to the Hoffmeister effect.[28, 29]

7.5 References

1. Clark, E., *Protein refolding for industrial processes.* Current Opinion in Biotechnology, 2001. **12**(2): p. 202-207.
2. Hermeling, S., et al., *Structure-Immunogenicity Relationships of Therapeutic Proteins.* Pharmaceutical Research, 2004. **21**(6): p. 897-903.
3. West, S.M., J.B. Chaudhuri, and J.A. Howell, *Improved protein refolding using hollow-fibre membrane dialysis.* Biotechnol. Bioeng., 1998. **57**(5): p. 590-599.
4. Li, M., G. Zhang, and Z. Su, *Dual gradient ion-exchange chromatography improved refolding yield of lysozyme.* Journal of Chromatography A, 2002. **959**(1-2): p. 113-120.
5. Jungbauer, A. and W. Kaar, *Current status of technical protein refolding.* Journal of Biotechnology, 2007. **128**(3): p. 587-596.
6. Stryer, L., *The interaction of a naphthalene dye with apomyoglobin and apohemoglobin.* Journal of Molecular Biology, 1965. **13**(2): p. 482-495.
7. Zhou, R., et al., *Hydrophobic Collapse in Multidomain Protein Folding.* Science, 2004. **305**(5690): p. 1605-1609.
8. Akiyama, S., et al., *Stepwise formation of α-helices during cytochrome c folding.* Nature Structural & Molecular Biology, 2000. **7**(6): p. 514-520.
9. Takahashi, S., et al., *Folding of cytochrome c initiated by submillisecond mixing.* Nature Structural & Molecular Biology, 1997. **4**(1): p. 44-50.
10. LeTilly, V. and C. Royer, *Fluorescence anisotropy assays implicate protein-protein interactions in regulating trp repressor DNA binding.* Biochemistry, 1993. **32**(30): p. 7753-7758.
11. Steiner, R.F., *Fluorescence Anisotropy: Theory and Applications*, in *Topics in Fluorescence Spectroscopy: Volume 2*, J.R.Lakowicz, Editor. 1991, Plenum Press: Baltimore.
12. Sadqi, M., L. Lapidus, and V. Muñoz, *How fast is protein hydrophobic collapse?* Proceedings of the National Academy of Sciences, 2003. **100**(21): p. 12117-12122.
13. Cheung, M., A. García, and J. Onuchic, *Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse.* Proceedings of the National Academy of Sciences, 2002. **99**(2): p. 685-690.
14. Akiyama, S., et al., *Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering.* Proceedings of the National Academy of Sciences, 2002. **99**(3): p. 1329-1334.
15. Shastry and H. Roder, *Evidence for barrier-limited protein folding kinetics on the microsecond time scale.* Nat Struct Mol Biol, 1998. **5**(5): p. 385-392.
16. Garel, J.-R. and R. Baldwin, *Both the Fast and Slow Refolding Reactions of Ribonuclease A Yield Native Enzyme.* Proceedings of the National Academy of Sciences, 1973. **70**(12): p. 3347-3351.
17. Day, I., et al., *Refolding of ribonuclease A monitored by real-time photo-CIDNP NMR spectroscopy.* Journal of Biomolecular NMR, 2009. **44**(2): p. 77-86.
18. Arai, M., et al., *Kinetic refolding of β-lactoglobulin. Studies by synchrotron X-ray scattering, and circular dichroism, absorption and fluorescence spectroscopy.* Journal of Molecular Biology, 1998. **275**(1): p. 149-162.

19.   Bhattacharjee, C., et al., *Structural Changes of β-Lactoglobulin during Thermal Unfolding and Refolding – An FT-IR and Circular Dichroism Study.* The Protein Journal, 2005. **24**(1): p. 27-35.

20.   Bushmarina, N., et al., *Cofactor effects on the protein folding reaction: Acceleration of α-lactalbumin refolding by metal ions.* Protein Science, 2006. **15**(4): p. 659-671.

21.   Schlepckow, K., et al., *Conserved Folding Pathways of α-Lactalbumin and Lysozyme Revealed by Kinetic CD, Fluorescence, NMR, and Interrupted Refolding Experiments.* Journal of Molecular Biology, 2008. **378**(3): p. 686-698.

22.   Murphy, R.M. and C.C. Lee, *Laser Light Scattering as an indispensible tool for probing protein aggregation*, in *Misbehaving Proteins - Springer ebook.* 2006.

23.   Wu, L. and P. Kim, *A specific hydrophobic core in the α-lactalbumin molten globule.* Journal of Molecular Biology, 1998. **280**(1): p. 175-182.

24.   Wang, W. and C. Roberts, *Aggregation of Therapeutic Proteins.* 2010: Wiley.

25.   Middelberg, A., *Preparative protein refolding.* Trends in biotechnology, 2002. **20**(10): p. 437-443.

26.   Tsumoto, K., et al., *Practical considerations in refolding proteins from inclusion bodies.* Protein Expression and Purification, 2003. **28**(1): p. 1-8.

27.   Gu, Z., Z. Su, and J.-C. Janson, *Urea gradient size-exclusion chromatography enhanced the yield of lysozyme refolding.* Journal of Chromatography A, 2001. **918**(2): p. 311-318.

28.   Baldwin, R.L., *How Hofmeister ion interactions affect protein stability.* Biophysical Journal, 1996. **71**(4): p. 2056-2063.

29.   Kunz, W., P. Lo Nostro, and B.W. Ninham, *The present state of affairs with Hofmeister effects.* Current Opinion in Colloid & Interface Science, 2004. **9**(1-2): p. 1-18.

Chapter 8 : **Conclusions**

Results presented in the chapter on aggregate nucleation indicate that the nucleation stage of aggregation in bovine serum albumin has a temperature dependant mechanism. Whilst the range of temperatures examined here are above the reported melting point for BSA, it is apparent that the protein is not completely unfolded and therefore the extent of unfolding will vary with temperature, and appears to have a significant effect on the nucleation mechanism. The mid range temperature appears concurrent with mechanisms postulated in the literature for other proteins; that the nucleation is a rapid establishment of a stable concentration of aggregate. However at the extremes of this range, it appears that the mechanisms deviates from this and that there may be clustering of highly reactive nuclei at high temperatures (formed from highly unfolded monomer), and continual formation of aggregate nuclei at low temperatures (formed from minimally unfolded monomer). From this work it is clear that minimal unfolding within the BSA monomer is sufficient to cause aggregation to the extent that they form large clustered agglomerates. A continuation of this work to yield results across an entire temperature range may yield enough data to form a coherent model.

Analysis presented on the growth of particulate aggregates show that the model of monomer addition to the aggregate nuclei appears to be a fitting description of the growth process. This differs from the model in the previous chapter in that during this growth period, there is a stable concentration of aggregates formed, therefore monomer-aggregate interactions are more prevalent that monomer-monomer interactions. This model is not just suitable for BSA studied in the previous chapter, but appears to be a generic form of aggregate growth applicable to proteins with a range of structural features. The detailed analysis from the UVLSS technique provides thermodynamic

values which appear in agreement with comparable literature values, and that of other studies used within this document. There is some apparent discrepancy in the state of the BSA molecule when it is aggregation prone. Kratky plots indicate that the protein molecule is only partially unfolded for aggregation to occur whilst the UVLSS model indicates that the BSA molecule is unfolded at 80 °C. The UVLSS studies however are purely an indication of when aggregation occurs; therefore what is referred to as the "unfolded" state is one that is sufficiently unfolded to initiate aggregation, which is not necessarily a completely unfolded molecule. Furthermore the technique provides a numerical method for examining the efficiency of aggregate preventing additives, and also illustrates the mechanism by which the additives prevent aggregation. A prime example of such is the ability of Tween 20 to reduce aggregation by reducing the collision constant between aggregate prone protein molecules.

Results presented in the chapter on aggregation during refolding indicate the use of fluorescence anisotropy to monitor the molten globule state during refolding of proteins. The mobility of fluorescent side chain residues has been shown to be prevalent during the refolding stage, with the mobility in existence for up to 30s, with some variance between protein species. Most strikingly, there is an obvious relationship between the mobility period of the protein and its propensity to aggregate. However, this mobility is also a necessary step in order for the protein to achieve its native structure, therefore a trade-off exists between achievement of native protein structure and aggregate prevalence.

Furthermore, this body of work has examined the role of salt using a number of these techniques, and whilst it appears that the charge shielding properties appear to stabilise the protein during hydrophobic collapse at low temperatures, at high concentrations and temperatures it exhibits a destabilising effect on the protein, and promotes aggregation.

219

This behaviour can be attributed to the Hofmeister effect, whose mechanism remains speculated upon in the literature. The sodium ions increase the solubility of the peptide residues; preventing aggregation. This may also reduce the number of successful collisions resulting in aggregation, as seen from the decreasing K4 value with increasing salt concentration in the UVLSS chapter. However, when presented with a folded protein, a sufficient concentration of sodium ions can cause the precipitation of hydrophobic groups, resulting in aggregation. This interaction with hydrophobic groups can explain the results seen from UVLSS data, where increasing salt concentration increases the rate of aggregation. This effect can appear counter-intuitive, and merits further investigation into the effects of salts on protein solubility.

Overall, this work provides insights into the nature of protein aggregation: illustrating that nucleation mechanisms are more complex than current literature suggests, presenting a technique for examining protein aggregation kinetics and thermodynamics ,presenting examples of analysis of aggregation preventing additives, and illustrating that the mobility of the protein has a direct correlation to its propensity to aggregate, whilst also highlighting areas of the research which merit further investigation.

Chapter 9 **: Future Work**

Below are listed future experiments relating to each section of work, and how the knowledge of each area could be expanded:

a) Aggregate nucleation using neutron scattering (Chapter 5) − Whilst the work here provides information on the nucleation of particulate aggregates, it fails to probe events on a single protein molecule size scale. Further experiments could be performed similar to those here, but at a lower temperature range, thus slowing down the nucleation, allowing for better statistics to be collected over the small time scales. As has been shown here, the behaviour of the nucleating monomers appears to change with temperature, therefore elucidating the mechanism of these higher temperature monomers may require a beamline with more flux or detector sensitivity than currently exists, and alternatively X-ray studies may be better suited.

b) Aggregation preventing additives studies using UVLSS (Chapter 6) − The work in this section paves the way for studies to be performed on all manner of aggregation preventing additives to elucidate their effectiveness and method of action. A prime example would be to test the use of amino acids as aggregate preventing additives which are stated in the literature.

c) Protein mobility and aggregation (Chapter 7) − In order to compliment the mobility work performed here, temperature studies could be performed using fluorescence anisotropy to assess the mobility of the protein during heating. In addition, single molecule studies such as FRET would provide much greater detail of exactly how this mobility is brought about.