A study of folded, denatured and aggregated states during the refolding of inclusion body proteins

A thesis submitted to the University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Engineering and Physical Sciences

2015

James Gilburt

School of Chemical Engineering and Analytical Science

List of contents

List of figures

List of tables

List of abbreviations

Arg HCl – arginine hydrochloride

Gdm HCl – guanidinium hydrochloride

Gdm – guanidine

GSH - gluatathione

GSSG – glutathione disulphate

HPLC – high performance liquid chromatography

HSQC - heteronuclear single quantum coherence

IBs – inclusion bodies

Word count: 43,965

Abstract

The need to high quality therapeutic proteins has grown significantly in the past 30 years. Recombinant proteins are often produced from vectors inserted into E. coli cell lines for large scale production. However, over-expression of the protein within the cell can lead to the formation of large, insoluble aggregates known as inclusion bodies. Native monomer protein can be isolated from inclusion bodies through a refolding process. This entails disruption of the aggregate structure with high concentrations of denaturant and renaturation in native-promoting solution. Our work characterises protein-protein interactions and aggregation between partially unfolded proteins during the refolding process. The protein-protein interactions are characterized in terms of the osmotic second virial coefficient (B_{22}). A positive value indicates repulsive interactions while a negative value indicates attractive interactions. Measurements are carried out for lysozyme, ribonuclease A and preproinsulin as a function of pH, ionic strength and denaturant concentration, alongside a range of known refolding excipients.

Past studies (Ho and Middelberg, 2004; Ho et al., 2003) have shown a link between higher B_{22} values in denaturant solutions and reduced aggregation during refolding. Our experiments have focused on the effects of urea and GdmHCl upon protein-protein interactions, alongside how ionic strength and refolding additives influence interactions between partially-folded states. At low ionic strength, solutions of urea increase net repulsive interactions compared to GdmHCl solutions through an attenuation of short-range attractive interactions. Electrostatic repulsive interactions are screened in solutions of GdmHCl due to the increased ionic strength of the solution; however short-range attractive interactions are also attenuated in a similar fashion to urea solutions. Protein-protein interactions in low and high concentration denaturant solutions have been shown to be highly sensitive to ionic strength and refolding experiments have shown that this correlates with increased aggregation during refolding. The solubilising additive Arg HCl has been shown to reduce short-range attraction between proteins in urea solutions, while the folding-promotor additives sucrose and hexylene glycol have been shown to have a more complex effect on protein-protein interactions in urea solutions dependent on denaturant concentration.

Within the wider context of the field of protein aggregation and refolding, the work conducted here will contribute towards the understanding of how denaturants and solutes influence attractive protein-protein interactions and aggregation behaviour between unfolded or partially folded proteins.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

My thanks to Dr Robin Curtis and Dr Jeremy Derrick for their guidance and patience.

My thanks to Dr John Liddell and the Fujifilm Diosynth staff for all their help and support.

My thanks to all those in both Robin and Jeremy's labs who gave advice and support during my experiments.

My thanks to Dr Davor Kovacevic at the University of Zagreb for his help and advice

My thanks to my partner Maeve, my family and my friends for all their love and support.

And finally, my thanks to beer and caffeine, the foundation stones of Western civilisation, for everything they've done.

Chapter 1

1.1 Introduction and literature survey

The demand for pure, high quality recombinant protein has increased dramatically over the past 30 years, as structural biology, the biopharmaceutical industry and other areas of biotechnology require large quantities of high quality, pure protein in the cheapest possible way. One of the most common methods for the production of recombinant protein is through the insertion of a DNA vector containing the desired gene into E.coli cells, the induction of overproduction of the protein through regulatory pathway stimulus, followed by cell lysis and several stages of protein purification. Often the recombinant protein is expressed within the cell as “inclusion bodies”, large aggregates of unfolded or misfolded protein several microns in diameter which can take up a large percentage of the cell volume. Though this may seem undesirable when the objective is to obtain a soluble monomer product, the formation of inclusion bodies can be practically beneficial (Bowden et al., 1991). They are easily separated from the cell lysis supernatant by centrifugation and subsequent chemical washing, requiring fewer post-lysis purification steps than with a soluble protein.

Insoluble inclusion body proteins are reconstituted by a process known as refolding, whereby the protein is solubilised, followed by changing the solution conditions to allow the protein to fold into a native conformation (Basu et al., 2011). The purified aggregates are put under mechanical and chemical pressures that break apart the ionic, hydrophobic and covalant binding that hold the construct together. The most commonly used method is through incubation of the inclusion body in a solution containing a high concentration of a chaotropic denaturant, such as Gdm HCl or urea. The chemical conditions are then altered to reduce the concentration of the denaturant to non-denaturing levels through dilution, dialysis or chromatographic systems, and correct native folding is promoted through a carefully controlled solute environment. Factors including pH, ionic strength, co-solvent selection and concentration, protein concentration, pressure and temperature all contribute to the refolding environment and the optimal environment differs for each unique protein sequence. The biggest loss of product during this process is due to the re-aggregation of protein upon refolding, which are either filtered or chromagraphically removed as a waste product.

On an academic scale this process may only require a few litres of bacterial growth medium and several labour-intensive days before 10-100 miligrams of the desired protein can be purified. In an industrial setting however, where tonnes of therapeutic proteins such as insulin or human-growth-factor-1 are produced, more high throughput and high volumes are required. Due to the large scale of these operations, a careful cost-benefit analysis is undertaken between a high recovery yield of soluble protein from the inclusion bodies and the constraints of high running costs and practical considerations. Greater understanding of how solution environment during refolding affects the overall aggregation and refolding kinetics will lead to more effective methods of optimising industrial-level refolding and protein purification.

Within the past decade, research has shown that there is a clear link between the level of aggregation during refolding and the second virial coefficient (B_{22}), the value which represents the averaged strength of protein-protein interactions between two bodies in solution. Previous findings have shown that the B_{22} value of a protein measured in a wide range of solution conditions can be a strong indicator for what conditions crystallisation and aggregation will occur (George and Wilson, 1994). Ho and Middelburg (2004) have shown that proteins under solution conditions that promote a more repulsive protein-protein interactions as measured through the B_{22} value in both native and denaturing conditions correlates with a higher yield during refolding (Ho and Middelberg, 2004; Ho et al., 2003). This evidence suggests that the monitoring and manipulation of the B_{22} of a solute protein can be used as a strong indicator for the aggregation propensity of a protein during refolding. However, much is still not known about the molecular interactions that control B_{22} in different denaturant solutions and how B_{22} is relates to aggregation.

The aim of this project is to investigate the effects of co-solvents and denaturants upon the protein-protein interactions of folded, unfolded and partially-folding proteins and elicit whether a link exists between the measurements of the strength of protein-protein interactions and the aggregation propensity of a protein. This will be achieved through characterising the protein-protein interactions in terms of the second virial coefficient (B_{22}) and mutual diffusivity (D_m) of the model proteins under a wide range of refolding and denaturing conditions, and the measurement of aggregation of partially-folded and natively folded proteins using a range of techniques. This work shall bring together current knowledge on the effects of pH, salts, and excipients upon folding, aggregation kinetics and protein-protein interactions and assess whether a viable method for predicting the “refoldability” of a given protein can be developed from current knowledge. Reliable methods for determining refolding yield, through qualitative and quantitative analysis shall also be developed, alongside measurement of the folded state of the purified protein.

This project aims to understand how protein-protein interactions during inclusion body refolding can influence aggregation during the refolding step. The solubilisation and refolding of inclusion body proteins has been intensely studied to ascertain the mechanisms which influence and alter the yield. Below is a survey of current literature focusing on several areas identified as important factors in protein refolding. The first section focuses on the effects of co-solvents, denaturants and solution conditions on protein-protein interactions and stability, which have a significant influence on refolding. The second section gives an outline of the current understanding of rapid aggregation and how aggregation is observed in refolding. The third section reviews the three model proteins that shall be used in the project, and the reasons for their use. The fourth section looks at current methods with which refolding is achieved, and how these have been modified with increasing knowledge of refolding. The fifth section outlines the analytical techniques employed to reveal the folded state of

purified proteins from refolding solutions. The final section describes how protein-protein interactions are characterised.

1.2: Co-solvents and denaturants

1.2.1Co-solvents used for protein refolding

The selection of the co-solvents used in the renaturation buffer has a vital influencing factor on the aggregation kinetics and yield of the refolding protein (Clark et al., 1999; Rudolph and Lilie, 1996). Common solvents that promote native folding and suppress aggregation include amino acids, polyols, sugars and detergents, as well as other families of small molecules (Arakawa and Timasheff, 1985). The additives are included in varying combinations and concentrations during a refolding experiment, the selection of which is specific to each protein of interest. Excipients that stabilise proteins against unfolding are, in general, preferentially excluded from the protein surface (Arakawa et al., 2007a; Arakawa et al., 2007b; Arakawa and Timasheff, 1982; Kita et al., 1994). A class of these additives are known as kosmotropes (Qu et al., 1998). A second class of excipients increase protein solubility through increasing protein-protein repulsion to reduce the attractive interactions between partially-folded proteins which lead to aggregation during refolding. Repulsion is promoted between proteins through preferential interaction of the excipient with specific amino acid side chains to reduce attractive short-range interactions such as hydrogen-bonding and hydrophobic attraction, and increasing the net charge of the protein, thus increasing double-layer electrostatic repulsion (Arakawa et al., 2007c; Wen et al., 2015).

Molecules such as proline, sucrose, glycerol and other small additives promote correct protein folding and structural stability through the contraction of expanded protein structures and exhibiting exclusion from the protein surface in both denatured and native states (Qu et al., 1998; Samuel et al., 2000; Timasheff, 1991). The exclusion of excipients from the protein surface leads to an increase in the chemical potential of the protein. A higher chemical potential on the surface of the protein leads to the preferential hydration of the protein. The effect of increasing the chemical potential of the protein through preferential exclusion increases in strength proportional to the surface area of the protein (Street et al., 2006). Therefore, a larger surface area will lead to greater hydration (Arakawa and Timasheff, 1982). Thus, because the transition between native and unfolded conformations leads to an increase in solvent-exposed surface, the chemical potential of the unfolded state becomes higher than that of the more compact folded structure when at high sugar or polyol concentration (Timasheff, 1998, 2002). A higher chemical potential of the unfolded state relative to the folded state leads to a contraction of the protein structure and therefore native folding. Also, a reduction in the population of intermediate folded states is observed as the protein structure contracts. Kendrick (1997) showed that sucrose reduces conformational mobility within an interleukin-1 receptor antagonist, and depopulates

partially-folded states within the protein solution (Chi et al., 2003; Kendrick et al., 1997). Nicoud et al (2015) performed a study comparing the effect on different sized polyol sugars on monoclonal antibody stability and aggregation (Nicoud et al., 2015). A link was found between polyol sugar size and the antibody unfolding rate, where a larger size and hydrogen-bond capacity of the polyol leads to an increase in the structural stability of the protein, though this effect diminished for sugars larger than sorbitol. Rapid structural collapse from a denatured to a native state and correct folding of the protein are essential during refolding, as intermediate folding structures are prone to aggregation,. Thus, using preferentially-excluded osmolytes such as sugars and polyols decrease the effective level of aggregation through promotion of correct protein folding, and increases protein-protein repulsion, as measured through the second virial coefficient (Liu et al., 2004).

The exclusion behaviour of many osmolytes contrasts with the behaviour of urea and guanidinium. Using denaturants promotes protein chain expansion through preferential binding of the denaturant to the protein backbone leading to solubilisation of hydrophobic residues (Bennion and Daggett, 2003). However, at low concentrations (0.25-0.5M), urea and Gdm HCl can be used as effective refolding additives. Urea and Gdm HCl both have strong solubilising effects on proteins. The aggregation suppression effect is due to the binding of denaturant molecules to hydrophobic residues on the protein surface and the reduction of the hydrophobic inter- and intra- molecular interactions (Rossky, 2008). At low (<1 M) urea concentrations, little structural unfolding occurs, but suppression of aggregation is observed (Hedoux et al., 2010; Qu et al., 1998).

Arginine is one of the most common additives used in refolding, predominantly used at concentrations greater than 0.5 M (Arakawa and Tsumoto, 2003). Protein aggregation during refolding is prevented by a different mechanism in arginine solutions than that of low concentrations of denaturants, sugars or polyols. Arginine has a unique guanino group as a side chain that is of the same structure as Gdm. Both arginine and Gdmshare the effect of suppressing aggregation (Arakawa and Tsumoto, 2003). The effectiveness of arginine to suppress aggregation at 0.5 M is due to the affinity of arginine for most side chains of amino acids through strong van der Waals forces (Wen et al., 2015). This behaviour is similar to Gdm HCl and leads to the preferential binding of arginine to the protein over water molecules (Kita et al., 1994). This binding is strongest to aromatic residues, such as tryptophan and tyrosine, where planer stacking of the guano group of arginine and the ring of the aromatic residues has been observed (Flocco and Mowbray, 1994). These residues are linked to the formation of attractive protein-protein interactions which lead to aggregation (Arakawa et al., 2007b). The binding of arginine to the protein surface is however limited in the number of contacts it can make with a protein compared to Gdm HCl due to the large size of the amino acid (Manikwar et al., 2013). Thus the denaturing effects observed in Gdm HCl solutions are not observed in arginine solutions. However, arginine has been shown to slightly decrease the structural stability of proteins (Manikwar et al., 2013). Through this limited binding to hydrophobic patches, arginine suppresses aggregation,

while not significantly affecting the structure of the protein itself (Arakawa et al., 2007a). Arginine is most commonly used in high pH solutions, as it acts as a strong base. To be used in low pH solutions, a counter-ion must be used to ensure pH stability, such chloride, which leads to arginine becoming an ionic excipient. The use of Arg HCl can therefore be detrimental to increasing protein-protein repulsion, as an increased ionic strength can lead to the screening of long range electrostatic repulsion between proteins (Arakawa et al., 2007a). Evidence has also been presented that suggests the aggregation-suppression behaviour of arginine is due to of cluster formation by arginine molecules. The formation of arginine clusters leads to an increased excluded volume, screening of short-range attractive protein-protein interactions and reduced diffusivity of the protein solution through an increased solution viscosity. This leads to a reduction in protein-protein collisions and thus reduces aggregation events. It has been hypothesised that this is the primary mechanism through which arginine inhibits aggregate formation, as opposed to direct protein-arginine binding (Schneider et al., 2011). This cluster formation has been shown to be dependent on the co-salt used with the arginine molecule. Arginine is most commonly used as a salt with chloride ions but other salts have been shown to have a significant effect on the preferential interactions of arginine with the protein surface, broadly following the Hoffmeister series. The use of kosmotrophic ions, such as phosphate (HPO_4^{2-}) and sulphate (SO_4^{2-}) lead to a reduction the binding affinity of arginine to the protein surface, while chaotrophic salts such as bromide (Br-) and thiocynate (SCN^-) increase the binding affinity of arginine. At high arginine-salt concentrations, the protein surface becomes saturated with arginine. This leads to an increase arginine-arginine cluster formation, which in term lead to increased excluded volume effects between proteins and decreases protein-protein collisions (Schneider et al., 2011). This new research suggests that arginine reduces aggregation through two mechanisms: through binding to the surface of the protein thus reducing hydrophobic attractive forces, and through cluster formation within the solution, creating a repulsive excluded volume force that reduces protein-protein collisions and suppresses aggregation.

1.2.2 The second virial coefficient (B_{22})

The second virial coefficient (B_{22}) is equal to the solvent-averaged interactions between a pair of proteins in a solution, derived from the virial expansion of the osmotic pressure (Π) exerted by the protein on the aqueous solution. The virial expansion of the osmotic pressure is defined as:

$$\frac{\Pi}{k_B T} = \rho_\rho + B_{22}(T, \mu_w, \mu_s)\rho_\rho^2 + B_{33}(T, \mu_w, \mu_s)\rho_\rho^3 + \cdots \qquad \text{Equation (1)}$$

Where kB is the Boltzmann constant, T is absolute temperature (K), ρ_ρ is the protein number concentration, μ_w and μ_s are the chemical potentials for water and co-solvent (i.e. salt) respectively and B_{22} and B_{33} are the second and third virial coefficients, respectively.

Within this equation, the virial expansions represent the interactions between protein molecules within the solution. B_{22} reflects the combination of all types of interactions including all electrostatic, hydrophobic, hydrogen-bonding, van der Waals, hydration and ionic forces. As all of these forces are affected by the solution conditions such as ionic strength and pH, the value of B_{22} changes according to how alterations in these parameters affect protein-protein interactions. The second virial coefficient is measured in units of mLMol/ g^2, with a positive value representing overall repulsive interactions between two proteins, and negative values representing attractive values between two bodies (Curtis and Lue, 2006; Curtis et al., 2001).

The value of B_{22} is related to the *two-body potential of mean force, w_2*, which is the free energy required to bring two protein molecules from infinite separation to a centre-to-centre separation, *r*. This relationship is represented in the following equation:

$$B_{22} = -\frac{1}{2}\int dr d\,\Omega_1 d\Omega_2\; x\; \left\{\exp\left[-\frac{w_2\,(r,\Omega_1,\Omega_2;T,\mu_w,\mu_s)}{k_B T}\right]-1\right\} \qquad \text{Equation (2)}$$

where Ω_1 and Ω_2 represent the relative orientations of proteins 1 and 2 respectively, and w_2 is the potential of mean force for the protein.

This equation links the total interaction forces between two proteins to the value of the second virial coefficient, changes in the chemical environment (i.e. temperature and pH) affect the potential of mean force.

1.2.3 Effects of pH and ionic strength on the B_{22} of proteins

The net charge of a protein is governed by the solution pH and the individual pKa's of the arginine, lysine, histidine, glutamate and aspartate amino acid residues on the surface of the protein. Soluble proteins retain a charge at physiological pH, and changing the pH of the solution away from the pI of the protein leads to an increase in the net charge of the protein. The charge of a protein is important for several reasons; structural stability can be regulated by charge-charge interactions within and between proteins, while the solubility of a protein is been strongly linked to the net charge (Hayakawa (Hayakawa and Nakai, 1985a, b; Shaw et al., 2001; Strickler et al., 2006). The charge of lysozyme

has been measured through ion-titration by Kuehner et al (1999) and shows charge can change significantly over a small pH range, changing from a charge of +14 to +11 between pH 4.0 and pH 5.0 (Kuehner et al., 1999).

Protein-protein interactions at low ionic strength can be rationalised by considering proteins as hard charged spheres (based on DVLO theory) where long range electrostatic repulsion between like-charged spheres dominates protein-protein interactions, with short-range attractive interactions and excluded volume forces have smaller contributions (Curtis and Lue, 2006). Electrostatic protein-protein interactions are sensitive to the ionic strength of the solution, with increasing ionic strength leading to reducing B_{22} values (Curtis and Lue, 2006; Dumetz et al., 2007; Muschol and Rosenberger, 1995; Velev et al., 1998). Muschol and Rosenberger (1995) used static light scattering to show that at pH 4.7 in 50 mM sodium acetate buffer, the B_{22} value for lysozyme solutions is highly positive. However, with increasing NaCl concentration, the values of B_{22} rapidly decrease and become negative at concentrations greater than 400mM. This trend is due to the increasing ionic strength of the solution, which screens long range electrostatic interactions by reducing the solution screening length, *k*, reducing the repulsion between proteins.

The magnitude of protein-protein repulsion is also dependent upon pH. The pH of a solution controls the net charge of the protein, and the repulsion potential between two proteins is proportional to net charge squared. Lysozyme has a basic isoelectric point (10.5). Therefore at pH 4.7, lysozyme carries a strong positive charge, leading to strongly repulsive double-layer forces. At higher pH, the protein has a reduced charge and weaker double-layer repulsion, and therefore a lower concentration of ions is required to screen this repulsion (Curtis et al., 2002; Elcock and McCammon, 2001; Narayanan and Liu, 2003; Velev et al., 1998). In the absence of strong double layer forces, protein-protein interactions are governed by shorter-ranged forces such as van der Waals, hydrogen bonding, hydrophobic and salt-bridging. Due to the combination of these forces, B_{22} of lysozyme becomes negative as the salt concentration increases.

A study by Roberts et al (2015) in which the B_{22} values and the interaction parameter k_D (an interaction parameter of proportional value to B_{22}) of a monoclonal antibody with an isoelectric point (pI) of 9.1 was measured using static and dynamic light scattering at various pH values and ionic strengths (Roberts et al., 2015). The results are presented in figure 2.1, and show that at low ionic strength, the repulsion exhibited between proteins is strongly correlated with pH. At pH 5, when the antibody is strongly positively charged, the B_{22} value measured is high, as double layer electrostatic repulsion is strong due to the high net charge. Increasing the pH of the solution towards the pI of the protein reduces the net charge of the protein and therefore a decrease in repulsion is observed. Close to the pI of the protein at pH 9.0, protein—protein interactions become attractive, as the net charge of

the protein reaches zero and oppositely charged residues lead to attractive protein-protein electrostatic interactions. At pH values where the protein has a net positive charge, increasing ionic strength leads to reducing B_{22} and k_D values, as the increased ionic strength screens electrostatic repulsive interactions and net repulsion between the proteins is reduced. At pH 9.0, attractive electrostatic interactions are screened with increasing ionic strength, so protein-protein repulsion increases and the B_{22} and k_D values increases. At 280 mM ionic strength, all B_{22} and k_D values are nearly equal at all pH values, indicating that electrostatic interactions no longer contribute significantly towards protein-protein interactions due to being completely screened by the high ionic strength.

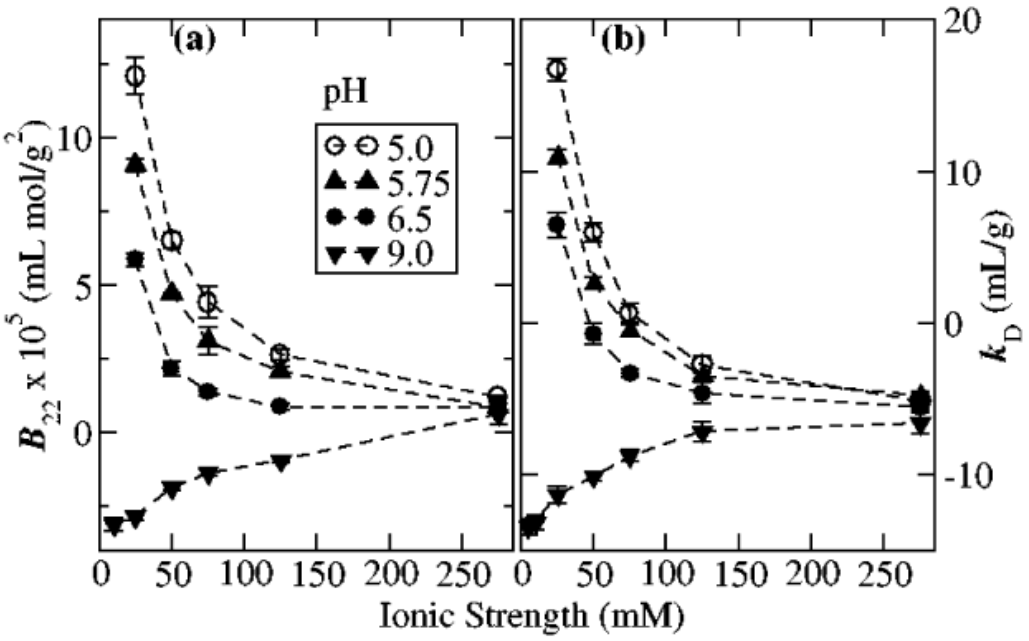


Figure 1.1. The B_{22} value (left) and k_D value (right) of an antibody at varying pH values as a function of ionic strength concentration) (Roberts et al, 2015).

In a bioprocessing context, methods of reducing repulsive interactions between proteins by increasing salt concentrations are primarily used for “salting out” proteins for either separation or crystallizing protein. In a refolding context, the presence of repulsive protein-protein interactions leads to a reduction in protein aggregation, so the use of high salt concentrations is undesirable. However, many refolding protocols require the use of a range of different salts and ionic strengths in the refolding buffer solution. This is due to the individual requirements of target proteins and shows the complexity of protein refolding.

1.2.4 Disulphide bonds: breaking and formation in refolding

Disulphide bonds are a key component of the native structure of approximately 65% of all secreted proteins in humans. Disulphides are formed between cysteine residues within the polypeptide chain and aid in the correct folding and structural stability of the protein (Mossuto et al., 2011). Correct bond formation is modulated by molecular chaperones and oxidoreductases that occur within both eukaryotic and prokaryotic cells. For example dsbA is a chaperone within E. coli and protein disulphide isomerase (PDI) by eukaryotes (Bardwell et al., 1991). Mutations or knock outs of the genes coding for chaperones lead to secretion of misfolded proteins with improper disulphide bond formation. Correct disulphide bond formation is essential for native tertiary protein folding within proteins containing one or more disulphide bonds, as the bonds can influence and guide hydrophobic collapse, backbone chain dynamics, packing of secondary structural components and inter-domain interaction (Silvers et al., 2012).

1.2.5 Disulphide bonds within Inclusion Bodies

Inclusion bodies contain aggregated proteins in both native and non-native conformations containing non-native intra- and inter chain disulphide bonds, alongside reduced cysteine residues (Schoemaker et al., 1985). The disulphide linkages covalently link the proteins within inclusion bodies that are predominantly held together with non-covalent interactions such as hydrophobic attraction and hydrogen bonding networks. The intermolecular disulphide linkages are problematic during the disruption of IB, as methods require the use of chemical reducing agents in addition to the addition of denaturants such as urea or Gdm HCl. The most commonly used reducing agents are dithiotheritol (DTT) and glutathione (GSH), which are strong small molecule reductants. These chemicals are used in denaturing solutions during the initial disruption of the inclusion bodies with the use of denaturants. The molar concentration used of the reductant should be between 5 and 10 times the molar concentration of cysteines within the inclusion body solution to ensure complete reduction of all cysteine bonds with the bulk protein solution.

1.2.6 Reforming of cysteine bonds upon refolding

Correct re-formation of intra-molecular cysteine bonds is an important step in the refolding process and thus choosing an appropriate redox environment is a key aspect of the refolding buffer (Basu et al., 2011). Several methods of oxidising have been shown to be effective at forming and shuffling

intra-molecular cysteine bonds until a native conformation is achieved. The most basic method is through air oxidation, in which an oxidising environment is achieved through bubbling oxygen gas through the solution. Though this method is by far the cheapest, air oxidation is not favoured by industry or academia as it allows very little control over the reaction rate or concentration of oxygen within the solution and a relatively low rate of shuffling of the cysteine bonds occurs. This leads to off-pathway intermediate formation and re-aggregation. The use of metal ions has been shown to provide a shuffling system, however due to the lack of control over the concentration of oxygen, the correct ratio of metal ion to oxygen is hard or impossible to achieve.

A more common method for protein oxidation utilises oxidising agents or redox couples in the refolding buffer, thus providing a controllable source of electrons for the free thiol groups. The choice of oxidising agent depends on several factors, most importantly protein-dependant oxidising efficiency. GSSG/GSH and cysteine/cystine couples have both been shown to significantly improve refolding yield of denatured lysozyme compared to air oxidation (Raman et al., 1996). However, the use of these ideal oxidising agents within large scale bioprocessing is limited by the prohibitive cost of GSSG and cysteine. Cheaper alternatives such as cystamine can be substitutes compared to the more efficient but costly GSSG and cysteine molecules. Short peptide disulphides, such as CGC and RKCGC, which mimic the active sites of prokaryotic isomerases such as PDI, can provide an effective form of oxidative bond shuffling (Wang et al., 2011). The use of peptide disulphides has been shown to increase yields by up to 10% compared to GSSG for model proteins. However, the production of such unique short peptides on an industrial scale is prohibitively expensive in comparison to the use of simpler oxidising agents.

1.2.7 Denaturant chemistry

The disruption of inclusion bodies requires the use of a strong chemical denaturant with which to solubilise the aggregated proteins. The two most common denaturants used are urea and Gdm HCl. When used at a high concentration and in conjunction with reductants, denaturants disrupt and solubilise inclusion bodies and creating a solution of denatured monomer proteins for the refolding process.

1.2.7.1 Guanidinium hydrochloride

Gdm is a small, highly polar molecule widely used as a protein denaturant, and is stored as a salt formulated with hydrogen chloride. Gdm HCl is highly soluble in water, allowing for concentrated solutions that are essential for protein denaturation. Protein denaturation occurs in solutions containing between 4 and 6M Gdm HCl (Liu et al., 2005). Though widely studied, a complete consensus on the exact mechanism of protein denaturation by Gdm HCl solutions is yet to be reached. However, a common theory relates to the preferential interaction of Gdm ions to hydrophobic residues and stabilisation of unfolded states through reduction in the hydrophobic interactions between core residues. Gdm ions have been shown to preferentially "stack" in a planer fashion to the surface of aromatic residues, and have a preferential interaction with aliphatic side chain residues, through van der Waals forces, displacing weakly-bound water molecules *(Dempsey et al., 2005; Mason et al., 2007; Shao et al., 2012)*. Dempsey et al (2004) showed that the denaturing effectiveness of Gdm HCl is enhanced considerably compared to urea when denaturing proteins whose structure is primarily reliant on Trp-indole hydrophobic/planar interactions (*Dempsey et al, 2004*). It is theorised that the favourable interactions between Gdm ions and aromatic residues may contribute towards the effectiveness of Gdm HCl at protein unfolded compared to urea. Though destabilisation of the protein structure through disruption of the hydrogen-bonding network is theorised by some (Dempsey et al., 2005; O'Brien et al., 2007), Gdm has been shown to have little significant hydrogen bonding with the amide backbone of a polypeptide chain directly, due to the weak hydrogen-bonding strength of the nitrogen atoms of the molecule (Lim et al., 2009). However, non-specific binding of Gdm to the peptide backbone through van der Waals forces has been reported. Gdm HCl also significantly increases the viscosity of the solution, which in turn reduces the end-to-end distance of polypeptide chains, as shown by Moglich et al *(Kawahara and Tanford, 1966; Moglich et al., 2005),*Overall, the binding of Gdm molecules to the hydrophobic residues of the protein weakens tertiary structure and solubilises the interior of the protein, displacing surface-bound water molecules and shifting the equilibrium of the system towards an unfolded state compared to a natively-folded state.

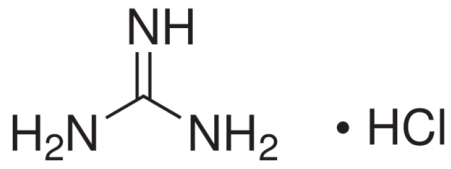


Figure 1.2. The chemical structure of Gdm HCl.

Gdm HCl has a measurable effect on the protein-protein interactions and hydrodynamic properties of the protein. Denaturation occurs to proteins in solution containing between 2 and 6 M Gdm HCl, as

observed through fluorescence, dependant on the pH of the denaturing solution and amino acid sequence of the protein (Liu et al., 2005). Lui and Cellmer (2005) performed multiple static and dynamic light scattering experiments using lysozyme as a model protein to determine the hydrodynamic radius and B_{22} of lysozyme in 0 to 8M Gdm HCl, without the presence of a reducing agent (Liu et al. 2005). The second virial coefficient of lysozyme solutions at pH 4.5 as a function of Gdm HCl concentration is shown in figure 2.2. The values of B_{22} indicate a reduction in the repulsive protein-protein interactions when increasing the Gdm HCl concentration to 1M, due to the screening of double-layer electrostatic repulsion between the proteins. However, interactions do not become attractive as is observed in solutions containing comparable NaCl concentration, due to the reduced hydrophobic attractive interactions between the proteins due to the presence of Gdm HCl (Curtis et al., 2001). The value of B_{22} increases with increasing Gdm HCl concentration above 1 M, indicating a maximum in protein-protein repulsion at 7 M denaturant. This increase in repulsion between proteins is due to the binding of Gdm ions to the surface of the protein, reducing hydrophobic attractive interactions. Small angle X-ray scattering analysis of lysozyme in various concentrations of Gdm HCl solutions by Javid et al (2007) show agreement with the data presented by Lui and Cellmer, showing an reduction in protein-protein repulsion at 500 mM Gdm HCl, and increased repulsion at higher concentrations (Javid et al., 2007). The hydrodynamic radius of lysozyme in Gdm HCl remained relatively constant at all concentrations, indicating little overall expansion of the structure while denaturing (Lui, 2005). This lack of expansion is due to the disulphide bonds within the protein structure, which covalently hold together distant sections of the peptide chain and therefore restrict expansion of the denatured structure.

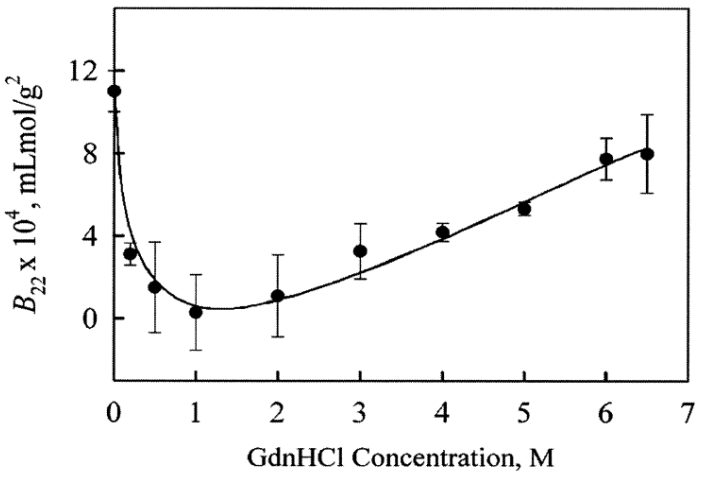


Figure 1.3. The B_{22} value of lysozyme at pH 4.5 as a function of Gdm HCl concentration. (Liu and Cellmer, et al. 2005).

1.2.7.2 Urea

Urea is a natural bi-product of the removal of ammonia from the body and is present in all cells in the body. Structurally similar to Gdm HCl, the molecules only differ by one atom, an oxygen atom in urea instead of a nitrogen atom in Gdm HCl. This small difference has significant effects upon the mechanism of the denaturant action on protein-protein interactions, as the oxygen atom provides a stronger hydrogen-bonding site than the nitrogen atoms, and reduces the planar structure of the molecule (Lim et al., 2009). Most noticeably, the concentration at which protein denaturation occurs is higher in urea solutions than with Gdm HCl. Denaturation of proteins occurs in solutions between 6 and 8 M urea, while 4 to 6 M Gdm HCl solutions are required for protein unfolding for the majority of proteins (Timasheff and Xie, 2003). Urea is also used at low concentrations (0.25 – 2 M) as a refolding co-solvent, as it has been observed to have aggregation-suppressing properties through the weakening of attractive protein-protein interactions and increasing protein solubility (Deyoung et al., 1993; Street et al., 2006; Winter et al., 2002; Zangi et al., 2009).

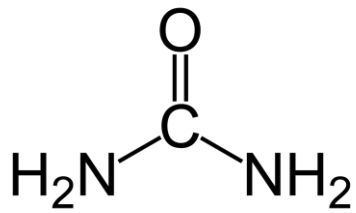


Figure 1.4. The chemical structure of urea.

Molecular dynamic simulations and experimental data have provided insights into the mechanism for unfolding of the polypeptide chain by urea. Though there is still considerable debate in the area, a commonly agreed mechanism states that protein unfolding in high concentration urea solutions occurs through a two-stage mechanism. Firstly, in an “initial binding” stage, urea interacts with the peptide backbone and hydrophobic residues on the protein surface through preferential van der Waals forces, lead to a shedding of weakly-bound water molecules from the primary hydration layer *(Bennion and Daggett, 2003; Hua et al., 2008; Rossky, 2008; Zou et al., 2002)*. Peptide-water hydrogen bonds are replaced by urea-peptide hydrogen bonds, though changes in the hydrogen-bonding network of the peptide is not believed to be the main thermodynamic driving force of unfolding in urea solutions (Lim et al., 2009). The water-shedding effect is enhanced around aliphatic and aromatic residues. Urea has been shown to significantly reduce the hydrophobic attraction between aliphatic surfaces due to the preferential interaction of urea with hydrocarbon chains in comparison to water molecules (Rossky, 2008). Secondly, in the denaturing stage, urea molecules penetrate the interior of the protein’s hydrophobic core through preferential interaction with aliphatic and aromatic residues. The binding of urea to the interior hydrophobic residues and peptide backbone

leads to an unfolding of the peptide chain and a shift in equilibrium to favour an unfolded state. This theory follows closely to Timesheff's description of preferential interactions between urea and amino acids in urea solutions increasing in strength proportional to urea concentration (Timasheff and Xie, 2003). Several articles have noted that between 0 and 2 M urea, proteins have been shown to change net charge, with Bovine Serum Albumin increasing in charge, and lysozyme decreasing in charge (Ortore et al., 2008; Sinibaldi et al., 2008). However, there is little literature regarding the effects of urea upon the short-range protein-protein interactions during denaturation as have been characterised with Gdm HCl, especially at low concentrations of denaturant (>2 M). Understanding the protein-protein interactions of partially folded states that occur in the presence of denaturants is important for the understanding and prevention of aggregate formation.

1.2.7.3 Sensitivity of protein unfolding in denaturants

The concentration of denaturant at which a protein denatures is dependent on several factors; the choice of denaturant, the pH of the solution and the biophysical characteristics of the protein. An early study by Greene and Pace (1974) compared the unfolding of various proteins in urea versus Gdm HCl solutions, revealing that proteins generally unfolded at a lower concentration of Gdm HCl than urea, the mechanistic origin of which is discussed previously (Greene and Pace, 1974). Later studies show the pH dependence of unfolding for lysozyme and ribonuclease A, amongst others, showing that the further the pH of the solution is from the pI of the protein, the lower the concentration of denaturant required to induce unfolding (Thomson et al., 1989; Timasheff and Xie, 2003). Timasheff (2003) showed that lysozyme (pI 11.3) remains structurally stable up to 9 M urea at pH 7.0, while denaturing at 3.7 M in solutions at pH 2.0. A similar pattern is observed with ribonuclease A, with denaturation occurring at 1.8 M Gdm HCl at pH 3.0, and at 3.0 M Gdm HCl at pH 7.0 (Pace et al, 1989). In both these examples, an increase in the net charge of the protein leads to an increased structural instability.

1.3 Protein aggregation: mechanism, prevention and prediction

The aggregation of polypeptides leads to the formation of inclusion bodies, amyloids and amorphous aggregates, both in vivo and in vitro. These structures are formed by the self-association of denatured, partially folded or natively folded proteins to form either soluble or insoluble higher-order oligomers (Finke et al., 2000; Jahn and Radford, 2008). The formation of aggregates can occur through several possible pathways, which are influenced by the sequence, structure and aggregation propensity of the protein. However, a simplified pathway of aggregation begins with the association of native or partially-folded proteins through attractive protein-protein interactions to form dimer. This is

known as the nucleation step, and is often rate limiting. The formation of dimers is often preceded by structural changes within one or both of the dimer-forming proteins, as partial-unfolding or secondary structure changes have been shown to increase the strength of attractive interactions between proteins and also to decrease the likelihood of dissociation. The pathways through which the aggregate grows from a dimer into large soluble oligomers or insoluble aggregates are numerous, and are heavily influenced by the solution conditions in which they occur, with pH, ionic strength, denaturant and excipient concentrations playing important roles.

Many aggregates are irreversibly formed, with the protein state unable to return to a monomer form under the solution conditions in which the aggregate was formed. Aggregation can occur in low-salt, native solute conditions, with no inducing factor other than the long term storage of the protein and the protein-protein interactions that occur between the natively and partially unfolded proteins (Shire et al., 2004). These native and partially-unfolded protein-protein interactions can be greatly influenced by small changes in the storage solution composition and the protein structure and sequence. Any alterations to the specific and non-specific protein-protein interactions will lead to changes in the level of short and long term protein aggregation. In a refolding context, aggregation is observed at two major steps, with the formation of the inclusion bodies in E. coli expression systems and during the refolding step where the denatured protein is reconstituted into a native structure (Clark, 2001).

1.3.1 Inclusion Body Formation and structure

Inclusion bodies (IBs) are large (~1 μm), insoluble heterogeneous oligomers of recombinant protein formed within the cell walls of E. coli (Bowden et al., 1991). All inclusion bodies share the same spherical amorphous appearance, regardless of the original primary peptide sequence. Inclusion bodies can represent over 50%of the total volume of a cell and have no intrinsic enzymatic or signalling activity. An early theory proposed by Bowden et al (1991) stated that the environmental stress of producing such high levels of a single protein kinetically induced amorphous unstructured aggregates to form (Bowden et al., 1991). This formation is primarily driven by the exposure of hydrophobic patches by a high population of folding intermediates within the cell and a lack of molecular chaperones which under normal conditions aid in the native folding of the protein, both of which increase stress within the cell. Misfolded or partially folded monomers become energetically trapped aggregates. A high cellular concentration of protein from the recombinant expression leads to the rapid expansion of the aggregate composed entirely of amorphous unstructured protein chains. More recent evidence suggests IBs have a more ordered structure. Ami et al have shown that IBs have significant β sheet content due to the absorbance in the amide I region in inclusion body samples, while also indicating some native α helical structure (Ami et al., 2005). NMR studies by Wang et al (2008) show significant amyloid-like structure within IBs (Wang et al., 2008). Overall, the evidence for both unstructured and secondary structure of protein chains within inclusion bodies

indicates a complex internal structure and formation process of inclusion bodies. Inclusion bodies are composed primarily of a core of cross-linking β sheet amyloid-like structures, with evidence for some proteins containing residual native structure from internalised intermediates. It should be noted that, though the predominant protein contained with inclusion bodies is the overexpressed recombinant protein, there are often a variety of other cellular proteins caught up in the aggregate, alongside other molecular contaminants such as DNA/RNA fragments and lipids.

1.3.2 Aggregation during refolding

During renaturation, the reduction of denaturant concentration causes the denatured protein to collapse upon its hydrophobic core, forming folding intermediates (Buswell and Middelberg, 2003; Finke et al., 2000). Kinetically, these intermediates can either continue to fold into the native conformation of the protein, or form soluble or insoluble aggregates through a series of aggregation stages. The behaviour and stages described below are a generalisation of observed protein aggregation pathways, and does not necessarily represent a universal pathway for aggregate formation as no such pathways exist (Andrews and Roberts, 2007; Roberts, 2007). Much of the literature regarding non-native aggregation is based upon initially folded monomer states, where the partial unfolding of the protein into a more aggregation prone state is a rate limiting step, and there is far less data on systems where aggregation and folding are competing in solutions with high populations of partially-folded and denatured proteins. There have been attempts to elicit the mechanisms and aggregation pathways that occur during protein refolding in regards to intermediate formation. However, much of the work is purely experimental data based on specific proteins, such as Lefebrve and Robinson (2003), and therefore no overall scheme for competing folding and aggregation kinetics has yet been proposed, distinct from the general process described below (Lefebvre and Robinson, 2003; Roberts, 2007).

In the initial self-association and nucleation step of non-crystallising aggregation, unfolded or partially unfolded polypeptides make initial contact and form a non-covalent dimer in solution $R_1 + R_1 = R_2$. This dimer is often transient and reversible in nature, as the protein-protein interactions which lead to the aggregate formation are weak hydrophobic interactions, van der Waals, hydrogen bonding or oppositely charged electrostatic attractive interactions. Therefore, initial dimer formation is potentially one of the rate limiting steps in the aggregation process (Andrews and Roberts, 2007). However, this initial binding does pose an energetic activation barrier in solutions where protein-protein interactions are repulsive. An important distinction between initially-native and denatured protein aggregation is that the formation of a dimer between proteins in a native-promoting solution require partial denaturation leading to aggregation-prone intermediates. During refolding, however, the polypeptide chain begins partially-folded and denatured and therefore begins as an intermediate aggregation-

prone structure. Therefore, the energetic activation barrier to the formation of dimers in non-refolding conditions is no longer present in refolding solutions and therefore the rate of dimer formation is increased. The rate of R_2 formation is proportional to the concentration of protein, as the relative probability of protein-protein interactions occurring is higher. Concentration therefore sets an important parameter for refolding, as concentration is carefully controlled to ensure low initial aggregation. Low final concentrations of <1 mg / mL are common in refolding methods, as it has been shown that there is a significant decrease in yield at final concentrations higher than ~1mg / mL.

In the propagation stage of aggregation, reversible dimers may undergo further misfolding or rearrangement of their structure, and bind with other dimers to form the nucleus of an irreversible aggregate. These irreversible aggregates are strengthened by more stable interactions between the polypeptide chains, such as regular hydrogen bonding and disulphide bonds, and gain a more regular structure through structural rearrangements, a precursor for β sheet amyloid-like polymerisation in the next stage. This irreversible aggregation stage is a potential rate limiting step, and is less controllable than the nucleation stage. Whether the nucleation or aggregation stages are the rate limiting step in the formation of insoluble aggregates is protein-specific.

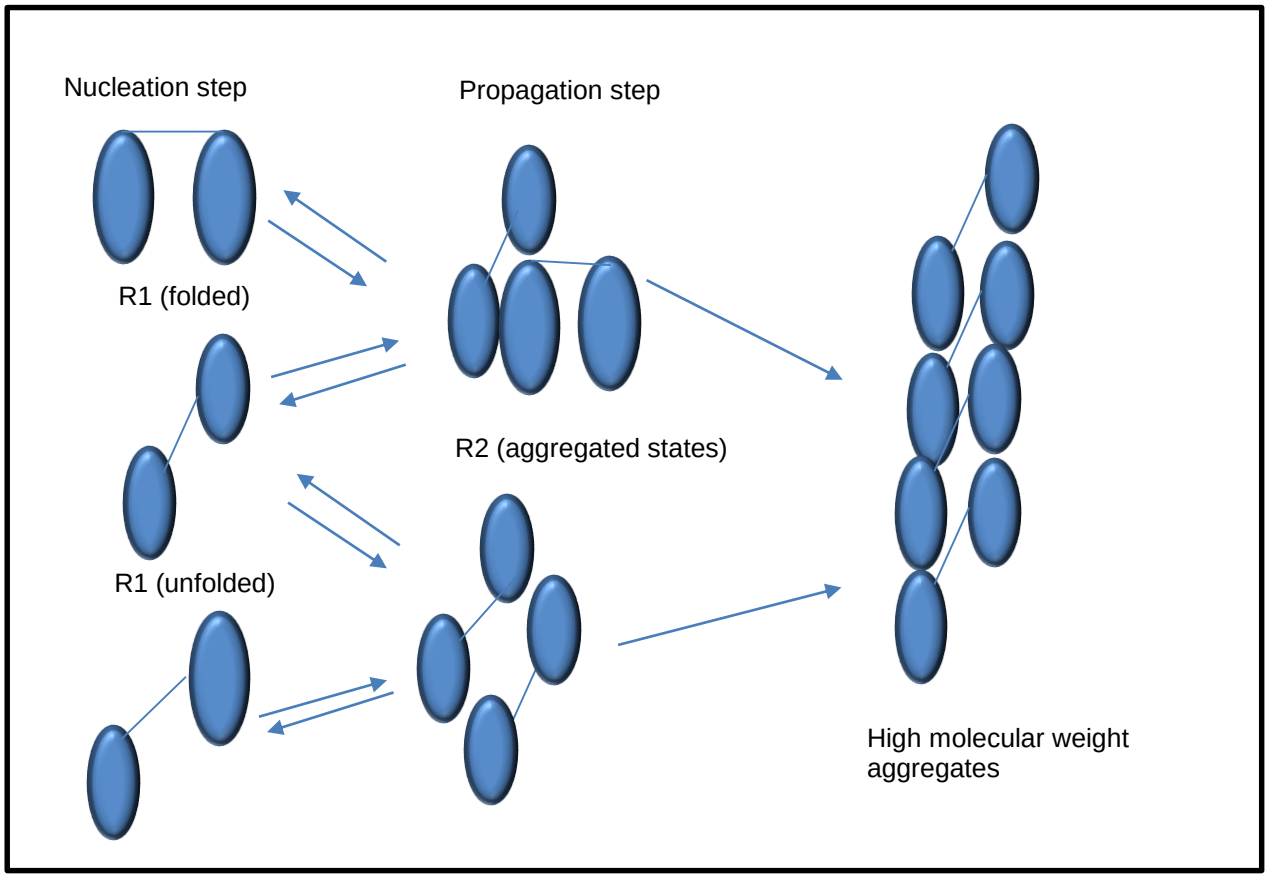


Figure 1.5. Diagram of the potential aggregation pathways during refolding, with proteins aggregating from both unfolded and folded states.

1.3.3 Aggregation and the second virial coefficient (B_{22})

Though long term aggregation, as described above, can lead to regular secondary structure, initial nucleation arises from non-specific attractive association. The likelihood of association can be evaluated by the osmotic virial coefficient (B_{22}). As the B_{22} represents averaged interactions, it has been shown to be a reliable indicator as the propensity of a protein to aggregate, crystallise or remain soluble in native-favouring buffer conditions (George and Wilson, 1994; Zhang and Liu, 2003). Previous evidence has shown that the value of B_{22} of proteins in denaturing conditions is linked to the yield obtained through refolding (Ho and Middelberg 2004, Ho et al. 2003). It is also theorised that aggregation during refolding can be linked to the B_{22} of the protein under refolding buffer conditions, and thus B_{22} can be used as a predictive indicator for how much native protein will remain soluble after refolding has occurred. Knowing how "refoldable" a protein is under a variety of buffer conditions is useful for bioprocessing of inclusion body proteins, as it allows for optimal buffer selection and fine tuning of the process. Middelburg et al showed that large positive values of B_{22} at high denaturant concentrations correlated with a higher refold yield for a selection of inclusion body proteins, while negative B_{22} values correlated with a high propensity for aggregation. Addition of known additives such as arginine and sucrose also has the effect of increasing the value of B_{22} which coincides with an increased final yield upon refolding. Understanding the link between B_{22} and refolding yield is crucial for the optimisation of refolding buffers and understanding the forces involved in the suppression of aggregation.

1.3.4 Prediction of aggregation from primary sequence

There is much interest in being able to predict protein aggregation propensity; more specifically, which regions within a protein sequence are most likely to aggregate. This interest has significance in medicine where aggregation-prone proteins in physiology have been found to be the cause of an increasing number of diseases. In the biotechnology industry, there is also a need to predict aggregation during storage and purification. It has been shown that certain regions within structured and unstructured protein sequences, known as "hotspots", are more likely to lead to aggregation. Hydrophobic patches made up of 3 or more non-polar amino acids, alongside low net charge and a low propensity to form α helixes and a high propensity to form β sheets are the major contributing factors that lead to amyloid fibril formation (Chiti and Dobson, 2006; Pawar et al., 2005). These simple rules have been used to create predictive software and algorithms that predict aggregation within proteins, such as AGGRESCAN (Conchillo-Sole et al., 2007; Tartaglia et al., 2008; Tartaglia and Vendruscolo, 2008).

However, these predictive systems are designed primarily for amyloid-forming protein aggregation, rather than initially-denatured refolding aggregation prediction. Further problems include that the location and structural relevance of the “hotspots” are not taken into account when predicting aggregation. Hydrophobic patches are common in the interior of stable proteins, and thus prediction of aggregation based on these hotspots may be useless. There are currently several refolding pathway and kinetic schemes for refolding (Buswell and Middelberg, 2003; Dong et al., 2004), however they do not take into account the primary sequence or biophysical information, simply describing 1st order folding vs 3rd order aggregation. Therefore, there is little in the way of describing clear predictors of protein yield from refolding based solely upon primary sequence. However, it could be theorised that predictions could be made using similar predictors: overall hydrophobicity, net charge and “hotspot” hydrophobic patches. A major obstacle is also the varying solution conditions which are common in the refolding process. The variation of ionic strength, pH, denaturant and appropriate refolding additives are all key factors in the aggregation propensity of a protein during refolding and a predictive model that takes into account all these factors does not yet exist.

1.4: Model proteins

For studying the effect of the solute environment on protein-protein interactions and aggregation, model proteins must be chosen. The criteria for choosing the model proteins are that they are structurally stable enough to remain soluble in a wide range of solute environments to allow a range of experimental measurements to be performed on the samples, to be readily available in sufficient quantities to allow for a large number of experiments to be performed, and preferably to have been discussed within the literature. Using these criteria, the proteins hen egg-white lysozyme, bovine pancreatic ribonuclease A and human preproinsulin were chosen.

Lysozyme is a 14.4 kDa globular protein, the structure, biochemistry and thermodynamics of which has been widely studied (Bye and Falconer, 2013; Hedoux et al., 2010; Hua et al., 2008; Liu et al., 2005). Lysozyme is also widely available at high purity and low cost. The protein contains two domains separated by a cleft, with both alpha helical and beta sheet secondary structure features and exhibits a two-stage unfolding mechanism with no folding intermediates between the folded and unfolded states (Uversky, 1993; Weiss et al., 2000). Ribonuclease A is a 13.7 kDa protein which is one of the most widely studied proteins in structural biology, the most famous which is the original work by Anfinsen et al on the early understanding of the kinetics of the folding of a polypeptide chain (Anfinsen, 1973; Anfinsen et al., 1961). Ribonuclease A consists of a single globular domain containing both alpha helices and beta sheet secondary structures. The folding and unfolding pathway of ribonuclease A has been heavily studied and has been shown to unfold in a two-stage pathway (Neira and Rico, 1997). Both lysozyme and ribonuclease A are positively charged at low pH,

due the their high pI values. Preproinsulin is a globular protein consisting of 110 amino acids, produced in the pancreas as a precursor of insulin (Oh et al., 1998). Preproinsulin is widely manufactured in modern biotechnology companies to be processed and sold as insulin for human consumption. Communications with the project's sponsors Fujifilm Diosynth have revealed that prepro-LysPro-insulin, a point-mutation mutant used by Fujifilm Diosynth for its increased structural stability, forms inclusion bodies when expressed in pAVEway expression system within e. Coli cells, the protein forms inclusion bodies. These inclusion bodies are processed by urea denaturation and dilution refolding into soluble, monomer preproinsulin. Therefore, preproinsulin provides an industrially relevant globular protein that is negatively charged at high pH due to its pI of 5.0.

1.5: Protein refolding methods

Once an inclusion body sample has been solubilising in a denaturing environment, the proteins must then be refolded to their native conformation with the removal of denaturant into conditions which favour native folding. This can be performed in several ways, each having merits and drawbacks.

1.5.1 Dilution refolding

One of the most common methods of refolding, the dilution method involves the addition of large volumes of optimal refolding buffer to a solution of concentrated denatured protein, thus diluting the denaturant to a non-denaturing concentration (Clark, 2001; Raman et al., 1996; Rudolph and Lilie, 1996; Vallejo and Rinas, 2004). This process can be done at several speeds. Rapid dilution is the addition of the full final volume of refolding buffer in one step, reducing the effective denaturant concentration from 4 M to 8 M to ~ 0.1 to 0.5 M. This leads to rapid protein refolding and has the advantage of being a rapid and simple method, while avoiding extended incubation of the protein solution at intermediate denaturant concentration, which can lead to promotion of aggregation.

Stepwise dilution involves the addition of refold buffer over several steps (Vallejo and Rinas, 2004; Winter et al., 2002). The solution is initially diluted to a moderate concentration of denaturant between 2 to 3 M. At this moderate denaturant concentration, protein will form a molten globule state. A change from reducing to oxidising solution conditions allows for the formation and reshuffling of disulphide bonds within the structure of the protein. The denaturant is then further diluted out by slow addition of the remaining refolding buffer over the course of several hours, to a final denaturant

concentration of between 0.1 and 0.5 M. This method is preferable if it has been shown that the protein of interest does not have a propensity to aggregate at intermediate denaturant concentrations. The use of Gdm HCl as the denaturant in this method is undesirable, as between 1 and 3 M Gdm HCl, the B_{22} of the denatured protein becomes negative in some cases. This has been shown to promote aggregation as attractive protein-protein interactions during refolding can lead to nucleation of aggregates, especially with higher protein concentration. This technique, though effective for some proteins, is impractical and not in use on an industrial scale.

The rapid dilution method of renaturation is common in both industry and academia, as the method is relatively simple to achieve and can produce high yields with the correct selection of refolding buffer. However, because aggregation follows higher order kinetics than folding, the solution must be diluted so that the final concentration of protein is relatively low, usually between 50-500mg/L. On an industrial scale, to be able to achieve large quantities of folded protein requires huge volumes and large quantities of the required solutes, leading to big waste and high costs.

1.5.2 Dialysis refolding

The concentration of denaturant can be reduced without significant changes to the volume by dialysing against a denaturant-free buffer, thus slowly introducing renaturation conditions (Basu et al., 2011). The dialysis membrane is selected to be permeable to molecules with molecular weights less than that of the target protein. The dialysate must be an optimal buffer chosen to promote native folding, similar to that used in dilution refolding. The buffer must ensure that protein-protein interactions remain repulsive and correct folding is promoted.

During dialysis refolding, the initial protein concentration is often low (<1 mg/ mL), as higher concentrations will lead to an increased molecular crowding effect and therefore elevated aggregation and precipitation, as the denaturant concentration is reduced (Tsumoto et al., 2010). Dialysis refolding suffers a similar problem to step-wise dilution; although both promote the formation of molten globule structures at intermediate denaturant concentration and thus correct disulphide formation, repulsive protein-protein interactions can be reduced at moderate denaturant concentrations most clearly characterised in Gdm HCl solutions. Due to this reduced repulsion, the aggregation propensity of the protein at moderate denaturant concentration is significantly increased. To achieve high yields with dialysis refolding, strong repulsive protein-protein interactions must be present, and therefore co-solvent selection could be guided by B_{22} values of proteins measured under native and denaturing/renaturing conditions. For large scale refolding, dialysis is an unfavourable choice, due to the relatively low potential yields and, large volumes of dialysis equipment being required.

1.5.3 On-column refolding

Refolding proteins in a column is an attractive option for industrial protein refolding, as the method can incorporate two steps at once: protein refolding and purification. Two general methods of column refolding can be used to refold denatured proteins; size exclusion chromatography (SEC) and matrix affinity (ion exchange/reverse phase) chromatography.

In SEC, a mixture of protein species is run through a gel filtration column. The protein species are separated according to their size; smaller proteins move at a slower rate as they partition into the pores of the gel beads. This process is adapted for refolding by equilibrating the gel column with refolding buffer devoid of denaturant and containing oxidising agents (Werner et al., 1994). A concentrated (up to ~ 80 mg / mL) denatured protein solution is run through the gel filtration column at a slow flow rate. The column reduces the effective local denaturant concentration of the protein environment as the denaturant moves more slowly through the stationary phase than the protein. Folding intermediates and aggregates formed by self-association and inter-molecular cysteine bond formation are separated from the correctly folded monomer by the gel filtration mechanism. This removes the aggregates from the refolding environment immediately and thus increases the probability of first order native folding by reducing competition from higher order aggregation. This process has the additional advantage of being a rough separation technique (Middelberg, 2002). Though the final solution may still contain non-target proteins, the solution will have few impurities and thus further purification will be made easier. Within an industrial context, any reduction in downstream processing is cost-saving. The major disadvantage is the low throughput as only small volumes (1-2 mL of the sample can be processed at any one time. In addition, the higher the sample concentration, the lower the final yield of protein is achieved due to increased molecular crowding within the column.

Ion-exchange (IEx) and reverse-phase (RP) chromatography are the methods of binding a sample of denatured protein to a resin matrix column, followed by reduction of the denaturant concentration by a gradient program (Middelberg 2002). By immobilising the protein on the column while the denaturant is removed, aggregation can be prevented as the protein is isolated through strong protein-matrix binding and thus preventing protein collisions that lead to aggregate-formation, and therefore favouring native refolding. Upon reduction of the denaturant concentration, the protein is eluted off the column in IEx with a high gradient of NaCl and for RP-HPLC by acetonitrile. Freydell et al (2010) undertook a fairly comprehensive analysis of the IEx process and found several problems (Freydell et al., 2010). Firstly, under denaturing and reducing conditions, the exposed hydrophobic residues of the denatured protein may interact with bound proteins on the column matrix, which can prevent some of the proteins from being eluted with the high ionic strength buffer. In this case, the protein must be eluted from the matrix with a mild denaturant. Second, the binding capacity of the column is greatly

reduced for denatured proteins, owing to an increased hydrodynamic size, leading to a decreased accessible binding surface. Finally, for denatured/reduced lysozyme, loading above 1 mg of protein onto the column significantly increases the level of aggregation upon refolding. This process may be used for small scale refolding of low protein concentrations, but for industrial scale purposes, the technology is developed enough that high yields can be achieved at high protein concentrations to make the approach a viable refolding method.

1.6: Methods for the analysis of refolding yield and protein structure

Once the components of a refolded solution have been separated, the folded and functional state of each fraction require analysed. This is necessary to determine the fraction of aggregated or native state, the refolded yield and the quality control of the final desired product. This can be achieved through a variety of assays and spectroscopic methods, depending on the specific protein in question. Below are the outlines of some of the techniques commonly used in industry and academia to elicit the refolding yield and the structure of the purified monomer protein, though they are not employed within the work described herein.

1.6.1 Enzyme activity and protein-specific assays

Many proteins have known enzymatic or binding properties specific to their function (Gillis et al., 1978; Shugar, 1952). A refolded solution where 100% of the target protein within the inclusion body is folded into a native conformation will have the same specific activity as a stock solution of the same protein concentration. Activity assays are employed before separation techniques are used to isolate the different components of the refolded solution. Further analysis of the separated components is essential to verify the correct sample retained as the final product.

Activity assays are generally reliable but are only applicable to proteins with known activity and assays. In the bioprocessing industry, many pharmacologically relevant proteins, such as antibodies, have no activity which can be readily assayed, or a reliable assay takes up too much resources or time. Other biophysical techniques must therefore be employed to verify the overall refolded yield and native structure.

1.6.2 Reverse-phase high pressure liquid chromatography (RP-HPLC)

Reverse phase chromatography is a method of separating out a mixture of proteins based on hydrophobicity. A solution containing a mixture of proteins is run over a column containing a high density of hydrophobic hydrocarbon chains, ranging from 4 to 18 carbons long. The proteins bind to the chains with an affinity based on hydrophobicity. The proteins are then eluted off the column with a gradient of acetonitrile concentration. The proteins are eluted off the column depending on the strength of their binding. This procedure allows for sharp separation of the same protein in different structural conformations and disulphide bond arrangements as each has a distinctive hydrophobic binding capacity (Huang et al., 2007). Therefore, RP-HPLC is commonly used to analyse refolded solution that may contain a mixture of native, mis-folded and soluble aggregated proteins. However, care must be taken before declaring which fraction of the RP-HPLC is the soluble native protein, as no structural detail is given by the spectra provided. Further structural analysis of the fractions must be undertaken to ensure correct determination of the tertiary structure of the proteins.

1.6.3 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy is a method of determining the secondary structure content of a protein by observing the differential absorption of left or right rotating components of polarised light. The differential absorption arises from the chromophores within a polypeptide chain in an asymmetrical environment (Greenfie.N and Fasman, 1969; Kelly et al., 2005; Louis-Jeune et al., 2012). The main optically active chromophores in proteins are contained in the peptide bond. The conformation of this bond produces a distinctive spectrum in the far-UV (180-240 nm). The secondary structures of proteins produce distinctly different spectra in this region, produced by two electronic transitions to lower energy levels by the amide bonds (Woody and Koslowski, 2002). These are an n-π* transition which is responsible for signals within the 215-230 nm region, and a π-π* transition which produces signals within the 185-200 nm region. The n-π* transition produces negative signals at 216-218 nm for β sheet structures and at 222 nm for α helix structures. The π-π* transition produces positive signal at 198 nm for β sheets, and for α helices produces a positive signal at 190nm and a negative signal at 208nm (see figure 2.3 for an illustration containing spectras corresponding to the secondary structure elements. CD spectroscopy is a useful technique to determine whether the final product of a refolding experiment contains the expected structural elements of the target protein. However, if the sample contains significant populations of protein molecules in different structural conformations, including aggregated or misfolded, CD spectroscopy can provide little accurate data regarding the composition of the sample, simply information of the bulk structural composition of the sample. Therefore it is not applicable to measuring the final refolding yield of a refolding experiment.

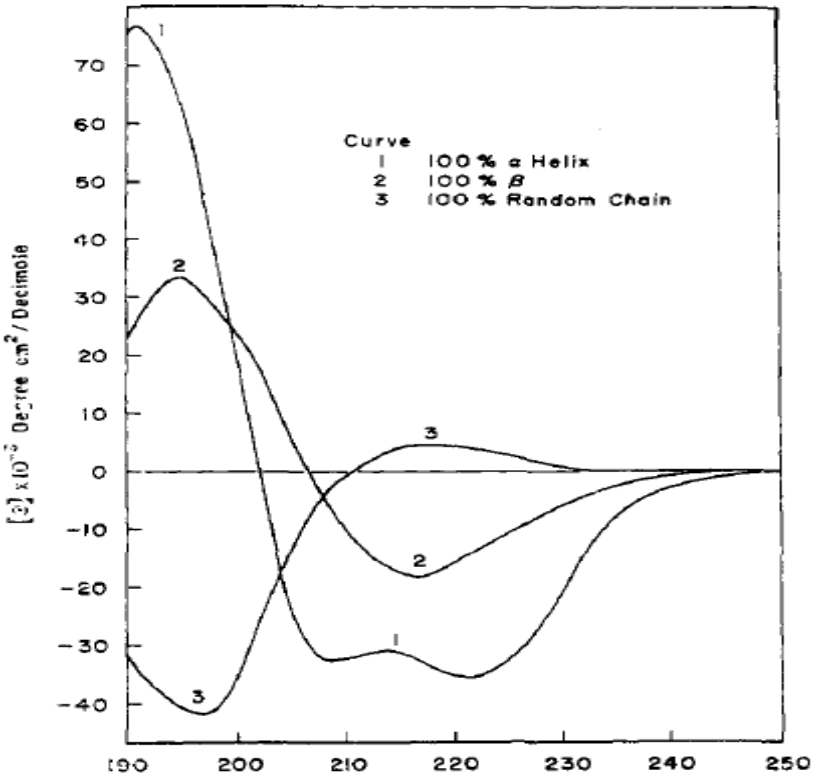


Figure 1.6 The CD spectra of pure α helixes (1) , β sheets (2) and "random coil" or "unstructured (3) proteins. Note the location of peaks described in the text (Greenfie.N and Fasman, 1969).

1.6.4 2D Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectroscopy is a powerful tool to investigate subtle differences in the conformational state of a polypeptide chain, based on the change in orientation of a nucleus under a strong magnetic field. A nucleus has a magnetic dipole moment, created by the quantum spinning charge of protons within the nucleus. Isotopes with a ½ spin that are relevant to protein chemistry are ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P. One dimensional NMR simply measures the chemical shifts within ^{1}H atoms, and is not often used for detailed protein studies. Two dimensional NMR is used to measure two ½ spin isotopes within the macromolecule. In protein chemistry the most commonly used isotopes are ^{1}H and ^{15}N, which provide a spectrum for the amide backbone of the polypeptide, where a signal is generated for each ^{15}N-^{1}H pairing. ^{15}N is found in relatively low abundance thus ^{15}N-labelled protein must be expressed using E. coli grown in a specific growth medium containing only ^{15}N nutrients.

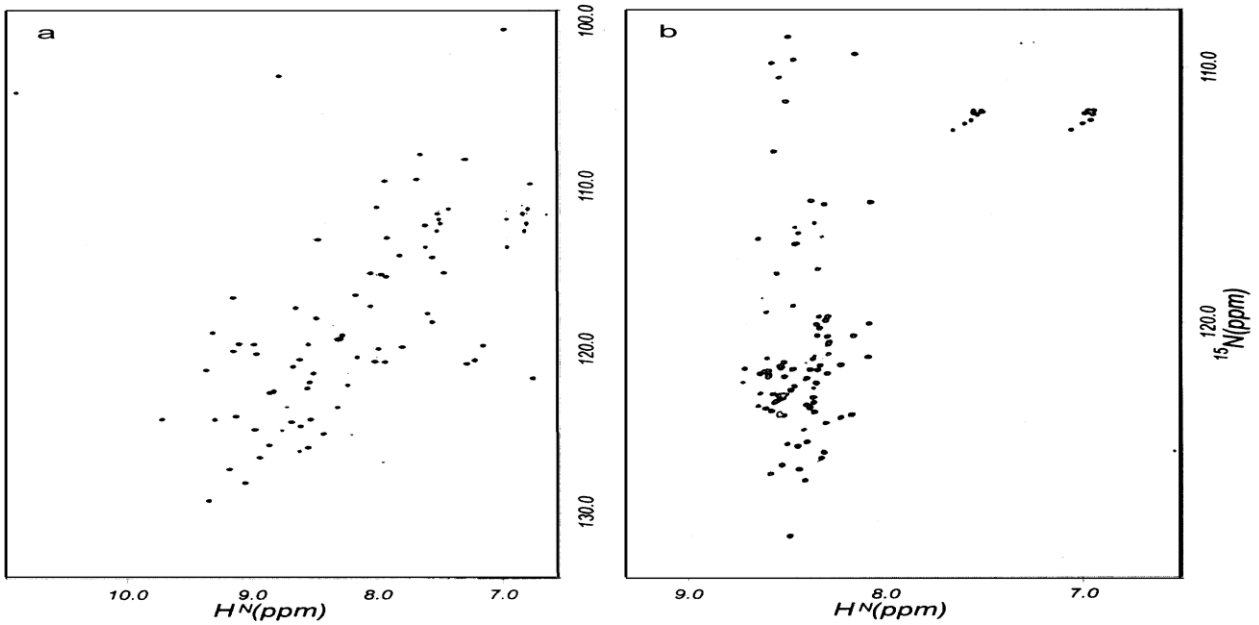


Figure 1.7. Comparative HSQC 1H – 15N spectra (Chatterjee et al. 2005). Both spectra are of a ^{15}N-labelled SUMO protein. On the left (a) is the native folded state, while on the right (b) the protein is suspended in 8M urea, and is in a denatured form. Of note is the closer clustering of the peaks within the 8.5ppm region (b), while (a) shows a far broader range of H signal, indicative of distinctive secondary structure.

The 2D NMR spectra of ^{15}N and ^{1}H, corresponds to the amide (N-H) groups on the peptide is known as a heteronuclear single quantum coherence (HSQC) spectra. Each point on the spectra represents one N-H bond. Differences in chemical shift occur due to the local magnetic environment, which in turn is determined by the structure of the protein. Chatterjee et al provide an example of how an experimentally derived spectrum can be used to analyse the folded state of a protein by comparing against a previously known NMR spectrum as a comparison, shown in figure 2.4 (Chatterjee et al., 2005). Using this analytical technique the level of secondary structure within a protein sample can be determined and a comparison made against a previously obtained HSQC spectrum of the folded protein. However, though this technique can provide information on the structure of a protein in solution, the stipulation that the protein be grown in a specialised media containing only ^{15}N sources of nitrogen is prohibitively expensive and impractical to large scale protein production, and thus is rarely employed in the analysis of refolding yields or final products.

1.7: Protein-protein interaction characterisation techniques

1.7.1 Static Light scattering

Static light scattering can be used to characterise several different aspects of a protein solution. The scattered light measured at a known protein concentration can be used to calculate the molecular weight of the protein. The B_{22} of a protein can be determined by performing a series of static light scattering measurements of different concentrations of the protein. Upon irradiation by light in the visible spectrum of a macromolecules in solution, the level of Rayleigh scattered light (R_θ, where θ is the scattering angle) can be measured. This can provide information regarding the molecular weight and protein-protein interactions in the form of the B_{22} of the macromolecule. Upon irradiation by incident light, electrons scatters light in all directions. The intensity of light scattered is proportional to the molecular mass of the macromolecule (as the molecule is assumed to be spherical in solution) and the protein concentration of the solution. In a solution of identical particles significantly smaller than the length of the incident light, the total scattered light is equal to the sum of all the particles, minus the background scattering of the solution. Therefore, R_θ can be related to the protein concentration and molecular mass. The interactions between the proteins can be measured by the differences in the intensity of light scattered by proteins at different protein concentrations. Attractive interactions lead to a relative increase in the intensity of light scattered with increasing concentration. The theory behind these measurements is discussed in a later chapter.

1.7.2 Dynamic light scattering, hydrodynamic radius and mutual diffusivity

Brownian motion describes the random diffusion of molecules through a solution due to interactions with solvent molecules; larger molecules move slower. The hydrodynamic radius (R_h) of the molecule can be calculated with Stokes-Einstein equation. The R_h is not the size of the protein, but the overall hydrodynamic structure that is diffusing through the solution. This therefore represents not just the protein, but also the ions and water molecules bound to the protein surface. Dynamic light scattering provides the means to calculate the hydrodynamic radius of a protein in solution by measuring the changes in the intensity of laser light scattered by a protein solution over a short length of time. These changes are proportional to the diffusion coefficient (D_m) of the protein and, if the viscosity of the solution is known, the R_h can be calculated. D_m is defined as the speed in which a molecule diffuses in solution relative to the other molecules in solution. The extrapolation to D_m to zero protein concentration gives infinite-dilution diffusivity, D_o, for the diffusion speed of a molecule in an infinitely dilute solution. Both D_m and R_h are useful in understanding the molecular and hydrodynamic interactions experienced by proteins in solution. The theory behind these measurements is discussed in a later chapter.

Chapter 2: Materials, methods and theory

2.1 Materials

lyophilised lysozyme powder, lyophilised ribonuclease A, hexylene glycol, sodium acetate (salt), sodium chloride (salt), sodium phosphate dibasic (salt), sodium phosphate basic (salt), ethanolamine (liquid), cysteine (crystal), urea (crystal), sucrose (powder) and sodium hydroxide (salt) were purchased from Sigma-Aldrich (UK). Syringe-top filters (0.22µm), Gdm HCl (salt) and Arg HCl (salt) were purchased from Milipore (Darmstandt, Germany). Tris base (salt) was purchased from Formedium (UK). Resource Q columns, Resource S columns and Superdex 200 gel filtration column were purchased from GE Healthcare Life Sciences (UK). Glycerol was purchased from VWR chemicals (UK). DTT (powder) was purchased from Alfa Aesar (UK). Hydrochloric acid and acetic acid were purchased from Fischer Chemicals (UK). 0.02 µm syringe-top filters were purchased from Whatman (GE Healthcare, UK). Pure water was obtained using a Mili-Q filtration system (Milipore, Darmstandt, Germany).

2.2 Methodology

2.2.1 Static and dynamic light scattering measurements

Protein was solubilised in 50 mM sodium acetate buffer (pH 4.5) for lysozyme and ribonuclease A to a concentration of 10 g/L, or 20 mM Tris-HCl (pH 9.0) for preproinsulin to a concentration of 5 g/L, and filtered with a 0.22 µm syringe-top filter. Samples of the protein solution were diluted to produce 10 protein concentrations between 1 g/L and 10 g/L for lysozyme and ribonuclease A, and 0.5 g/L and 5 g/L for preproinsulin. Stock urea/NaCl solutions were prepared from the same buffers and filtered in the same manner. The protein and urea/NaCl solutions were injected simultaneously to a 1:1 ratio through 0.02 µm syringe-top filters using syringe pumps (Harvard apparatus, UK) at a total flow rate of 0.3 mL/min and mixed using a three-way tubing confluence into a Dawn Eos light scattering instrument (Wyatt, CA, USA), UV-975 UV detector (Jasco, UK), and RI detector Optilab T-rEX (Wyatt). For dynamic light scattering measurements, the samples were instead injected into Dawn Treos (Wyatt, CA, USA) and UV detector. Protein concentration was calculated from UV absorbance at 280 nm (extinction coefficients for each proteins were as follows; lysozyme, 2.64 M^{-1} cm^{-1}, ribonuclease A, 1.58 M^{-1} cm^{-1}, preproinsulin, 1.44 M^{-1} cm^{-1}) Data was collected using the Astra 6 software, and analysed using Microsoft Excel software. Additional measurements of high concentrations of protein solutions were taken by diluting a sample of a protein solution by a factor of 1:50 with the appropriate buffer and measuring the UV absorbance of the sample at 280 nm using the UV-1600PC spectrophotometer (VWR, UK).

2.2.2 Laser Doppler electrophoresis: varying urea concentration

Lyophilised lysozyme powder was solubilised into 50 mM acetate (pH 4.5), 10 mM phosphate (pH 6.4) or 20 mM Tris-HCl (pH 9.0) to a concentration of 10 g/L. Each sample was dialysed for 1 hour against the appropriate buffer at a ratio of 1:50 to remove impurities. Stock 8 M urea solutions were created for each pH from the dialysis solutions. The pH of each sample was ascertained prior to measurement. The urea concentration was altered by addition of varying volumes of the 8 M urea stock to samples of the protein solution to achieve the desired urea concentration.

2.2.3 Laser Doppler electrophoresis: varying pH

Lyophilised lysozyme powder was solubilised in pure water to a concentration of 10 g/L and dialysed into 500 mL of pure water to remove impurities. The pH of the solution was adjusted by addition dropwise of either 0.01 M NaOH or 0.01 M HCl while monitoring pH with a pH probe.

2.2.4 Laser Doppler electrophoresis: measuring the zeta potential

The zeta potential of the protein solution was measured using the ZetaPlus Zeta Potential Analyser (Brookhaven Instruments Corporation, USA). A diode was immersed in 1.5 mL of the solution, and 10 readings were taken. Readings of the zeta potential and mobility of the protein were averaged to produce the final values.

2.2.5 SEC-MALS

Size exclusion chromatography with online multi-angle light scattering (SEC-MALS) was performed using a superdex 200 gel filtration column connected to a DAWN Eos light scattering instrument (Wyatt Technology, CA, USA), UV-975 UV detector (Jasco, UK), and Optilab T-rEX refractive index detector (Wyatt Technology, CA, USA). The column was equilibrated with 50 mM acetate buffer (pH 4.5), which had been filtered with a 0.22 µm syringe-top filter. 500 µL of sample protein was injected onto the column using a Rheodyne model 7010 injector at 0.5 mL/min. Data was collected using ASTRA 6 software and analysed using Microsoft Excel software.

2.2.6 Ion exchange chromatography

Ion exchange chromatography (IEx) was performed using either a 1 mL ResourceQ (for use with lysozyme) or 1 mL ResourceS (for use with preproinsulin) anion/cation exchange column. All samples were dialysed against either 50 mM acetate (Resource Q) or 20 mM Tris-HCl pH 7.5, which was used to equilibrate the column. 10 mL of the sample protein solution was injected over the IEx column at a

flow rate of 1.0 mL/min, followed by a wash of 5 column volumes (CV). The bound protein was eluted from the column using a continuous gradient of NaCl between 0 and 1 M in the appropriate buffer, up to a maximum concentration of 1 M over 40 CV. The samples were collected in a 96 well plate at a volume of 1 mL per sample.

2.2.7 Dilution refolding

Lyphilised lysozyme was solubilised in denaturant, either 8 M urea or 8 M Gdm HCl, 10 mM DTT and 50 mM acetate buffer (pH 4.5) at a concentration of 10, 20 or 30 g/L and incubated at room temperature for 60 minutes. The sample is then rapidly diluted to a ratio of 1:30 with the refolding buffer and incubated at room temperature for 18 hours. Refolded samples were filtered with a 0.22 µm syringe filter and the protein concentration measured before and after filtering through measurement of the UV absorbance 280 nm using the UV-1600PC spectrophotometer (VWR, UK). The aggregated state of the filtered protein sample was measured through SEC-MALS, to determine total refolded yield. Turbidity of the pre-filtered solution was determined through measurement of the absorbance of light at 600 nm using the V-760 spectrophotometer (Jasco, UK).

2.2.8 Dialysis refolding

Lyophilised lysozyme was solubilised in denaturant, either 8 M urea or 8 M Gdm HCl, 10 mM DTT and 50 mM acetate buffer (pH 4.5) at a concentration of 20 g/L and incubated at room temperature for 60 minutes. The sample was transferred to 8000 Da dialysis tubing and immersed in the refolding buffer at a ratio of 1:30 for 18 hours. The refolding buffers all contained 3mM cysteine alongside the desired refolding additives. Upon extraction from the refolding buffers, the samples were filtered with a 0.22 µm syringe-top filter and the protein concentration measured before and after filtering through UV 280 nm absorbance using the UV-1600PC spectrophotometer (VWR, UK). The pre-filtered protein concentration of the solution was calculated by re-solubilising a sample of the refolded protein solution with the denaturing buffer, followed by measurement of the protein concentration of the denatured protein solution. If UV absorbance surpasses 1 AU, the sample was diluted by a factor of 1:50 to determine an accurate reading. The aggregated state of the filtered protein sample was measured through SEC-MALS, to determine total refolded yield.

2.2.9 Optim 1000

The Optim 1000 system (Forte Bio, UK) was used to monitor aggregation and unfolding of proteins between 25°C and 90°C. Lysozyme was solubilised in 50mM acetate buffer (pH 4.5) to produce a 2 g/L solution, which was filtered with a 0.22 µm syringe-top filter. Samples were mixed with stock excipient solutions to produce samples at 1 g/L protein with the desired excipient concentration. The

samples were run in duplicate in the Optim 1000, where they were heated 1°C stepwise, and the fluorescence and SLS at 266nm were monitored to determine the protein unfolding and aggregation respectively. Analysis was performed using the Optim 1000 software.

2.2.10 Preproinsulin inclusion body dilution refolding

The purified inclusion body paste was refolded according to the Fujifilm Diosnyth dilution refolding protocol. 1.0 grams of inclusion body slurry were solubilised with denaturing buffer containing 4.5M urea or 6 M Gdm HCl, 10 mM DTT, 365 mM ethanolamine at pH 10.5. A ratio of 1.0 grams of inclusion body paste to 8.0 mL denaturing buffer was used. Solution was stirred at room temperature 60 minutes. The denatured protein solution was then rapidly diluted by addition of the refolding buffer; 10% hexylene glycol, 20 mM ethanolamine, 288 µM cystamine, used at a ratio of 17.0 ml refolding buffer to 1.0 ml denatured protein solution. Solution was then stirred overnight at room temperature

2.2.11 Purification of preproinsulin from inclusion bodies

The preproinsulin (λ17NproLispro) was produced for biophysical analysis through expression in a BL21 *E. coli* strain controlled by pAVE0111 vector. The inclusion bodies produced were washed with UHQ-water and purified through centrifugation. The inclusion body slurry were solubilised with denaturing buffer containing 4.5M urea, 10 mM DTT, 365 mM ethanolamine (pH 10.5). A ratio of 1.0 grams of inclusion body paste to 8.0 ml denaturing buffer was used. Solution was stirred at room temperature 60 minutes. The denatured protein solution was then rapidly diluted at a ratio of 1:17 by addition of the refolding buffer; 10% hexylene glycol, 20 mM ethanolamine, 288 µM cystamine. Solution was then stirred overnight at room temperature. The pH of the solution was brought down to pH 9.0 using 5 M HCl. Upon acidification, the refold was filtered first through a course depth filter (3M Zeta plus EXT 60ZA05A media), secondly through a 0.2µm filter (Sartorius Sartopore 2). Refold was then passed over an AIEX chromatography column for primary capture of the refolded preproinsulin. The AIEX column was previously equilibrated with sodium borate (pH 9.0), and the protein was stripped from the column with a gradient of 0.1 M NaCl to create a purified protein solution.

2.3 Theory

2.3.1 Measurement of the mobility and zeta potential through laser Doppler electrophoresis

Laser Doppler electrophoresis works through the analysis of the Doppler shift in a beam of light shone through a solution when an electric field is applied to charged particles in said solution. The Doppler shift of the light, Δf, is then related to mobility, u as:

$$u = \frac{\lambda \Delta f}{E sin\theta}$$ Equation (3)

Where λ is the wavelength of light, E is the electrical field between the electrodes, and θ is the scattering angle the light is measured at (in this case, 90 degrees). The zeta potential of the particle can then be calculated from the observed mobility using the Henry equation below:

$$\zeta = \frac{3\eta}{2e_0 e_{rf(\kappa a)}} u$$ Equation (4)

Where ζ is the zeta potential, η is the viscosity, ε_0 is the permittivity of free space, ε_r is the relative dielectric constant of the medium, κ is the inverse screening length, *a* is the radius of the protein and *f* is the Henry function, given as:

$$f(\kappa a) = 1 + \frac{1}{2}\left[1 + \left(\frac{2.5}{\kappa a(1+2\exp(-\kappa a))}\right)\right]$$ Equation (5)

The Henry equation is valid only to values of ζ <50 mV.

2.3.2 DVLO theory

The behaviour and aggregation of proteins in solution can be modelled using Derjaguin-Landau-Verwey-Overbeek (DLVO) theory, a method for modelling the interactions of colloidal bodies in electrolyte solutions as rigid charged spheres. With DLVO theory, interactions between colloids arise from attractive short range van der Waals forces and long range electrostatic repulsion (sometimes referred to as double layer forces) (Curtis and Lue, 2006). DLVO theory is most commonly used in biological systems to predict the behaviour of micelles and vesicles, although the theory can be applied to proteins. In the latter case, however, protein-specific interaction forces, specifically hydrodynamic and solvation forces, also need to be included in the model.

At low ionic strength, the main contribution to colloidal interactions according to DLVO theory is the electrostatic repulsion, which can be indirectly probed through the measurement of the zeta potential

of the charged particle. The zeta potential is the potential difference between the solution ions bound to the particle and the bulk electrolyte solution, and is directly related to the net charge of the particle. Repulsion arises from charge repulsion and osmotic pressure between the double ion layers of particles. The structure of the layer of bound surface ions is defined by the distance from the surface of the particle and the size of the bound ions. The shortest distance an ion can be bound to the surface of the particle is known as the Stern plane, and the Stern layer is the closest layer of oppositely charged bound ions to the particle surface. Beyond this is the more diffuse layer of ions, known as the diffuse layer. The slipping plane forms where the ion density changes to that of the bulk solution. It is at the slipping plane that the zeta potential is measured, though the double layer potential can also be written as the diffuse layer potential. However, for the sake of calculations, the zeta potential and the diffuse layer potential are treated as the same. The zeta potential can be used to calculate the contribution of the electrostatic B_{22} which is the experimentally accessible parameter that describes the interaction between two charged bodies.

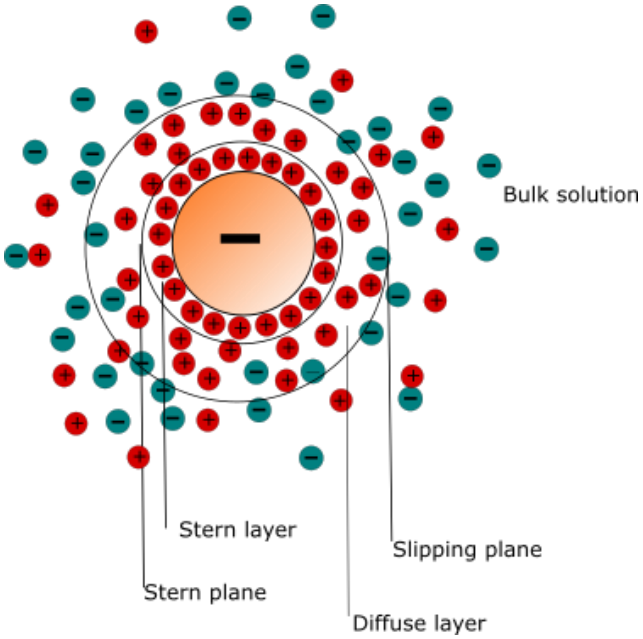


Figure 2.1. The structure of the electrolyte solution surrounding a charged protein in solution. Zeta potential is measured at the slipping plane between the diffuse layer and the bulk solution.

DVLO theory can also account for other interaction forces that can contribute towards attraction or repulsion between particles. The excluded volume is the volume of space a particle cannot move into because of the presence of another identical molecule. This force is based solely upon the particle volume and is independent of other solution forces. A separate force is derived from the variety of short ranged interactions that occur between proteins, including; hydrophobic interactions, hydrogen bonding, van der Waals forces, and salt bridges. However, the extent and strength of these interactions is not well defined, and vary significantly with solution conditions. The Hamaker

dispersion potential is used to define dipole-dipole dispersion interaction forces between two bodies, i.e. van der Waals forces.

Finally, hydration forces between the proteins can also contribute towards protein-protein repulsive interactions. Evidence has been published suggesting that the accumulation of strongly hydrated ions at the surface of the protein at high concentration salt solutions (0.5 M) can lead to a short-range repulsion between proteins, leading to a barrier to protein aggregation (Valle-Delgado et al., 2011; Zhang et al., 2007). This type of short-range repulsion is not predicted in traditional DVLO theory, and is instead seen only with proteins with certain salts, notably NaCl.

Overall, DLVO can be used to account for all solution behaviour and gives the net B_{22} as the sum of electrostatic repulsive forces, excluded volume forces and short-range interaction forces,

$$B_{22} = B_{22,elec} + B_{22,exl} + B_{22,sri} \quad \text{Equation (6)}$$

2.3.3 Calculating double layer force from charge and zeta potential

The determination of the double layer force, W_{el}, requires knowledge of zeta potential of the protein ζ, using equation 4.

$$W_{el} = 4\pi e_0 e_r \frac{a^2}{r} \zeta^2_{dexp[-\kappa(r-2a)]} \quad \text{Equation (7)}$$

Where *r* is the distance from the protein centre. The inverse screening length, κ, is defined as the distance from a charge carrier in solution where the effects of the electrostatic charge is no longer felt, and is related to the ionic strength of the solution, IS, as shown in equation 2.

$$\kappa^2 = \frac{2N_{AV}ISe^2}{e_0 e_r k_b T} \quad \text{Equation (8)}$$

Where k_B is the Boltzmann constant, N_{AV} is Avargardo's number, e is one electronic charge and T is absolute temperature. If the values for constants in the above equation are factored in as K_B = 1.3806 × 10^{-23} J/K, e = 1.6 × 10^{-19} C, e_0 = 8.547 × 10^{-12} C^2/(J-m) and e_r = 78.3, equation 7 is simplified as:

$$\kappa(nm^{-1}) = 3.29\sqrt{IS(M)} \qquad \text{Equation (9)}$$

Therefore, if the ionic strength and zeta potential are known, the double layer force can be calculated. This value is then used with equation 4 to find the contribution of this force to the second virial coefficient.

$$B_{22} = -\frac{1}{2}\int drd\,\Omega_1 d\Omega_2\; x\left\{\exp\left[-\frac{w_2\,(r,\Omega_1,\Omega_2;T,\mu_w,\mu_s)}{k_BT}\right]-1\right\} \qquad \text{Equation (10)}$$

Where r is the distance between the two proteins, Ω_1 and Ω_2 represent the orientations of proteins 1 and 2 respectively, w_2 is the potential of mean force as defined as W_{el}, μ_w and μ_s are the chemical potentials for water and solute respectively.

Alternatively, W_{el} can be calculated using the protein charge, Z_p, using equation 9.

$$\frac{W_{el}}{k_BT} = \frac{Z_p^2\lambda_B}{(1+\kappa a)^2}\frac{\exp[-\kappa(r-2a)]}{r} \qquad \text{Equation (11)}$$

Where λ_B is the Bjerrum length, the distance at which the coulomb interaction between a pair of charged ions is equal to k_BT. At 25°C in water, this value equals 0.7 nm.

2.3.4 Calculating excluded volume contributions

Excluded volume is the inaccessible solution volume in which a molecule cannot move into due to the presence of other identical molecules, in this case, the volume of a protein. The contribution of excluded volume to the second virial coefficient is relatively small compared to electrostatic repulsion, but its calculation is still important.

$$B_{22ex} = \frac{\left(\frac{16}{3}\right)\pi r^3 N_{AV}}{M_w^2}$$ Equation (12)

Where *r* is the radius of the protein, and M_w is the molecular weight of the protein.

2.3.5 Measurement of second virial coefficient through static light scattering

Though individual contributions to the second virial coefficient can be calculated as described above, the net protein-protein interactions are measured through static light scattering. The molecular mass and second virial coefficient were calculated using the Raleigh scattering intensity at 90°, R_θ. The molecular mass is derived from equation 13:

$$R_\theta = KcM$$ Equation (13)

Where *c* is concentration in g/L, *M* is molecular mass and *K* is a constant defined in equation 14.

$$K = \frac{2\pi^2 n_0^2 (\frac{dn}{dc})^2}{\lambda^0 N_{AV}^4}$$ Equation (14)

Where λ^o is the incident wavelength, N_A is Avogadro's number, n_o is the refractive index of the solvent and *(dn/dc)* is the refractive index increment of the solvent-solute complex with respect to concentration (g/ml). This equation allows for the determination of the molecular weight of a molecule, at a known concentration and refractive index increment (for proteins, this is usually 0.185). The refractive index of water (1.33) was used for calculations when the refractive index of the solution could not be calculated.

The second virial coefficient is determined from equation 15.

$$\frac{Kc}{R_\theta} = 1/M(1 + 2B_{22}Mc + \cdots)$$ Equation (15)

The virial expansion of which is:

$$\frac{Kc}{R_\Theta} = \frac{1}{MP_{(\Theta)}} + 2\,B_{22}C \qquad \text{Equation (16)}$$

Where B_{22} is the second virial coefficient and $P(\theta)$ is the partial scattering function, which is dependent upon the particle shape. P is equal to 1 for particles less than 1/20th of the size of the wavelength used, in this case 658 nm, i.e. 35 nm. The proteins being studied have a diameter of between 2-5 nm, so P is equal to 1.

A graphical plot of KC/R_θ against concentration (known as a Zimm plot) will give the intercept at c = 0 as 1/M and half the gradient will give the value of B_{22}. The R_θ can be derived from the measured light scattering intensity of the same, I_θ, with equation 17:

$$R_\Theta = \left(I_\Theta - I_{\Theta,solvent}\right)r^2/I_0V \qquad \text{Equation (17)}$$

Where $I_{\theta,solvent}$ is the scattering intensity of the solvent, r is the distance between the sample and the detector, I_o is the intensity of the incident beam and V is the volume of the medium. Equation 16 compensates for the various geometric factors (i.e. R) which may affect the intensity of the scattered light that are not related to the sample solution. Adjustment of this equation gives:

$$R_{90} = A_{cscc}\left(\frac{V_{90} - V_{90,dark}}{V_{laser} - V_{laser,dark}}\right) \qquad \text{Equation (18)}$$

Where A_{cscc} is the configuration specific calibration constant, a constant specific to each detector, V_{90} is the scattering intensity, $V_{90,dark}$ is the dark offset, V_{laser} is the baseline laser intensity and $V_{laser,dark}$ is the dark offset for the baseline. All measurements at 90^o with static light scattering are derived from this equation.

2.3.6 Dynamic light scattering measurement of particle size

Dynamic light scattering was used to measure the size and molecular diffusivity of proteins in solution. Molecular diffusivity arises from Brownian motion, which describes the random diffusion of solutes occurring due to collisions with solvent molecules. Diffusivity is defined with solute diffusion coefficients, which are related to the mean square displacement of solutes over a set time. The hydrodynamic radius of the molecule can be calculated from the diffusivity using Stokes-Einstein equation:

$$R_h = k_B T / 3\pi\eta D_0 \qquad \text{Equation (19)}$$

Where R_h is the hydrodynamic radius, k_B is the boltzmann constant, η is the solution viscosity and D_m is the diffusion coefficient. Note that the R_h corresponds to the overall hydrodynamic structure that is diffusing through the solution and is a measure of the protein with bound ions and water molecules. Solution viscosity for urea samples was taken from

The diffusion coefficient can be measured from correlations of fluctuations in the light scattered by the particles in solution. More specifically, the diffusion coefficient is related to the auto correlation function (g), given as:

$$g^2(q,t) = < i(0)i(t) > I < i(\infty) >^2 \qquad \text{Equation (20)}$$

Where *i(0)* is the intensity at time *0*, *i(t)* is the intensity at time "t" and i(∞) is intensity at infinite time.

Intensities separated by a short time interval will be very similar. With increasing time, however, there will be no correlation between the intensities. The correlation function $g^2(t)$ is a measure of how quickly the intensities loses their correlation, and can be described by an exponential decay, with a decay constant Γ. The decay constant for a monodispersed sample is given by:

$$\Gamma = D_m q^2 \qquad \text{Equation (21)}$$

Where D_m is the mutual diffusivity and *q* is the scattering wave vector given as:

$$q = 4\pi n_o/\lambda_o \sin(\frac{\theta}{2})$$ Equation (22)

Where n_o is the solutions refractive index, λ_o is the wavelength and θ is the scattering angle. For all measurements, the refractive index of the sample used was 1.33, as the accurate refractive index could not be determined for all samples.

Particles that diffuse more rapidly correspond to a more rapid decay which can be used to define the diffusion constant of the particles in solution. Fitting of the correlation function to the decay constant is performed by the programs associated with the dynamic light scattering equipment, using a CONTIN algorithm (Provencher, 1982) or cumulant method (Koppel, 1972).

The mutual diffusivity of a protein is defined as:

$$D_m = D_0[1 + k_D\, c]$$ Equation (23)

Where D_o is the infinite dilution diffusivity, k_D is the interaction parameter and c is concentration.

D_o is the diffusivity of a protein in infinite dilution, defined by Stokes-Einstein equation and takes into account only hydrodynamic interactions that influence diffusion. D_o can be calculated by graphically extrapolating measurements of D_m vs protein concentration to zero protein concentration. K_D as the interaction parameter defines the protein-protein interactions that also influence the diffusivity of the protein. Here it is defined as:

$$k_D = 2B_{22}M_w/v$$ Equation (24)

Where B_{22} is the second virial coefficient, M_w is the molecular weight, and v is the partial specific volume of the protein.

Since this parameter takes into account protein-protein interactions in the form of B_{22}, the slope of D_m vs protein concentration will show how protein-protein interactions influence diffusivity.

Chapter 3: The effect of denaturants on long range electrostatic interactions and aggregation of proteins

3.1 Introduction

The most common method of refolding inclusion bodies is through solubilisation with a high concentration of denaturant, usually urea or Gdm HCl (6-8 M), followed by reducing the denaturant concentration (<1 M) to induce native folding. The additives used to aid in the resolubilisation process are chosen empirically through high throughput screening, with little understanding of the excipient's molecular influence on protein-protein interactions or aggregation. In solutions containing Gdm HCl electrostatic repulsion between proteins is screened as reflected by an ionic strength induced reduction in the second virial coefficient (Liu et al., 2005), and in some instances using Gdm HCl can accelerate aggregation (Vernaglia et al., 2004). However, there have not been any comprehensive studies of how urea influences protein-protein interactions and the link to its impact on protein refolding yields. Ho and Middelburg found a link between the second virial coefficient of proteins under denaturing concentrations and refolding yields, as more repulsive protein-protein interactions correlated with increased refolding yield (Ho and Middelburg, 2003). However, the origin of any increase or decrease in protein-protein repulsion under denaturing conditions and how denaturants influence aggregation during refolding is poorly understood. The effects of urea and Gdm HCl on protein-protein interaction and aggregation are compared in this chapter to better understand how these similar and widely used molecules may differ in their effect on protein-protein interactions and the link to refolding aggregation.

The mechanisms of denaturation by urea and Gdm HCl have been extensively studied. There is still no clear consensus on the mechanistic action of urea on protein structure, but a two-stage "dry molten globule" mechanism is the most commonly put forward (Hua, 2008). In the initial surface-solubilisation stage, urea preferentially binds with the peptide backbone and non-polar residues through stronger Van der Waals forces compared to water, thus replaces water molecules from around the protein surface (Timasheff, 2003, Hua, 2008, Rossky, 2008). Urea replaces peptide-water hydrogen bonds with peptide-urea hydrogen bonds, and solvates hydrophobic residues (Zangi et al., 2009). Urea has, however, been shown to be preferentially excluded from hydrophilic residues, namely asparagine and glutamate, and to a lesser extent arginine and lysine (Stumpe and Grubmuller, 2007). In the denaturation stage, the increased solubilisation of the surface of the protein by the preferential binding of urea increases the population of partially folded intermediates, allowing urea to penetrate the hydrophobic interior of the protein. The hydrophobic residues buried inside the protein are solubilised through Van der Waals interactions with urea, creating a "dry" molten globule structure, called so due to the absence of water molecules in the interior of the protein. The internal

secondary structure is thus weakened without hydration by water molecules (*Hua, 2008)*. There is also evidence that urea forms hydrogen bonds with the peptide-backbone and disrupts the hydrogen bonding network that stabilises the secondary structure motifs (Lim et al., 2009; Stumpe and Grubmuller, 2007). However, due to the relatively small differences in energy between protein-water and protein-urea hydrogen bonds, disruption of the hydrogen bond network does not seem to be the main energetic driving force of denaturation, and instead the formation of urea-protein hydrogen bonds may serve instead to avoid a large loss in enthalpy during unfolding (Lim et al., 2009; Zangi et al., 2009; Zhang and Cremer, 2010). The overall strong preferential interaction of urea with aliphatic and aromatic residues and the peptide backbone replaces the hydrophobic and hydrogen-bonding forces within the protein and stabilises the unfolded states of the protein.

Gdm HCl has been shown to reduce inter- and intra-molecular hydrophobic attractive interactions and unfold proteins through preferential Van der Waals interaction with hydrophobic residues, molecular "stacking" at aromatic residues such as tryptophan, and attenuation of the intramolecular hydrogen bonding of the carboxyl backbone atoms (O'Brien, 2007, Mason, 2007, Dempsey 2004, Shao 2012). The binding of Gdm ions leads to the displacement of water molecules from the surface of the protein, especially around negatively charged and, aliphatic and planar residues (Shao 2012). The tertiary structure of the protein is destabilised through weakening of hydrophobic interactions and the protein unfolds as the equilibrium shifts towards unfolded states. Though the mechanisms of urea and Gdm HCl appear similar, Gdm HCl is a more effective denaturant for several reasons. Firstly, the presence of Gdm or chloride ions negates structurally stabilising intramolecular electrostatic interactions (Dempsey, 2005). Secondly, the stacking mechanism of guanidine ions to the surface of aromatic residues does not occur with urea molecules, which has been shown to significantly increase the effectiveness of unfolding by Gdm HCl in aromatic-rich model proteins (Dempsey, 2005).

Second virial coefficient and hydrodynamic measurements of solutions containing lysozyme between 0 and 7 M Gdm HCl at low pH indicate how Gdm HCl affects protein-protein interactions (Lui et al, 2005). With increasing Gdm HCl concentration, an initial decrease in protein-protein repulsion is observed leading to a maximum protein-protein attraction at a Gdm HCl concentration of 1 M. This is due to the screening of electrostatic repulsion by the increased ionic strength of the solution. With increasing Gdm HCl concentrations above 1 M, an increase in repulsion is observed as the hydrophobic attraction between the proteins is reduced by the binding of the Gdm ions to the hydrophobic residues of the protein. It is theorised that this minima in repulsion at 1 M Gdm HCl contributes towards increased amyloid formation compared to low or high Gdm HCl concentrations in previous studies by (Vernaglia et al., 2004). The effects of urea on protein-protein interactions have yet to be clearly characterised. The main driving forces behind aggregation during refolding are the short-range hydrophobic attractive interactions and hydrogen bonding, and current literature suggests that both urea and Gdm HCl disrupt these types of interactions.

The solvent-averaged interaction between two identical proteins can be characterised in terms of the second virial coefficient. All protein-protein interaction forces contribute towards the value of the second virial coefficient, also described as the B_{22} value. The contribution of the different forces to the B_{22} value of a protein in solution can be written as a sum (25).

$$B_{22} = B_{22elec} + B_{22exl} + B_{22sr} \qquad \text{Equation (25)}$$

In equation (25), the second virial coefficient is related to the sum total of three terms. The first term is related to the electrostatic repulsive force, B_{22elec}. The electrostatic repulsive force is dependent on the square of the net charge of the protein and the screening length of the solution, which is inversely proportional to the square root of the ionic strength of the solution. Thus an increasing protein net charge leads to increased electrostatic repulsion between like-charged proteins. Increasing ionic strength weakens the effect of the electrostatic interaction as the range of the potential is reduced due to ionic screening. The B_{22elec} can be estimated from the measured zeta potential or effective net protein charge. The second term is the excluded volume contribution, B_{22exl}, which can be calculated from equation (12) using measurements of the protein hydrodynamic radius. The excluded volume contribution corresponds to the volume about a protein that is made inaccessible to other proteins. All other interactions are expected to be short-ranged and included in the last term of the equation, B_{22sr}. The short range interactive force is expected to have a range corresponding to one or two solvent layers surrounding the protein and is derived from forces such as hydrophobic attraction, hydrogen bonding, Van der Waals attractive interactions, and hydration forces. Using equation (25), the short range interactive force contribution to B_{22} can be calculated if the values for B_{22ex} and B_{22elc} are known.

In this chapter, we have performed static and dynamic light scattering experiments on lysozyme, ribonuclease A and prepro-LysPro-insulin (PPI) to determine how urea influences the protein-protein interactions of both net-positively (lysozyme, ribonuclease A) and net-negatively charged proteins (PPI) in solutions with low ionic strength buffers. The aggregation behaviour of the proteins was also analysed through dilution and dialysis refolding methods to determine whether there is a link between differences in protein-protein interactions under different denaturant conditions and aggregation behaviour. At low ionic strength conditions, protein-protein interactions depend sensitively on electrostatic contributions. Thus, to determine whether urea influences the double-layer charged repulsion between proteins, laser Doppler electrophoresis was used to measure the zeta potential of lysozyme between 0 and 2 M urea over a range of pH conditions.

3.2 Results and Discussion

3.2.1 Increased protein-protein repulsion is observed in proteins in solution with low to high concentrations of urea

The aim of this work is to understand the influence of urea on protein-protein interactions in low ionic strength solutions where electrostatic interactions are significant between strongly charged proteins. Protein-protein interactions were characterised in terms of the second virial coefficient from static light scattering. Static light scattering experiments were performed using a dual-syringe delivery system in which high concentration urea solutions were mixed in-line with the corresponding protein solution into online static light scattering and UV detectors following the procedures outlined in chapter 3. Figure 3.1 shows examples of the Kc/R plots used to calculate the second virial coefficient and molecular weight of the proteins under different solution conditions. The data corresponds to experiments carried out on lysozyme in solutions containing 50 mM sodium acetate at pH 4.5, with urea concentrations of 0, 2 or 4 M, and for preproinsulin in solutions of 20 mM TRIS at pH 9.0 and 0.5 M urea. The inverse of the *y* intercept is equal to the apparent infinite dilution molecular weight of the protein, while half the gradient of the line is equal to the second virial coefficient. The values of B_{22} and M_w shown in the figure legend have been regressed using a least squares method based on the LINEST function in Excel. The error bars shown in for the B_{22} values derived from Zimm plots correspond to the standard deviation in the slope and the intercept values.

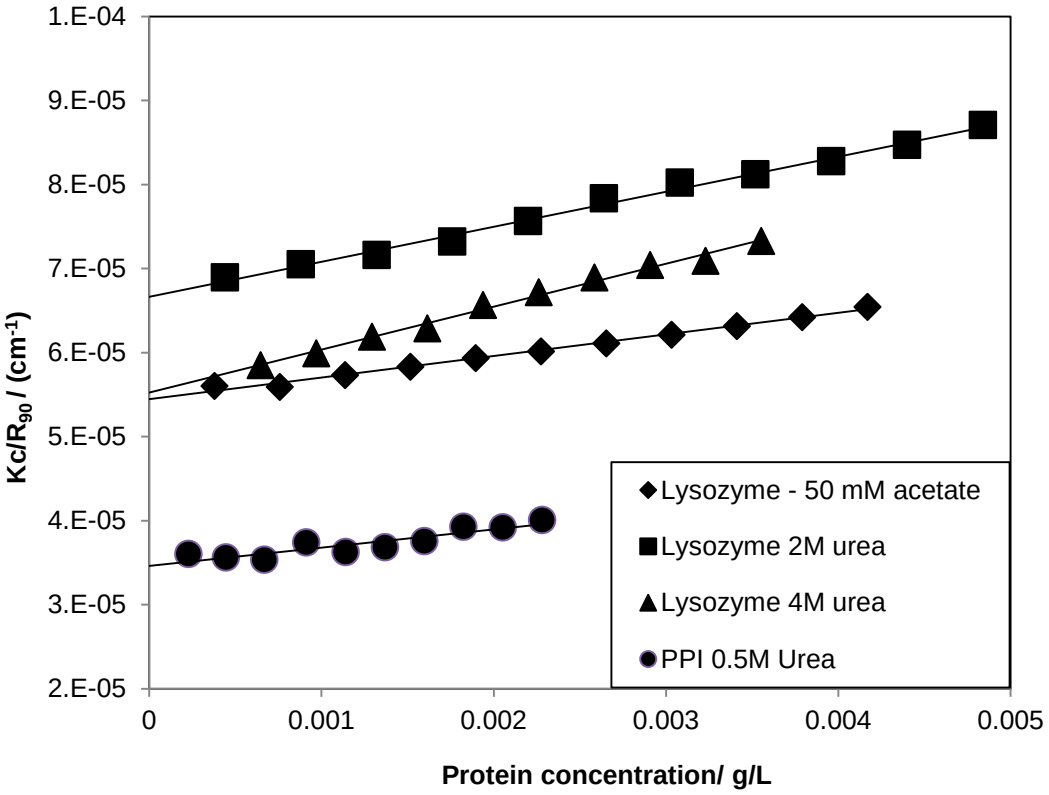


Figure 3.1. Examples of Zimm plots of experimental data for lysozyme in 50 mM sodium acetate solution in 0 M urea (B_{22} = 1.24 x10^{-3} mL Mol/g^2, Mw = 18.3 kDa), 2 M urea (B_{22} = 2.0 x10^{-3} mL Mol/g^2, Mw = 15.0 kDa) or 4 M urea (B_{22} = 2.55 x10^{-3} mL Mol/g^2, Mw = 17.1 kDa) solutions at pH4.5, and preproinsulin in 0.5 M urea solution in 20 mM Tris buffer at pH 9.0 (B_{22} = 1.08 x10^{-3} mL Mol/g^2, Mw = 28.8 kDa).

The second virial coefficient (B_{22}) values of lysozyme at 25°C were measured for solutions containing 50 mM sodium acetate buffer at pH 4.5 for urea concentrations ranging from 0 to 6 M. The results are shown in Figure 3.2. Also included in the figure are the B_{22} measurements for lysozyme in solution with sodium chloride at concentrations ranging between 0 and 1 M measured through static light scattering, and in solutions containing 0 to 7 M Gdm HCl concentrations as performed by Lui *et al* (2005). At low ionic strength and low pH with no additives present, the B_{22} value is large, equal to 1.24 x 10^{-3} mL Mol/g^2. This value is higher than the with B_{22} values for lysozyme in similar solution conditions form the literature, with Lui and Cellmer (2005) recording a value of 1.1 x 10^{-3} mL Mol/g^2 at pH 4.5 with 50 mM NaCl, and Curtis et al (2002) recording a value of 1.0 mL Mol/g^2 at pH 4.5 with 170 mM NaCl. Experimental error and differences in the techniques used may account for this error, though the deviation from the literature values is not significant. At pH 4.5, lysozyme has a reported net charge of 11 in solution with 0.1 M potassium chloride (Kruehner et al, 1999). However, a lower effective charge of 6 has been reported in low ionic strength citrate buffer solution at pH 4.5 (Javid et

al., 2007; Narayanan and Liu, 2003). In low ionic strength solutions, there are strong repulsive double layer forces between the lysozyme molecules due to the high net charge and large screening length.

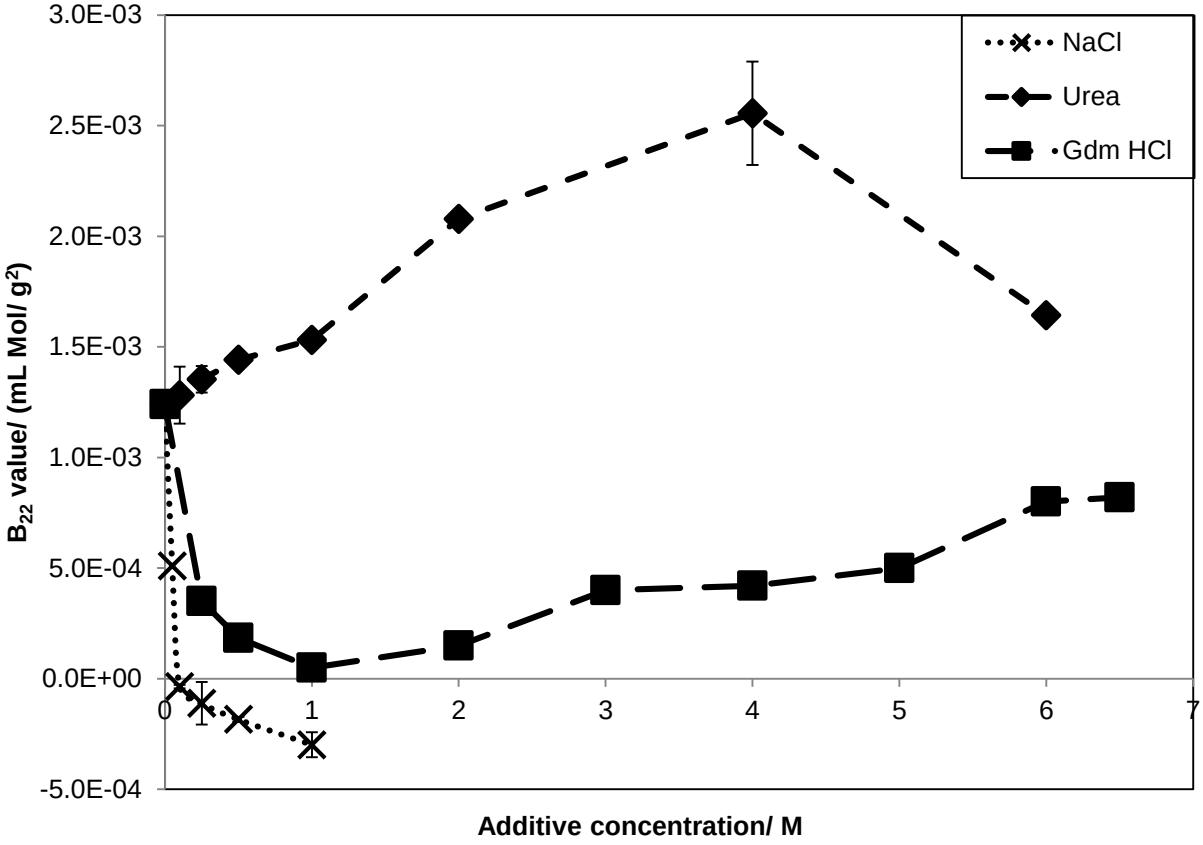


Figure 3.2. The B_{22} values of lysozyme in 50 mM sodium acetate buffer solution at pH 4.5 with either 0 to 6 M of urea, 0 to 7 M guandinium HCl or 0 to 1 M NaCl. GdmHCl solution values taken from Lui et al, (2005).

Increasing NaCl concentration leads to an increase in ionic strength and a decrease in protein-protein repulsion as electrostatic double layer forces are screened and attractive short-range interactions become more dominant. Likewise, in solutions with increasing concentrations of between 0 and 1 M Gdm HCl, net repulsion decreases as the Gdm and chloride ions screen electrostatic repulsive interactions. This decrease in repulsion, however, does not lead to a negative B_{22} value, as occurs with comparable NaCl concentrations. At the same ionic strength, interactions are more repulsive in Gdm HCl versus NaCl solutions indicating weaker short-range attractive forces in the presence of Gdm ions than are present in NaCl solutions. Increasing Gdm HCl concentration above 1 M further reduces short-range attraction between the proteins, indicated by increasing B_{22} values. The net change in B_{22} values of 7.5×10^{-4} mL Mol/g^2 with increasing Gdm HCl concentrations between 1 and 6 M indicates a significant reduction in protein-protein attraction. The cause of this decreased

attraction between proteins is due to the solubilising effects of Gdm ions, which has been attributed to binding of the denaturant to aliphatic and aromatic residues on the protein surface (Cellmer et al, 2005, Shao, 2012). This in turn may reduce the strength of attractive hydrophobic interactions between proteins and increase the net B_{22} value.

B_{22} values increase with increasing urea concentration and protein-protein repulsion reaches a maximum in solutions at 4 M urea. A lower B_{22} value of lysozyme in solution with 6 M urea in comparison to 4 M urea indicates protein-protein repulsion has been attenuated at high urea concentrations. Between 0 and 4 M urea, the value of B_{22} increases by 1.3×10^{-3} mL Mol/g^2. Unlike Gdm HCl, urea is non-ionic and thus increasing urea concentration does not screen double layer electrostatic repulsion. In contrast to salt solutions, strong repulsive interactions are retained with increasing urea concentration. However, this effect does not account for the increase in protein-protein repulsion inferred from the B_{22} measurements. At higher urea concentrations, it is expected that attractive hydrophobic interactions between proteins are reduced due to the denaturing effect of urea. This loss in short-range attraction could account for some of the net increase in B_{22} values at higher urea concentrations. At a urea concentration of 6 M, lysozyme has been shown to undergo unfolding at low pH, and thus the measurement of the B_{22} value here will be taken from a heterogeneous sample containing both folded and unfolded protein. This means that the value presented here may be inaccurate, as the value of B_{22} is derived from a model of homogeneous solutions. Therefore, the value must be understood to be only a guideline of the interactions that are occurring between lysozyme particles in solution with 6 M urea, rather than a definitive reading. Protein-protein repulsion also increases at low (<1 M) urea concentrations where urea denaturant effects are expected to be smaller. Since double-layer forces are large at low ionic strength, another possibility for the urea-induced repulsion might be attributed to changes in protein electrostatic properties. This possibility is probed later in this chapter through the measurement of the zeta potential of lysozyme in the presence of urea at varying pH and solution conditions.

The increase in protein-protein repulsion between lysozyme molecules observed in the presence of urea requires further investigation to determine whether the effects are non-specific and occur in other protein solutions. Ribonuclease A was chosen as a model protein to repeat static light scattering measurements under the same solution conditions as those of the lysozyme measurements described above. At low ionic strength and no additives present, ribonuclease A is net positively charged, with a predicted charge of +12e at pH 4.5 from the amino acid sequence (Protein Calculator v3.4, Life Protein 2013, Protein/Peptide property calculator, 2015) and has been shown to have a net charge of +11e through polyelectrolyte titration at pH 4.0 (*Tanford and Hauenstein, 1956, Horn and Heuck, 1982, Tessier, 2003*).

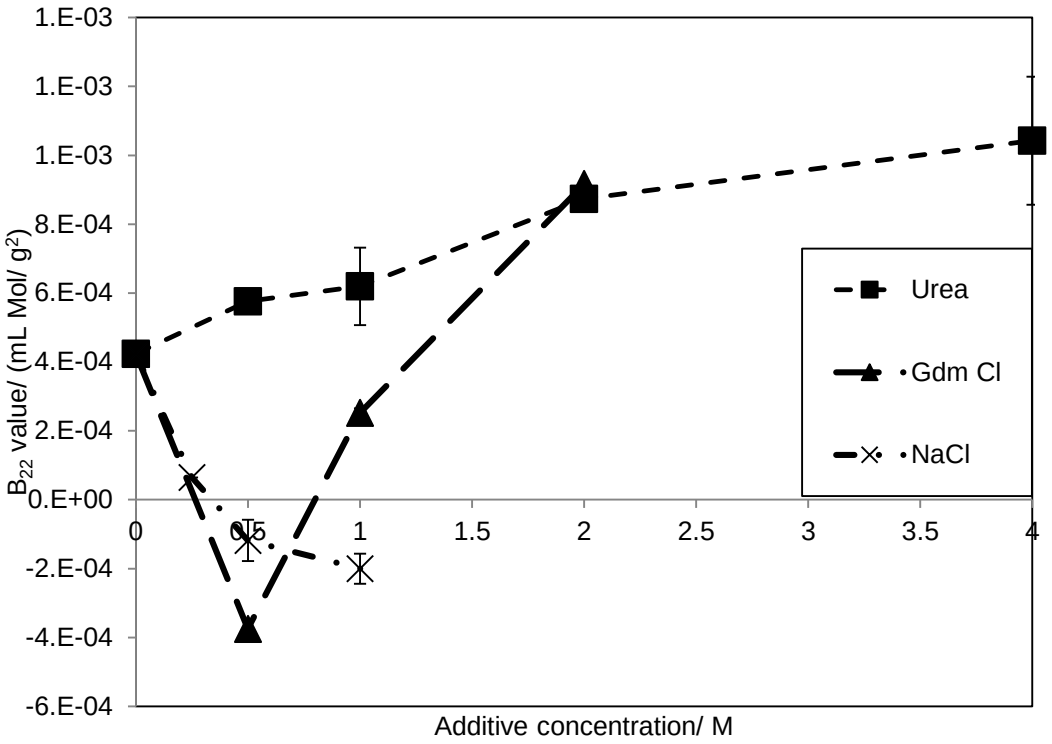


Figure 3.3. The B_{22} values of ribonuclease A in 50 mM sodium acetate buffer solution at pH 4.5 with either 0 to 4 M of urea, 0 to 2 M guandinium HCl or 0 to 1 M NaCl.

Figure 3.3 shows the B_{22} values of ribonuclease A in 50 mM sodium acetate solution at pH 4.5 with either urea, Gdm HCl or NaCl. The measured B_{22} value of ribonuclease in 50 mM sodium acetate solution is 4.2×10^{-4} mL Mol/g^2, reflecting weaker net protein-protein repulsion than measured between lysozyme particles under the same solution conditions. Self-interaction chromatography measurements of the B_{22} of ribonuclease A were performed by Tessier et al (2003), and report a B_{22} value of 3.46×10^{-4} mL Mol/g^2 at pH 4.0 at low ionic strength, which is less than the value measured with SLS at pH 4.5 here (Tessier et al., 2003). However, the similarity of the strength of the values indicates that the measurements are in general agreement as to the strength of the protein-protein interactions. The double-layer electrostatic repulsion contribution towards to the second virial coefficient for ribonuclease A may be lower than that calculated from the charge derived from titration experiments due to charge neutralisation by the buffer ions, or the net charge of the protein may be less than calculated. The net repulsion of ribonuclease A increases with the addition of urea, up to peak repulsion at a concentration of 4 M. The value of B_{22} increases by a value of 6.1×10^{-4} mL Mol/g^2 between 0 and 4 M urea, which is a smaller increase in B_{22} value to that observed for lysozyme solutions over the same change in urea concentration. Increasing NaCl concentration leads to screened repulsive electrostatic interactions and increased attraction between proteins, becoming negative in value in solutions of 0.5 M NaCl and decreasing in value at 1 M. The values calculated here are more negative in value than those calculated by Tessier at comparable NaCl concentrations.

The pattern of protein-protein interactions of lysozyme in solution with Gdm HCl is replicated with ribonuclease A. At low concentrations of Gdm HCl, repulsive double layer electrostatic interactions are screened due to increasing ionic strength. A maximum in attractive protein-protein interactions occurs at a Gdm HCl concentration of 0.5 M. Above 0.5 M, increasing Gdm HCl leads to more repulsive protein-protein interactions, as hydrophobic attraction between proteins is reduced by the binding of Gdm molecules to the hydrophobic surface patches of the protein.

There are two qualitative differences in the ribonuclease A data and the data collected from lysozyme in the same solution conditions. Firstly, the value of B_{22} for ribonuclease A in solution with 0.5 M Gdm HCl is more negative than that for solutions of 0.5 M NaCl. This implies that once double layer forces are screened, unlike in lysozyme solutions, the presence of Gdm ions leads to more attractive protein-protein interactions than with sodium chloride at the same ionic strength. The cause of this difference is not readily explainable, as the concentration of denaturant is too low for unfolding to occur and increase short-range hydrophobic attractive interactions.

Secondly, the B_{22} value of ribonuclease A in solutions at 2 M Gdm HCl is similar to that of ribonuclease A in a 2 M urea solution. In contrast, the protein-protein interactions are much more repulsive for lysozyme in urea-containing solutions when compared to Gdm HCl solutions at the same denaturant concentration. Net protein-protein repulsion in ribonuclease A solutions is equal to that observed in low ionic strength urea solutions, even though the double-layer electrostatic repulsive forces are screened at 2 M Gdm HCl, but not in the urea solutions. Thus, there is a net repulsive force in Gdm HCl solutions that is comparable in magnitude to that of the double-layer electrostatic repulsive force. The short range force contribution to the B_{22} value of ribonuclease A in 50 mM sodium acetate buffer with 2 M Gdm HCl is calculated from equation (25), with the assumption that electrostatic repulsive force contributions are effectively screened and are therefore close to zero. The net experimental B_{22} value (equal to 9.1×10^{-4} mL Mol/g^2) minus the B_{22exl} for ribonuclease A (equal to 3.2×10^{-4} mL Mol/g^2 calculated from equation (12) using an R_h equal to 1.9 nm *(Noppert et al., 1996),* gives a B_{22sr} of 5.9×10^{-4} mL Mol/g^2. This positive value for B_{22sr} indicates net repulsive short range interactions. This suggests that while Gdm HCl is expected to attenuate the hydrophobic short-range attractive interactions, in solutions at high concentrations of Gdm HCl there is an additional, as yet unknown, short-range repulsive force

A negatively charged protein preproinsulin was studied here to determine whether the effects of urea are generalisable across proteins with different charge polarities. Figure 3.4 shows the second virial coefficient of preproinsulin as a function of urea concentration for solutions containing 20 mM TRIS buffer at pH 9.0. The positive B_{22} value of preproinsulin in low ionic strength buffer equal to 5.3×10^{-4} mL Mol/g^2 indicates weak net repulsive interactions. The pI of preproinsulin is close to 5.0. Thus, in

solutions at pH 9.0, preproinsulin carries a net negative charge, leading to a double-layer electrostatic repulsion in low ionic strength solutions. This repulsion is increased significantly as urea is added up to a concentration of 2 M. Above this value, the plot of B_{22} exhibits a plateau between 2 and 4 M urea. The B_{22} values for solutions of preproinsulin molecules increase by a value of 1.1 x 10^{-3} mL Mol/g^2 between 0 and 4 M urea solutions. This trend is mirrored by solutions of lysozyme and by ribonuclease A, suggesting there may be a common mechanism for increasing net repulsion between proteins.

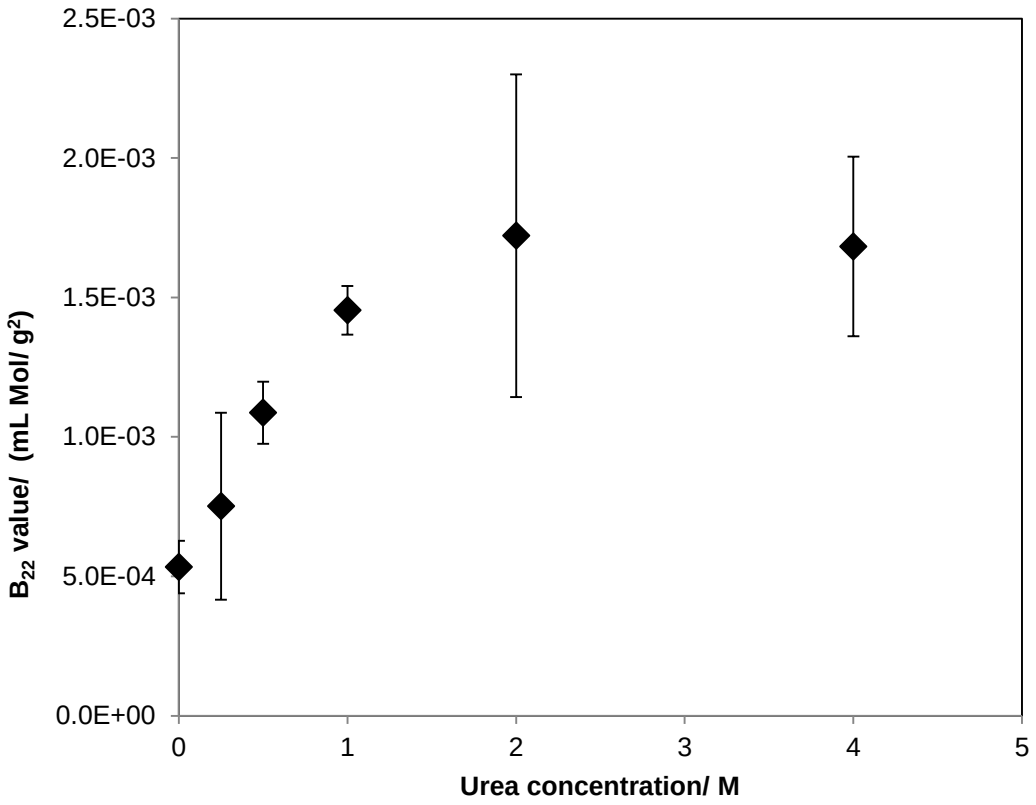


Figure 3.4. The B_{22} value of preproinsulin solutions at pH 9.0 with 20 mM Tris buffer and between 0 and 4 M urea

Ribonuclease A and lysozyme are net positively charged and preproinsulin is net negatively charged in the conditions studied here, indicating that the increase in repulsion induced by urea is independent of the sign of net protein charge. However, the origin of the urea-induced increase in protein-protein repulsion cannot be deduced from the measurements. The origin of the increased protein-protein repulsion could be derived from increased electrostatic double-layer repulsion or from attenuation of short-range attractive protein-protein interactions such as hydrophobic or Van der Waals forces with increasing urea concentration. However, in order to discriminate between these effects additional

parameters need to be measured as the second virial coefficient only provides an averages protein-protein interaction force.

3.2.2 Molecular weights

3.2.2.1 Lysozyme

The measured molecular weight from static light scattering can be used to assess the accuracy of the measurements. Deviations from the known monomer molecular weight (in the case of lysozyme, 14.3kDa) indicate the presence of impurities or systematic errors arising from approximations used in the light scattering equation. In figure 3.5a the molecular weights calculated from the static light scattering measurements of lysozyme as a function of urea or NaCl concentrations. In solutions containing 50 mM sodium acetate buffer, the molecular weight of lysozyme in calculated to be 18.3 kDa. This is above the actual molecular weight of lysozyme by 4 kDa, and the result suggests that there may be a small fraction of higher order aggregates, dimers and trimers for example, or reversible protein-protein associations that may be increasing the observed molecular weight. The molecular weight of lysozyme in NaCl solutions begins at 14.8 kDa in 50 mM NaCl before dropping below monomer to 12.5 kDa at 0.25 M, then returning to above monomer at higher NaCl concentrations. This variation, due to its non-linear nature, suggests experimental error, rather than any significant trend related to the protein.

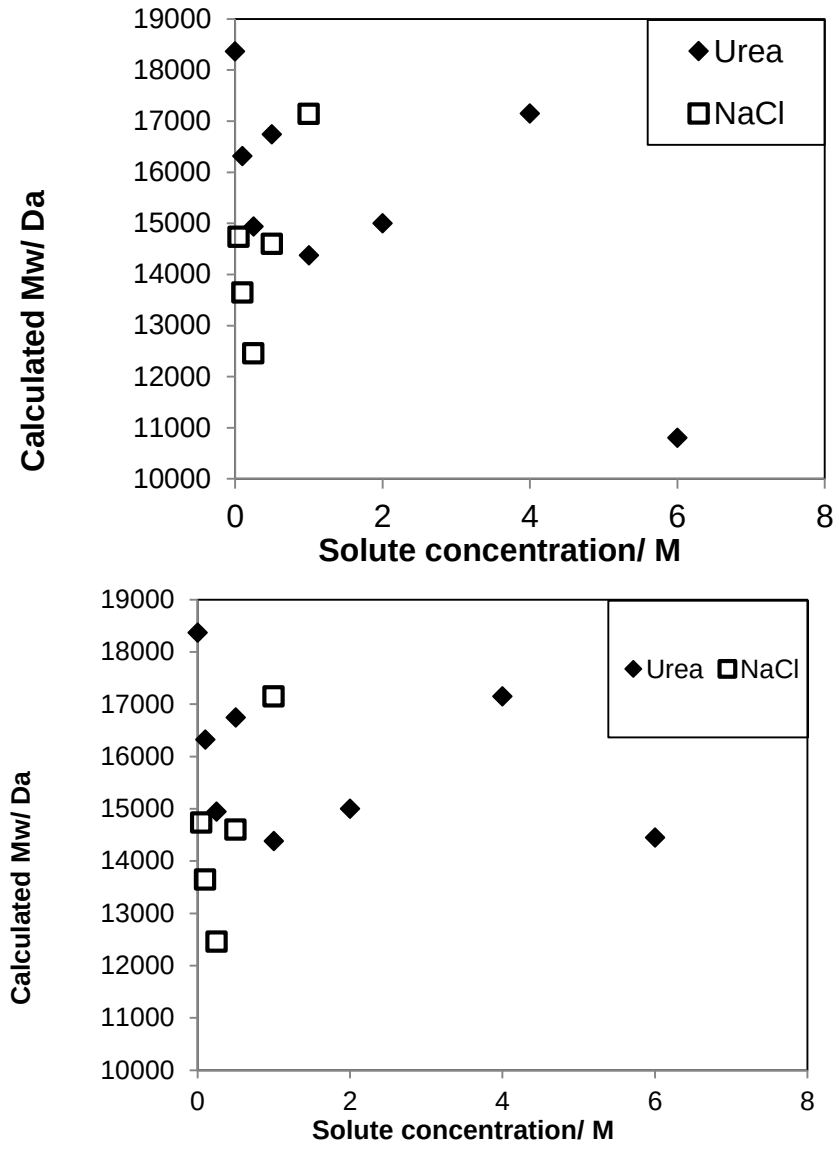


Figure 3.5 (a) The calculated molecular weight of lysozyme in urea or NaCl solutions measured through SLS. (b) The adjusted molecular weights of lysozyme in urea or NaCl solutions through adjustment of the dn/dc value used during the calculation.

The molecular weights calculated for lysozyme as a function of urea begin above monomer at 16.2 kDa at 100 mM urea. However, increasing urea concentration sees a downward trend in the calculated molecular weights to below monomer at 10.5 kDa in 6 M urea solution. This decrease in molecular weight is incremental and is not reducing monotonically, thus suggesting the trend of reducing observed molecular weight below monomer as a function of urea concentration must be an artefact of the calculation. The most likely cause of the decreasing molecular weight is due to a changing dn/dc constant as a function of solute (urea) concentration. The refractive index of the water (1.33) was used for the calculation of the B_{22} and molecular weight when the refractive index of the

solution could not be calculated. As the refractive index of high concentration urea solutions is higher than water solutions, this could also account for changes in the molecular weight below monomer. To account for this, the data analysis of SLS experiments where the calculated molecular weight of the protein significantly below that of monomer has been altered by changing dn/dc to a value such that the molecular weight is equal to monomer value. This altering of dn/dc value has an effect on the second virial coefficient, therefore the values of B_{22} presented are the values produced from the alteration of the dn/dc value. Figure 4.4b shows the molecular weights of lysozyme as a function of urea and NaCl concentrations derived from altering dn/dc values to obtain the correct molecular weight.

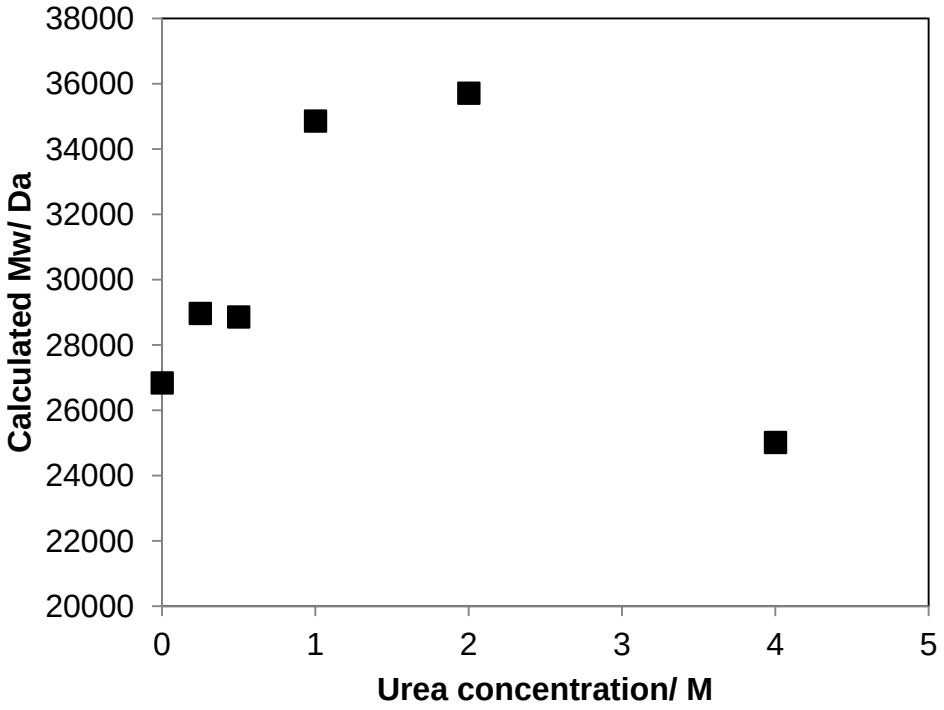


Figure 3.6. The calculated molecular weight of preproinsulin in urea solutions measured through SLS

3.2.2.2 Preproinsulin

The molecular weights calculated from the static light scattering measurements of preproinsulin as a function of urea concentration are shown in figure 3.6. The molecular weight of preproinsulin in 20 mM TRIS buffer is observed to be 26 kDa, which is over twice the monomer sequence molecular weight of 11.3 kDa. The high measured molecular weights reflect the presence of high molecular weight aggregates present alongside the monomer preproinsulin within the sample. With increasing urea concentration, the observed molecular weight increases at 36 kDa in 2 M urea solution, before reducing to 24 kDa in 4 M urea solutions. The observed molecular weights of preproinsulin in 1 and 2

M urea solutions appear to have risen monomerically in comparison to the molecular weight measured in urea-free solutions. This suggests the presence of a trimer aggregate in the preproinsulin solution.

3.2.3 Measurement of protein diameter and interaction parameter K_D of lysozyme in urea solutions through dynamic light scattering

The measurement of the diffusivity of lysozyme in solutions of urea at pH 4.5 using dynamic light scattering was used to determine an additional protein-protein interaction parameter, k_D, to verify the finding of a urea-induced increase in protein-protein repulsion. An example plot of the measured diffusivity verses concentration of lysozyme in 50 mM sodium acetate solution at pH 4.5 is shown in figure 3.7.

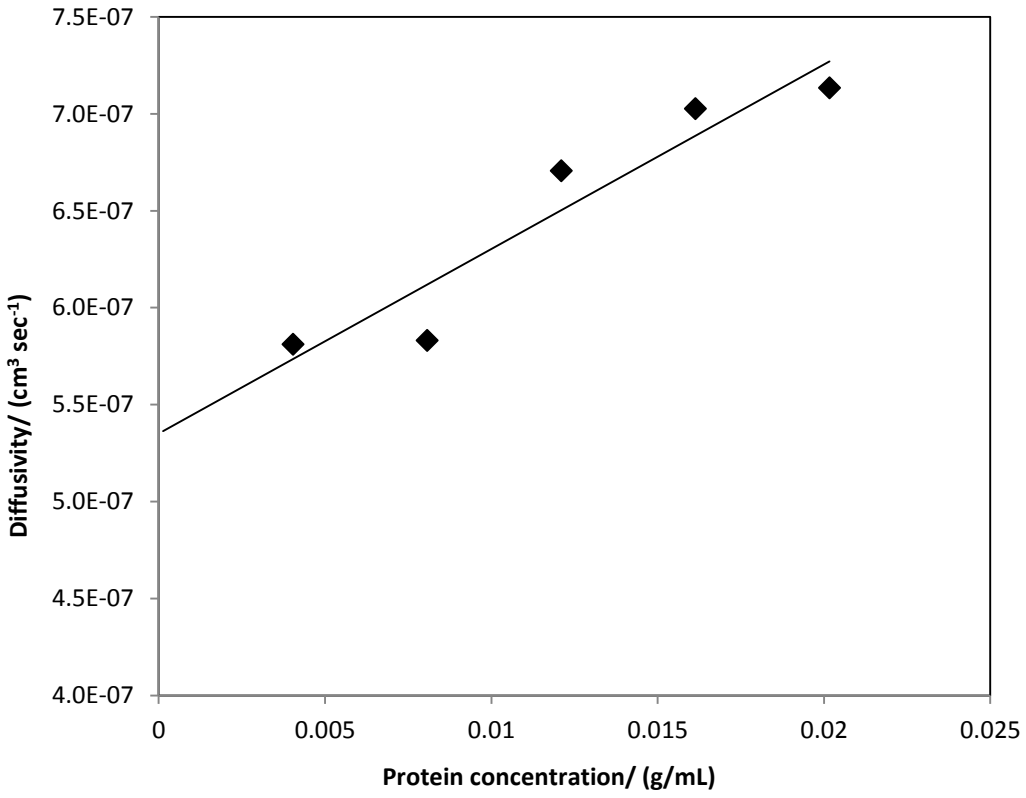


Figure 3.7. The diffusivity of lysozyme in 50 mM sodium acetate buffer solution at protein concentrations between 0.004 and 0.02 g/mL

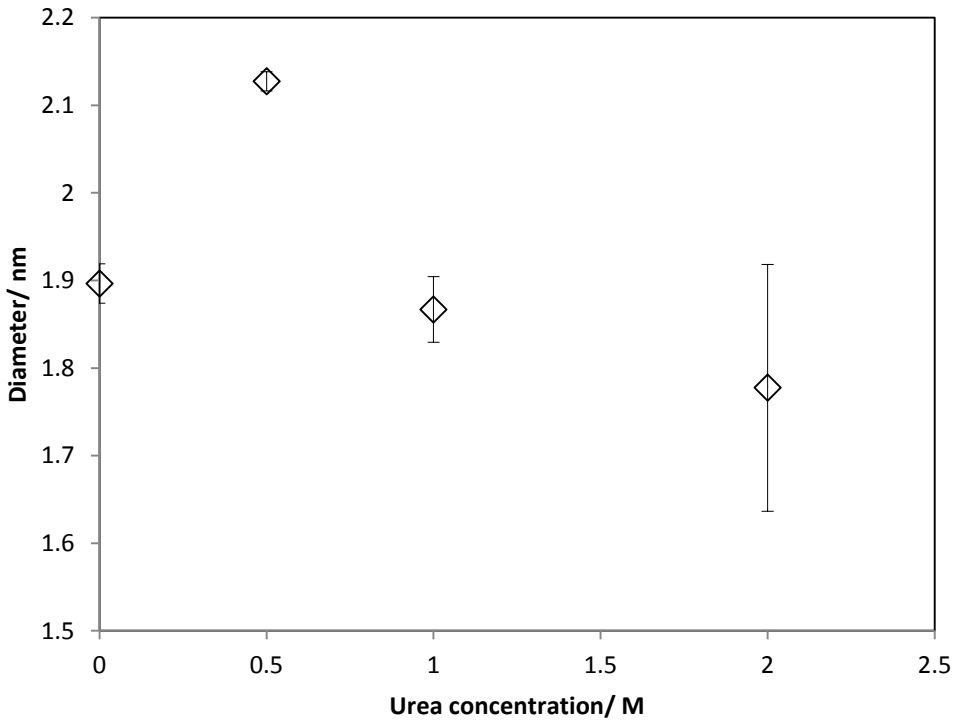
The relationship between the diffusivity, D_m, of the protein and k_D is represented in equation (23).

$$Dm = D_0[1 + k_D c] \qquad \text{Equation (23)}$$

Where D_0 is the infinite dilution value of the diffusion coefficient and c is the protein concentration in g/mL. D_0 is equal to the y intercept of the line of best fit for a plot of D_m verses c. Plotting D_m/D_0 verses c, the value for k_D is derived from the gradient of the line of best fit. The radius of hydration, R_h, of the protein is derived from D_0, using the Stokes-Einstein equation, equation (19), shown below.

$$R_h = k_B T / 3\pi\eta D_0 \qquad \text{Equation (19)}$$

where k_B is the Boltzmann's constant, T is the absolute temperature, and η is the solution viscosity.



size measured by Lui and Cellmer (2005) of 1.87 nm, and the size of 1.89 nm measured by Pamar and Muschol at low ionic strength (*Lui and Cellmer, 2005, Pamar and Muschol, 2009*). With increasing urea concentration, the size of the protein does not change significantly from the diameter measured in urea-free solutions. This is consistent with studies that have shown that lysozyme does not change in hydrodynamic size in the presence of denaturing concentrations of Gdm HCl (Lui et al, 2005).

From the diameter of the protein, the excluded volume contribution to the second virial coefficient, B_{22ex}, can be calculated using equation (12).

$$B_{22ex} = \frac{\left(\frac{16}{3}\right)\pi r^3 N_{AV}}{M_w^2}$$ Equation (12)

Where *r* is the radius of the protein, N_{AV} is Avogadro's number and M_w is the molecular weight of the protein. Using the value of 1.9 nm for the radius and 14600 as the molecular weight of the protein, an excluded volume contribution of 3.245 x10^{-4} mL Mol/g^2 is calculated for lysozyme in solutions containing between 0 and 2 M urea concentrations.

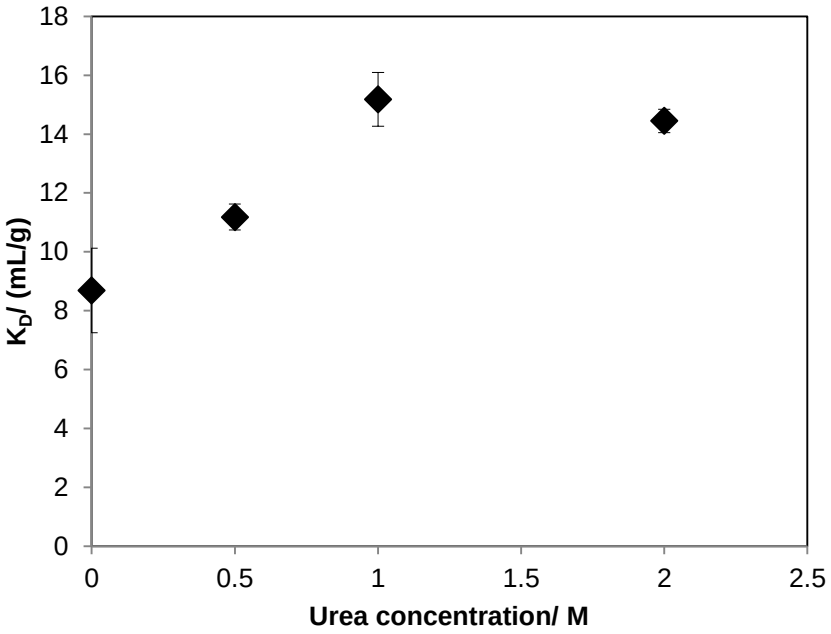


Figure 3.9. The k_D of lysozyme in 50 mM sodium acetate buffer solution at urea concentrations between 0 and 2 M.

K_D reflects the effects of protein-protein interactions on protein diffusivity. The interactions included both hydrodynamic and thermodynamic contributions. Previous studies have shown a linear relationship between k_D and the second virial coefficient (Muschol and Rossenberger 1995, Roberts, 2014, Saluja, 2010, Connolly, 2012). Positive values of k_D indicate repulsive protein-protein interactions while negative values indicate attractive protein-protein interactions. Figure 3.9 shows the K_D values of lysozyme in 50 mM sodium acetate solutions with 0, 0.5, 1, and 2 M urea concentrations. The K_D value of lysozyme in the urea-free solution is 8.7, reflecting repulsive protein-protein interactions. This value is significantly lower than that obtained by Muschol and Rossenberger (1995) for lysozyme under similar solution conditions. Their value is 33.1 is significantly higher than that obtained here. With increasing urea concentration, K_D increases in value, indicating an increase in the protein-protein repulsion, as is observed with second virial coefficient measurements at comparable urea concentrations.

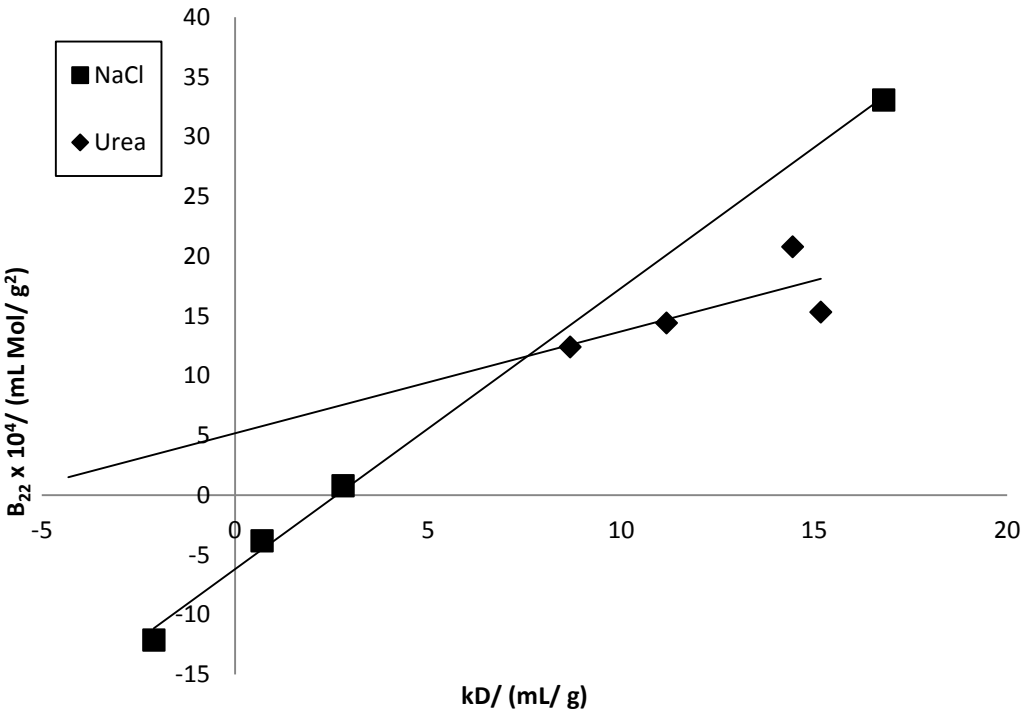


Figure 3.10. A plot of k_D verses B_{22} values measured through DLS and SLS, respectively. Values for NaCl solutions are taken from Muschol and Rossenberger, (1995).

Muschol and Rossenberger (1995) presented B_{22} and k_D values for lysozyme solutions containing sodium chloride concentrations, and found a linear relationship between B_{22} and k_D (Muschol and Rosenberger, 1995). The measurements of the values of B_{22} and k_D of lysozyme in solutions of

and the value of k_D, similar to that measured by Muschol and Rossenberger, though the line for the urea values do not fit the line given for NaCl solutions. This may be attributed to different measurement and analysis techniques producing lower values of B_{22} or k_D, or inaccurate measurements due to the presence of urea causing additional background scattering in DLS measurements. However, a positive correlation can be observed for the values of B_{22}/k_D in urea solutions, indicating that k_D and B_{22} have a linear relationship and increasing kD values are indicative of increasing repulsive protein-protein interactions. However, the B_{22}/k_D measurements presented in this chapter does not follow the same relationship as defined in equation (24), suggesting inaccuracies in measurements compared to those presented by Muschol and Rossenberger.

3.2.4 Urea induces a small reduction in double-layer electrostatic repulsion

Zeta potential measurements characterised by laser Doppler electrophoresis (see chapter 2 for experimental details) were used to determine the effect of urea on the double layer potential (or the net charge) of lysozyme, which in turn provides a measure for the effect of urea on the double layer force between proteins. Figure 3.11 shows the measured zeta potential of lysozyme at pH 4.5, 6.4 and 9.0 at urea concentrations ranging between 0 and 2 M. At each pH value, there is a small reduction in net surface charge potential of the protein, but provides no evidence that urea has a significant effect on the protein-protein pair electrostatic properties. In fact, the reduction in zeta potential indicates that increasing urea concentration should weaken the double layer repulsion between lysozyme molecules. Thus, an effect of urea on the double-layer electrostatic repulsion is not the main contributing factor to the increased B_{22} values measured in urea solutions.

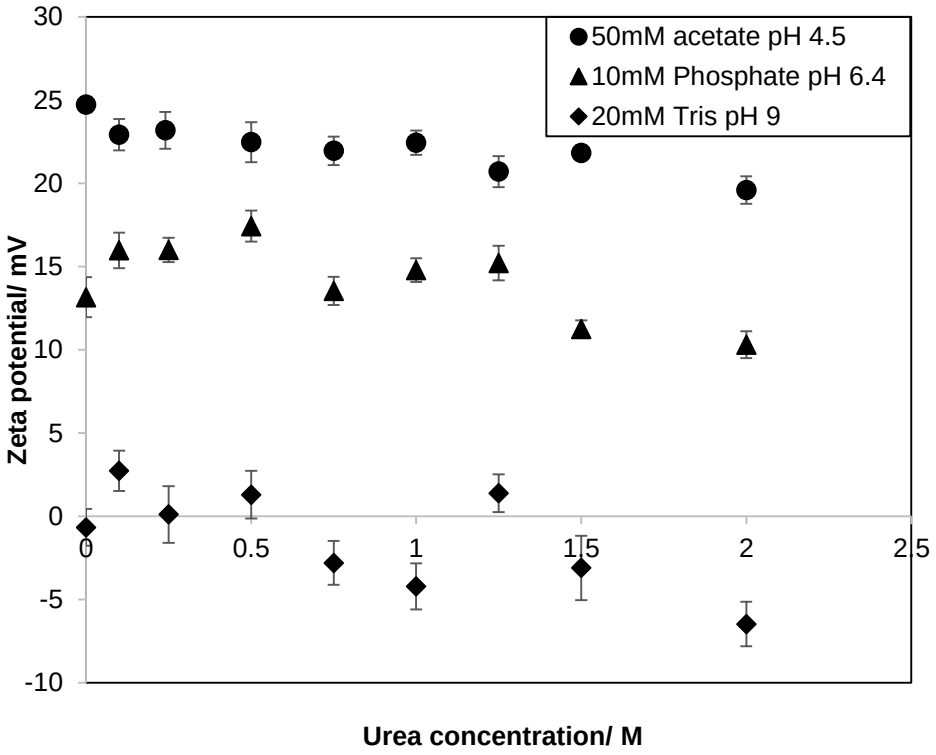


Figure 3.11. The measured zeta potential values of lysozyme solutions in 50 mM sodium acetate (pH 4.5), 10 mM phosphate (pH 6.4) and 20 mM Tris (pH 9.0) and between 0 and 2 M urea, using measurements of electrophoretic mobility through laser Doppler electrophoresis.

The effective charge of lysozyme can be calculated from the mobility measured through laser Doppler electrophoresis using equations (3), (4), (5) and (11). The results are presented in figure 4.10.At pH 4.5, the net charge of lysozyme in urea-free solutions is calculated as +7.1e. This value is significantly lower than the value of +11e calculated from titration experiments (Kruehner et al, 1999). The difference in the measured charge of the protein may be due to a number of factors. Firstly, preferential anion binding to positively charged amino acid residues may be occurring, leading to neutralisation of the charge of the protein and a lower net charge. Secondly, approximations used for the measurement of the mobility of the protein within the instrumentation may lead to a lower calculated charge value than that measured through a different method.

At both pH 4.5 and pH 6.4, lysozyme loses approximately 2 positive charge units when increasing urea concentration from 0 to 2 M. This reduction in protein charge leads to a reduction in electrostatic repulsion between lysozyme particles. The reduction in the charge of the protein as a function of urea is hypothesised to derive from changes in the pKa values of individual amino acids on the surface of

the protein due the increased structural flexibility in the presence of urea. Increased structural flexibility may also be responsible for increased anion binding to the protein surface, causing further neutralisation of the positive charge on the protein surface.

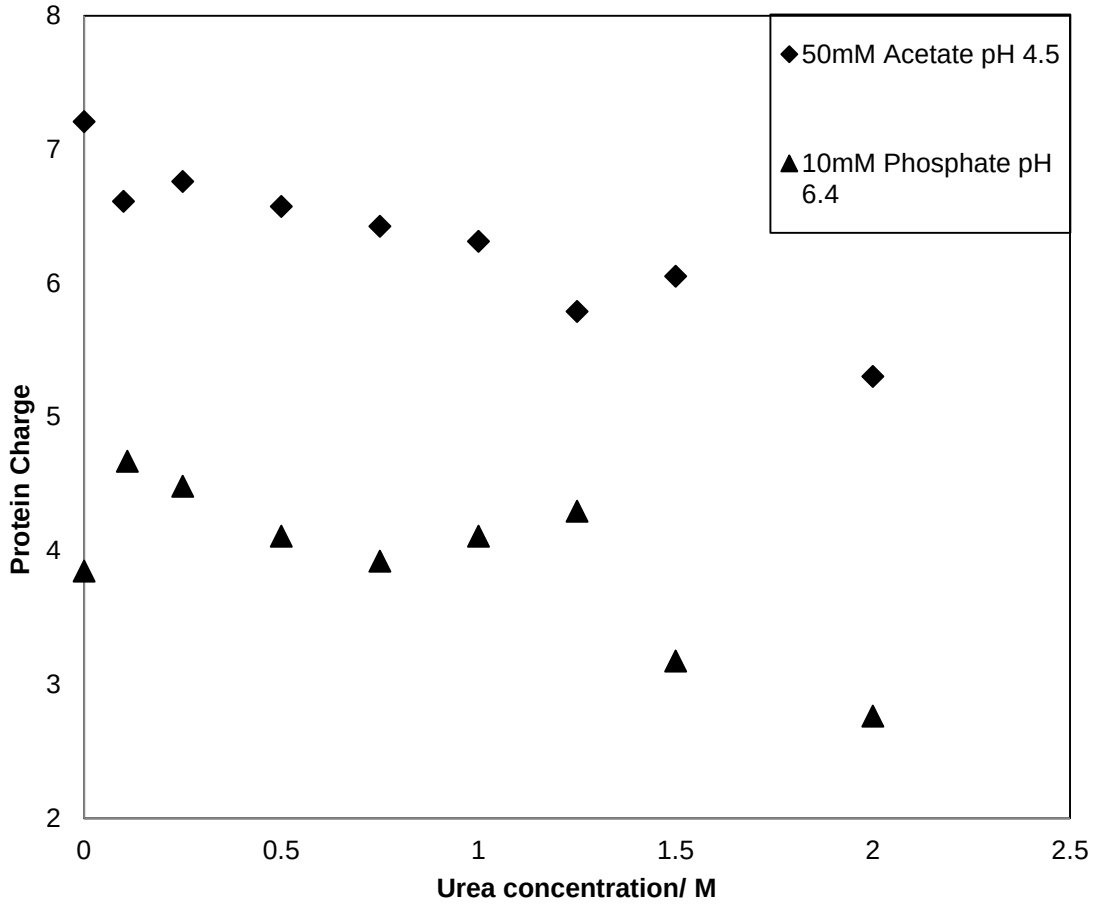


Figure 3.12. The calculated charge values of lysozyme solutions in 50 mM sodium acetate (pH 4.5) and 10 mM phosphate (pH 6.4) between 0 and 2 M urea, calculated from measurements of the electrophoretic mobility of the protein.

The zeta potential of lysozyme in solutions with or without urea was measured as a function of pH with incremental addition of either HCl or NaOH solutions. Figure 3.12 shows the zeta potential of lysozyme between pH 3.0 and pH 9.0 at set urea concentrations of 0, 1 or 2 M. Though there is little difference in zeta potential in solutions at 0 M and 2 M urea for all pH values, there is a slight reduction in the charge of lysozyme in solution with 1 M urea between pH 4 and pH 7 compared to urea-free solutions. A previous study by Ortore et al, 2008, using SANS to measure the effective charge of lysozyme in solutions of between 0 and 3 M urea, has shown comparable results, observing a decrease in net protein charge in 1 M urea solution, while at 2 and 3 M urea protein charge was equal to urea-free solutions (Ortore et al., 2008). It appears that the presence of urea has a slight effect on the charged properties of lysozyme. It has been hypothesised that in solutions containing 1

M urea, increased structural flexibility of the protein leads to downward shifts in the pKa of positively charged residues, which has the effect of reducing the net charge on the protein at moderate pH. The non-monotonic effect needs further investigation, which is beyond the scope of this work.

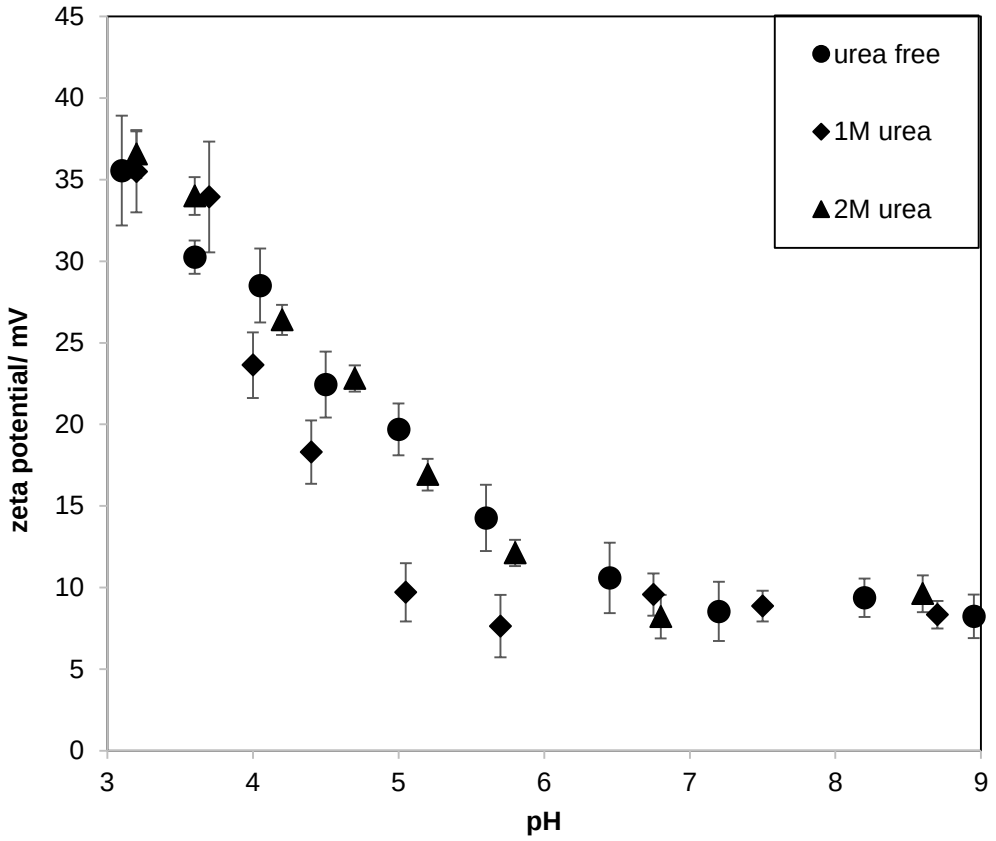


Figure 3.13. The calculated zeta potential of lysozyme in solutions of 0, 1, or 2 M urea as a function of pH.

3.2.5 Calculation of the B_{22} contributions of protein-protein interaction forces

The finding that urea does not increase the protein double layer potential suggests urea instead alters non-electrostatic, short-ranged interactions. Short-range interactions include excluded volume and attractive interactions, such as hydrophobic forces or hydrogen-bonding interactions. Dynamic light scattering measurements of lysozyme in solutions between 0 and 2 M urea indicate there is no statistically significant change in the hydrodynamic radius of the protein, thus the excluded volume contribution, B_{22exl}, does not change with increasing urea concentration. Thus, we expect the main effect of urea is to alter the short range interactions between proteins as denoted by B_{22sr} in equation (24).

Figure 3.13 shows the calculated B_{22} contribution of the different forces involved in protein-protein interactions of lysozyme in solutions between 0 and 2 M urea. The excluded volume contribution, B_{22exl}, has been calculated from the measured hydrodynamic size of 1.9 nm, using the Stokes-Einstein equation (19).The contribution of double-layer electrostatic repulsion to the B_{22} value, B_{22elec}, has been calculated using equation (11), with the effective net charge of the protein measured from the electrophoretic mobility of the protein in urea solutions from 0 to 2 M at pH 4.5 (figure 3.9). The B_{22sr} value is calculated from the B_{22} values displayed in figure 4.2, minus the B_{22exl} and the B_{22elec} values calculated for each urea concentration.

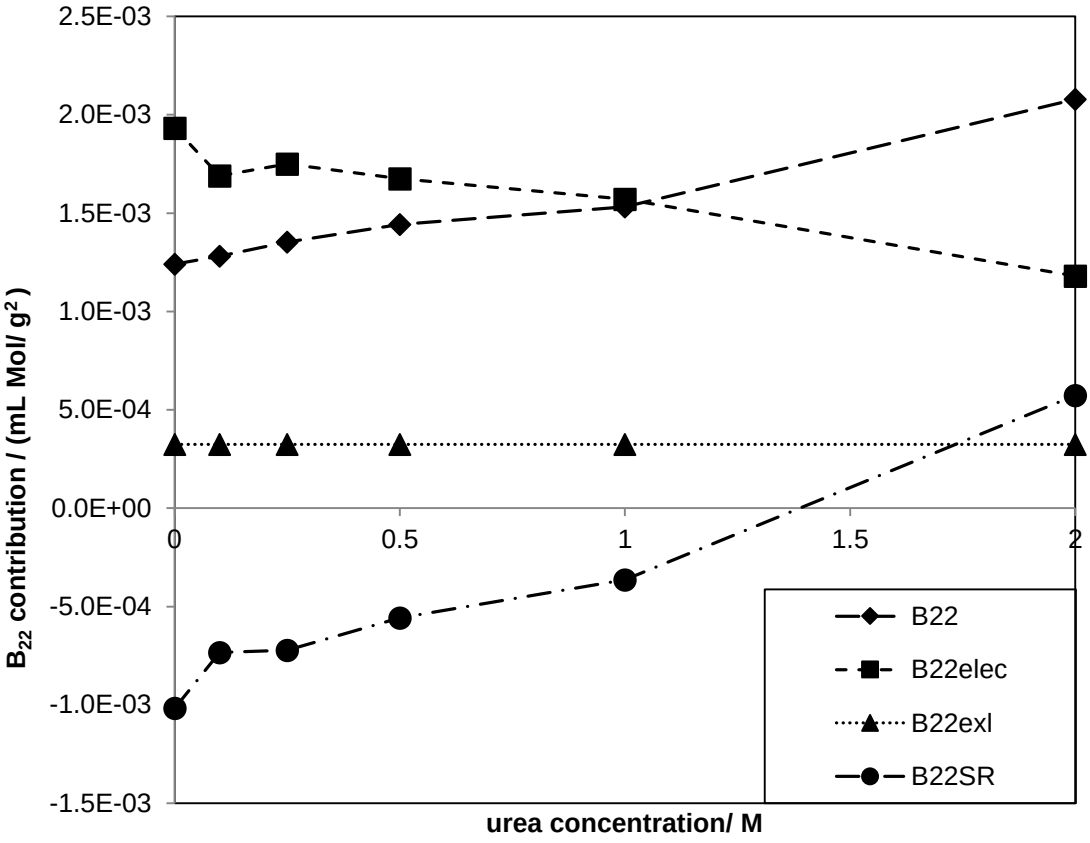


Figure 3.14. The calculated individual contribution towards the B_{22} value of lysozyme in solutions with 50 mM sodium acetate at pH 4.5 between 0 and 2 M urea.

The B_{22sr} value for lysozyme is negative in urea-free solutions, indicating net attractive short-range interactions, and B_{22sr} values increase monotonically with increasing urea concentration. In solutions of 2 M urea, B_{22sr} becomes positive, indicating urea induces net repulsive interactions between proteins. The decreased protein-protein attraction is likely due to the ability of the urea to bind to

hydrophobic residues on the surface of the protein and shield protein-protein hydrophobic attraction. However, the source of the repulsive short-range interaction at 2 M urea is currently unknown, and is not easily explained by DLVO theory, though a hypothesis is that strong hydration forces at the surface of the protein may lead to repulsive interactions, as the accumulation of urea at the surface of the protein will lead to the formation of an energetic barrier to attractive protein-protein interactions. The forces that contribute to the denaturation of the protein also determine the increased repulsion between protein molecules as a function of urea.

The calculation of the individual contributions to the second virial coefficient using equation (25) can be applied to solutions of lysozyme and Gdm HCl. The B_{22} values for lysozyme at pH 4.5 as a function of Gdm HCl are taken from Liu et al, (2005). The double-layer electrostatic contributions, B_{22elec}, for lysozyme in 50 mM sodium acetate solutions at pH 4.5 with no denaturant was calculated from the effective net charge of 7.2, as measured by Laser Doppler electrophoresis (figure 3.9). The high ionic strength of Gdm HCl solutions leads to a short screening length and thus protein-protein electrostatic repulsive force is close to 0 in Gdm HCl solutions above 250 mM. Liu et al (2005) showed that the hydrodynamic radius of lysozyme remains constant at all Gdm HCl concentrations in non-reducing conditions, indicating that the B_{22exl} remains constant. The calculated B_{22} contributions are presented in figure 3.14 for lysozyme in Gdm HCl solutions between 0 and 2 M. The B_{22sr} increases monotonically with increasing denaturant concentration and becomes positive at a denaturant concentration between 2 and 3 M Gdm HCl. The reduction in the B_{22sr} as a function of Gdm HCl concentration is a result of the Gdm ions reducing the hydrophobic attraction between the protein molecules through solubilisation of the aliphatic and aromatic residues on the protein surface. The positive values of B_{22sr} observed above 3 M Gdm HCl indicate the presence of repulsive forces, potentially similar to that observed at urea concentrations above 2 M. The origins of both of the theorised forces are as yet unknown, and it is unclear whether they are related or have separate origins in each denaturant. However, it is clear that the decrease in short range attractive protein-protein interactions is linked to the solubilising effect of urea and Gdm HCl solutions.

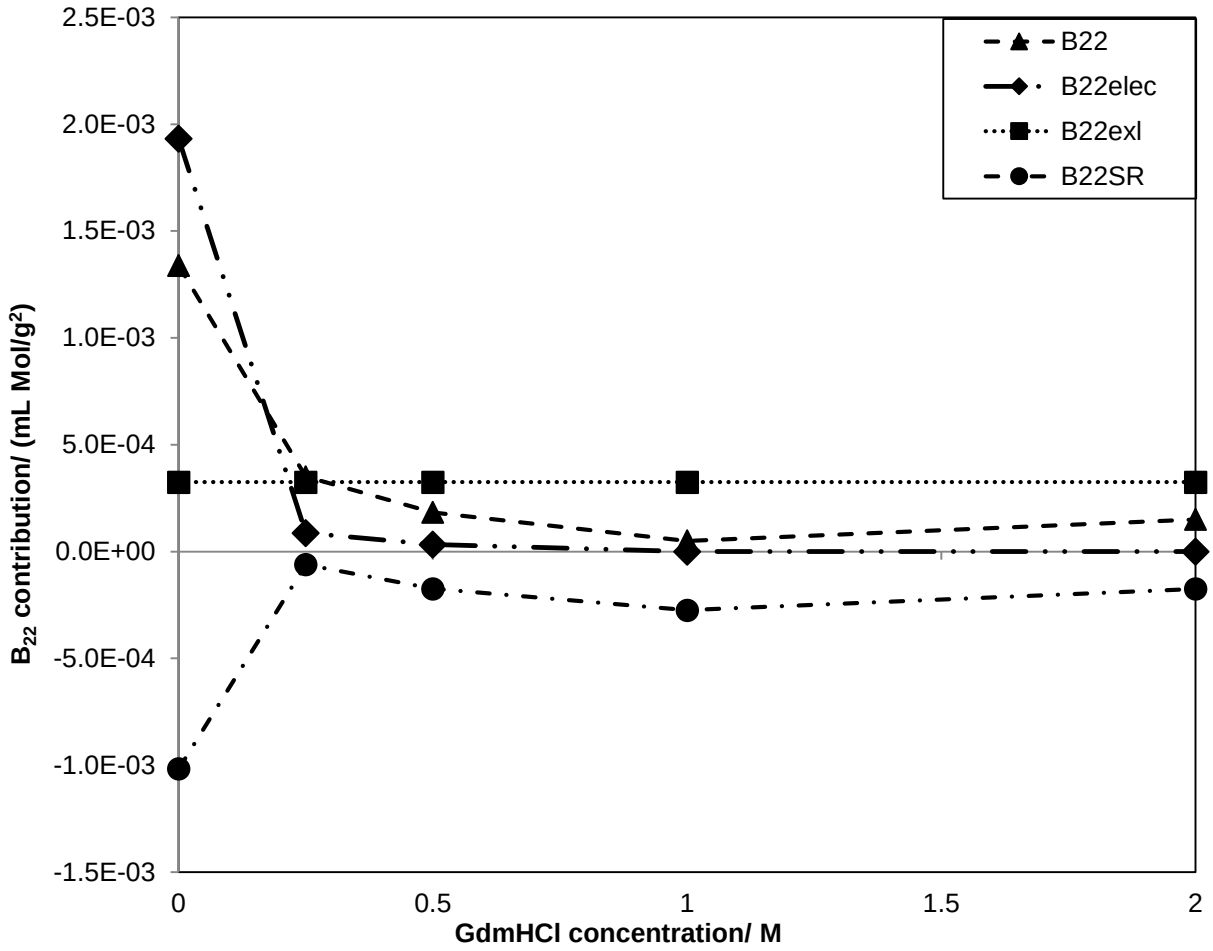


Figure 3.15. The calculated individual contribution towards the B_{22} value of lysozyme in solutions with 50 mM sodium acetate at pH 4.5 between 0 and 2 M Gdm HCl.

A comparison of the calculated values of B_{22sr} for lysozyme in solutions of either urea or Gdm HCl is shown in figure 3.15. In both urea and Gdm HCl solutions, protein-protein repulsion increased with increasing denaturant concentration. In solutions containing between 0 and 1 M of denaturant, protein-protein attractive interaction are weakest in Gdm HCl solutions compared to urea solutions, suggesting that Gdm ions attenuate short-range attraction at a lower concentration compared to urea. However, the attenuation of protein-protein attraction by Gdm HCl solutions plateaus at 0.25 M and a slight increase in attraction is observed between 0.5 and 1 M. It is unlikely that the B_{22sr} values for lysozyme in solutions of between 0.25 and 0.5 M Gdm HCl presented here are representative of the complete attenuation of all hydrophobic attractive interactions, as if this were the case the data would indicate that Gdm HCl would denature lysozyme at a much lower concentration than between 5 and 6 M. A hypothesis for the high B_{22sr} at 0.25 M Gdm HCl is presence of preferentially bound positively charged Gdm ions on the surface of the protein causing increased short-range repulsive interactions. However, this is unsubstantiated. It would instead be expected that B_{22sr} should increase

monotonically with increasing Gdm HCl concentration. The gradient of the curve for Gdm HCl concentrations of 1 M and above suggest that this is the case. On the other hand, B_{22sr} increases linearly with increasing urea concentration, and becomes greater than the corresponding value for Gdm HCl above 1 M denaturant concentration.

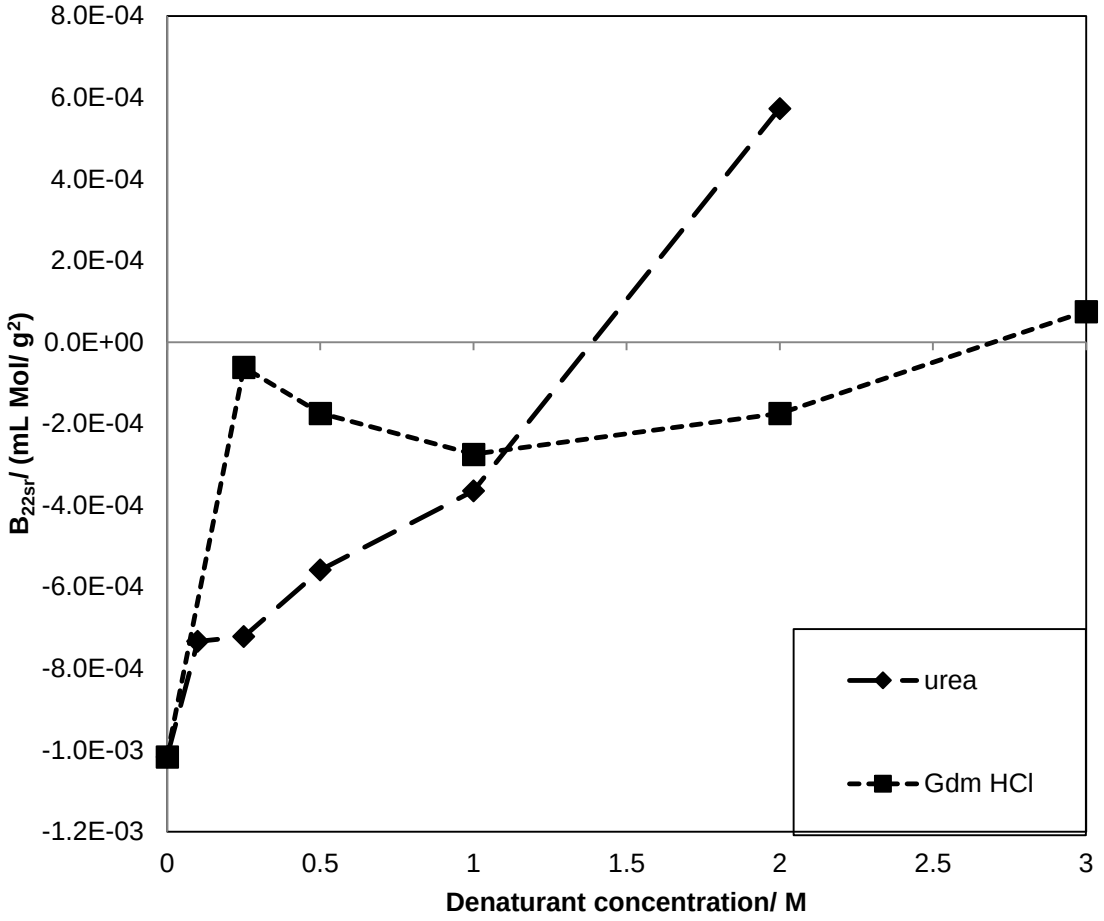


Figure 3.16. A comparison of the calculated individual contribution towards the B_{22} value of lysozyme in solutions with 50 mM sodium acetate at pH 4.5 between 0 and 2 M GdmHCl or between 0 and 2 M urea.

Though Gdm HCl solutions denature lysozyme at a lower molarity compared to urea, the calculated B_{22sr} value of lysozyme in urea solutions becomes net repulsive at a lower denaturant concentration (1-2 M) than Gdm HCl solutions (2-3 M). This is a striking difference that suggests that there are distinct differences through which the denaturants influence short-range protein-protein interactions at low ionic strength urea compared to high ionic strength Gdm HCl. The mechanism through which urea and Gdm HCl each influence protein-protein interactions cannot be determined from the data presented here. It has been hypothesised that urea and Gdm solutions denature proteins through the weakening of the hydrophobic attractive force, and disruption of the hydrogen bonding network.

However, each denaturant has been shown to bind and interact with the protein structural elements with different affinities. Gdm HCl has been shown to screen hydrophobic attraction by "stacking" in a planar fashion with aromatic residues (Dempsey 2004, Mason, 2008), while urea has been shown to disrupt the hydrogen-bonding network of the protein more effectively than Gdm (*Lim et al, 2008*). These key differences may explain the differences in B_{22sr} of lysozyme in solutions containing non-denaturing urea and Gdm concentrations. What is clear is that both urea and Gdm HCl solutions above 2 M induced repulsive short-range protein-protein interactions in lysozyme at low pH, with a significant increase in the value of B_{22} observed between 0 and 4 M urea in lysozyme (1.3×10^{-3} mL Mol/g^2), ribonuclease A (6.1×10^{-4} mL Mol/g^2) and preproinsulin (1.1×10^{-3} mL Mol/g^2) solutions. The net increase in B_{22} value may be derived predominantly from the reduction in the short-range attractive interactions, while increased hydration forces may also contribute. This leads to the hypothesis that the binding of both urea and Gdm to the surface of lysozyme lead to a reduction in hydrophobic and hydrogen-bonding attractive interactions and the formation of repulsive protein-protein hydration forces, which may contribute towards the increased solubility of proteins in both urea and Gdm HCl solutions.

3.2.6 Differences in aggregation behaviour of denatured lysozyme and preproinsulin in urea and Gdm HCl during refolding

There are significant differences in protein-protein interaction when comparing solutions at low to medium concentrations of urea or Gdm HCl. The next stage is to determine whether the net increase in repulsion observed in solutions with concentrations of urea between 0 and 6 M urea compared to those with 0 – 7 M Gdm HCl is translated into reduced aggregation behaviour. Both dilution and dialysis refolding experiments were conducted to compare the aggregation of lysozyme and preproinsulin in solutions where the denaturant was either urea or Gdm HCl.

3.2.6.1 Dilution refolding

Dilution refolding of lysozyme was conducted in solutions with 100 mM Tris-acetate buffer at pH 8.1. Lysozyme was solubilised at a concentration of 10 g/L in solutions containing either 6 M Gdm HCl or 8 M urea as the denaturant and 10 mM DTT as a reducing agent to a volume of 5 mL, denaturant concentrations lysozyme has been shown to unfold at within this pH range, and incubated at room temperature for 1 hour (Chen et al., 2014; Fang and Yi, 2003; Greene and Pace, 1974). The solutions were then diluted to a denaturant concentration of 0.25M urea or 0.188 M Gdm HCl with 100 mM TRIS-acetate buffer and 3 mM cysteine as an oxidising agent at pH 8.1, to a volume of 160 mL and

incubated for 16 hours at room temperature. A higher pH and ionic strength buffer was used for lysozyme refolding to increase the levels of aggregation. Previous experiments at lower pH did not produce quantifiable aggregation during dilution refolding and the effects of the denaturants on refolding yield could not be distinguished.

Dilution refolding for preproinsulin inclusion bodies used the refolding protocol provided by Fujifilm Diosnyth. The protein inclusion body slurry was solubilised at pH 10.5 with either 4.5 M urea or 6 M Gdm HCl solution with 10 mM DTT as a reducing agent and 365 mM ethanolamine as a pH regulating agent, at a ratio of 1.0 g of slurry to 8 mL of denaturant solution. A lower concentration of urea was used in this protocol due to the high solubility of preproinsulin inclusion bodies at pH 10.5 in urea. Gdm HCl was used at 6 M to ensure full solubilisation of the inclusion bodies comparable to that of the urea solution. The solubilised preproinsulin solution was then diluted to a final denaturant concentration of 0.25 M with the addition of 17 mL of refolding buffer per 1 mL of denatured protein solution. The refolding buffer contained 10% hexylene glycol as a refolding agent, 288 µM cystamine as an oxidising agent and 20 mM ethanolamine as a pH regulating agent, at pH 10.5. The composition of each solution is summarised in table 1.

Table 1 Contents of solubilising and refolding buffers used during lysozyme and preproinsulin inclusion body refolding

Solution	**Solutes**	**pH**
Lysozyme solubilisation solution (urea)	8 M urea, 10 mM DTT, 100 mM Tris-acetate	8.1
Lysozyme solubilisation solution (GdmHCl)	6 M Gdm HCl, 10 mM DTT, 100 mM Tris-acetate	8.1
Preproinsulin solubilisation solution (urea)	4.5 M urea, 10 mM DTT, 365 mM ethanolamine	10.5
Preproinsulin solubilisation solution (GdmHCl)	6 M Gdm HCl, 10 mM DTT, 365 mM ethanolamine	10.5
Lysozyme refolding solution (urea)	0.25 M urea, 100 mM Tris-acetate, 3 mM cysteine	8.1
Lysozyme refolding solution (GdmHCl)	0.188 M Gdm HCl, 100 mM Tris-acetate, 3 mM cysteine	8.1
Preproinsulin refolding solution (urea)	0.25 M urea, 10% hexylene glycol, 20 mM ethanolamine, 288 µM cystamine	10.5
Preproinsulin refolding solution (GdmHCl)	0.35 M urea, 10% hexylene glycol, 20 mM ethanolamine, 288 µM cystamine	10.5

Soluble refolding yield was estimated through measurement of protein concentration after filtering the refolding solution through a 0.22 µm syringe-top filter. The protein concentration of samples before filtration was calculated from measurement of the protein concentration of the solubilising solution before the addition of the refolding buffer.

Table 2 shows the percentage of total protein lost due to filtration, quantified by measurement of protein concentration before and after filtering the solutions through a 0.22 µm poresize filter, with the final soluble yields visualised in figure 4.15. The total protein lost due to aggregation was calculated as the fraction of post-filtered protein concentration divided by pre-filtered protein concentration.

% of total protein lost = (pre-filtered concentration - post-filtered concentration) / pre-filtered concentration x 100

Table 2 Pre- and post-filtered protein concentrations of refolded lysozyme and preproinsulin inclusion body solutions

Sample/ denaturant	**Concentration pre-filtering/ mg/ml**	**Concentration post-filtering/ mg/ml**	**% protein lost due to filtering**
Preproinsulin / Urea	0.77	0.65	15
Preproinsulin /Gdm HCl	0.63	0.48	24
Lysozyme/ Urea	0.36	0.31	14
Lysozyme/ Gdm HCl	0.33	0.21	37.5

Soluble refolded yield was calculated as the protein concentration remaining after filtration divided the pre-filtered protein concentration, to give a final refolding yield percentage. The results are shown in figure 3.17. Lysozyme solubilised in urea has a soluble refolded yield 23% higher than that of Gdm HCl. Similarly, preproinsulin inclusion bodies solubilised in urea have a soluble refolded yield 9% higher than that of inclusion bodies solubilised in Gdm HCl.

Aggregation occurs more readily for both proteins when in solution with Gdm HCl than with urea during dilution refolding. This behaviour correlates with the effect of the denaturant on the protein-protein interactions as quantified by the second virial coefficient measurements. In urea solutions, there is strong protein-protein repulsion between partially unfolded and molten globule states formed during the refolding process. The repulsion is due strong long range electrostatic interactions and weakened short range attractive interactions. On the other hand, Gdm HCl solutions screen protein-protein electrostatic repulsion, and though, short-range attractive interactions are attenuated, aggregation is significantly increased. It is clear from the refolding data presented here that the presence of strong long-range electrostatic repulsive interactions significantly reduces the aggregation propensity of both positively and negatively charged proteins.

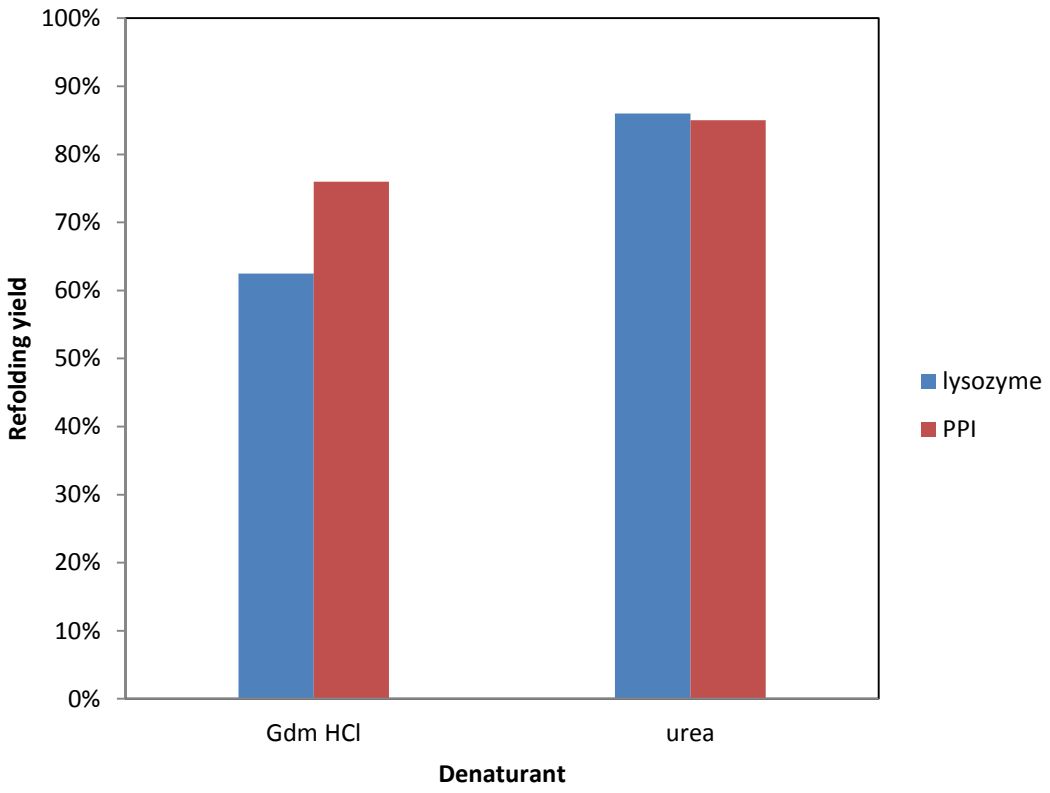


Figure 3.17. The soluble yields of lysozyme and preproinsulin dilution refolding experiments in either urea or Gdm HCl solutions.

3.2.6.2 Dialysis refolding

For the dilution refolding experiments, the pH and ionic strength of the solutions were higher than those used to determine the second virial coefficient of lysozyme. Dilution refolding at pH 4.5 did not lead to enough aggregation to separate the effects of using different denaturants. The protein concentration is so low upon dilution that refolding is preferred to aggregation since refolding is a first order process, whereas aggregation is second or higher order in protein concentration. Thus, we also determined the influence of denaturants of the refolding yield using dialysis refolding, which maintains the protein concentration at the value used in the solubilisation step. The concentration during the refolding step is much higher than occurs with dilution experiments, leading to an increased chance of aggregation due to increased molecular crowding within the sample. For these experiments, lysozyme was solubilised in either 8 M GdmHCl or 8 M urea solutions, with 10 mM DTT at a protein concentration between 20 and 60 g/L and incubated at room temperature for 1 hour. The protein solutions were dialysed using 8000 kDa molecular weight cut-off tubing into 50 mM sodium acetate buffer at pH 4.5 with 10 mM cysteine at a volume-to-volume ratio of 1 to 30 for 16 hours. The concentration of protein after dialysis was measured through absorbance at 280 nm. Samples were taken before and after filtration with a 0.22 µm poresize filter. The filtered product was analysed for aggregates using SEC-MALLS analysis (experimental procedure detailed in chapter 3). All samples were found to be monomer post-filtering, with a single SEC-MALLS peak. Soluble yield was calculated as the concentration of protein post-filtration divided by the concentration of protein pre-filtration within the solution. The resulting soluble refolded yields are shown in figure 3.18, with soluble yield plotted against the final initial pre-filtration protein concentration.

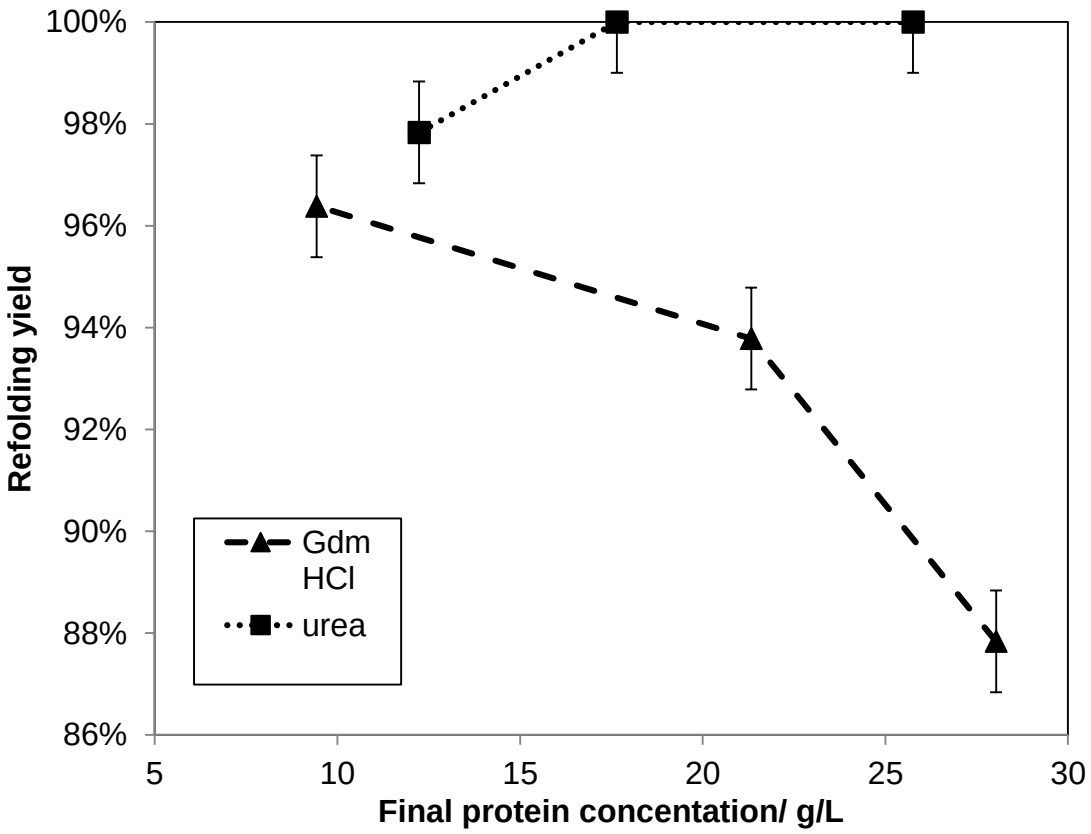


Figure 3.18. The soluble yields of lysozyme dialysis refolding experiments in either urea or Gdm HCl solutions as a function of protein concentration.

The results show increasing protein concentration leads to a higher level of aggregation in Gdm HCl solutions versus those with urea, even though protein aggregation levels are low in solutions of either denaturant at low protein concentrations (<5%). The refolding yields obtained at high protein concentration correlate well with the effect of the denaturant on protein-protein interactions. A higher protein concentration leads to increased molecular crowding within the solution. With increased molecular crowding, attractive protein-protein interactions in Gdm HCl solutions manifest as an increased protein aggregation propensity. On the other hand, refolding yields of almost 100% are observed in urea solutions, indicating little measureable aggregation. The dialysis refolding results reflects the trend observed with the dilution refolding of lysozyme discussed previously, with an increased level of aggregation observed with in the Gdm HCl solutions compared to the low ionic strength urea experiments.

Conclusions

In this chapter, it has been shown through biophysical measurements that urea induces significantly different protein-protein interactions between proteins in the molten globule and partially denatured states found during refolding compared to Gdm HCl, which is translated into significantly different propensities to aggregate. Urea has been shown to not significantly affect the long range electrostatic repulsion between proteins, while at the same time reducing net short-range attraction between the proteins through the denaturing mechanism of the molecule. This leads to an increase in net protein-protein repulsion at all concentrations of urea, including at concentrations where the protein is partially or completely unfolded compared to denaturant-free solutions. This leads to the suppression of aggregation between partially-folded intermediates during the refolding process, and thus an increase in refolding yield compared to Gdm HCl solutions, where electrostatic repulsive interactions are screened by the high ionic strength and thus lead to more attractive protein-protein interactions.

Chapter 4: Short-range protein-protein interactions in urea-NaCl solutions

4.1 Introduction

During the refolding of inclusion bodies, the target proteins are dissolved in solutions with a wide range of denaturant concentrations. Solubilisation is accomplished using solutions of urea and Gdm HCl at concentrations ranging from 5 to 8 M. The concentration of the denaturant is then reduced to between 0.1 M and 0.5 M by either dialysis or dilution refolding, to promote native structure. The final refolding step allows the denatured polypeptide chain to collapse and correctly fold into a native structure. Correct folding is promoted and aggregation is suppressed during refolding by the presence of co-solvents in the dialysis or dilution buffer solution.

In the previous chapter, evidence was presented that showed a urea-induced net increase in protein-protein repulsion in solutions at low ionic strength. The increased protein-protein repulsion in solutions of urea was correlated with lower levels of aggregation during refolding compared to solutions containing Gdm HCl, conditions where protein-protein interactions are much more attractive at comparable denaturant concentrations. However, it is not clear how urea induces the increase in repulsion as it was shown that urea does not significantly increase the electrostatic repulsion or excluded volume force between proteins. It is hypothesised that urea attenuates hydrophobic attractive forces, disrupts the hydrogen bonding network and weakens Van der Waals attraction, leading to an increase in net protein-protein repulsive interactions. These forces are exclusively short-range and “sticky” in nature and are masked by the long range double-layer electrostatic repulsion between proteins.

Work previously performed by Curtis (2002) and Grigsby (2001) have shown that short range interactions can be understood from studies using moderately concentrated salt solutions to screen electrostatic repulsion and enhance other short-range interactions (Curtis et al., 2002; Grigsby et al., 2001). A commonly used salt to probe short-range interactions is sodium chloride. Both sodium and chloride ions lie in the middle of the Hoffmeister series, and therefore have moderate “salting-out” effects on proteins in solution. Chloride ions are preferentially excluded from non-polar protein surfaces, but form weak complexes with positively charged amino acid side chains. These lysozyme-chloride complexes, due to the ion exclusion from the non-polar surfaces, enhance short-range hydrophobic attractive interactions and thus increase net protein-protein attraction.

In this chapter, osmotic second virial coefficients are measured by static light scattering (SLS) for pH 4.5 solutions of lysozyme or preproinsulin over a range of low to high concentrations of NaCl and of urea measurements. The Optim 1000 was used to measure thermally-induced unfolding and aggregation of lysozyme in solutions of NaCl and urea at pH 4.5. Urea was used as a denaturant during dialysis refolding experiments to measure the aggregation of lysozyme in partially folded and denatured states, which were performed with a dialysis buffer containing a range of NaCl concentrations between 0 and 0.5 M.

4.2 Results

4.2.1 Low concentrations of urea induce short-range protein-protein attraction with increasing NaCl concentrations

The second virial coefficient of lysozyme in solutions at pH 4.5 with 50 mM sodium acetate buffer was measured at urea concentrations of 0, 0.1, and 0.25 M and as a function of NaCl concentrations between 0.05 M and 1 M. The SLS results are shown in figure 4.1. In urea-free solutions, increasing NaCl concentration leads to screened electrostatic repulsion and increased attraction between proteins. The B_{22} values become negative in solutions containing 0.1 M NaCl and most protein-protein electrostatic repulsion is screened in solutions with 0.25 M NaCl as the B_{22} values become insensitive to increasing NaCl concentration. The values measured here correlate well with literature values from Curtis et al and others, where decreasing repulsive interactions were observed in solutions of lysozyme with increasing concentrations of NaCl (Curtis et al., 2002; Grigsby et al., 2001).

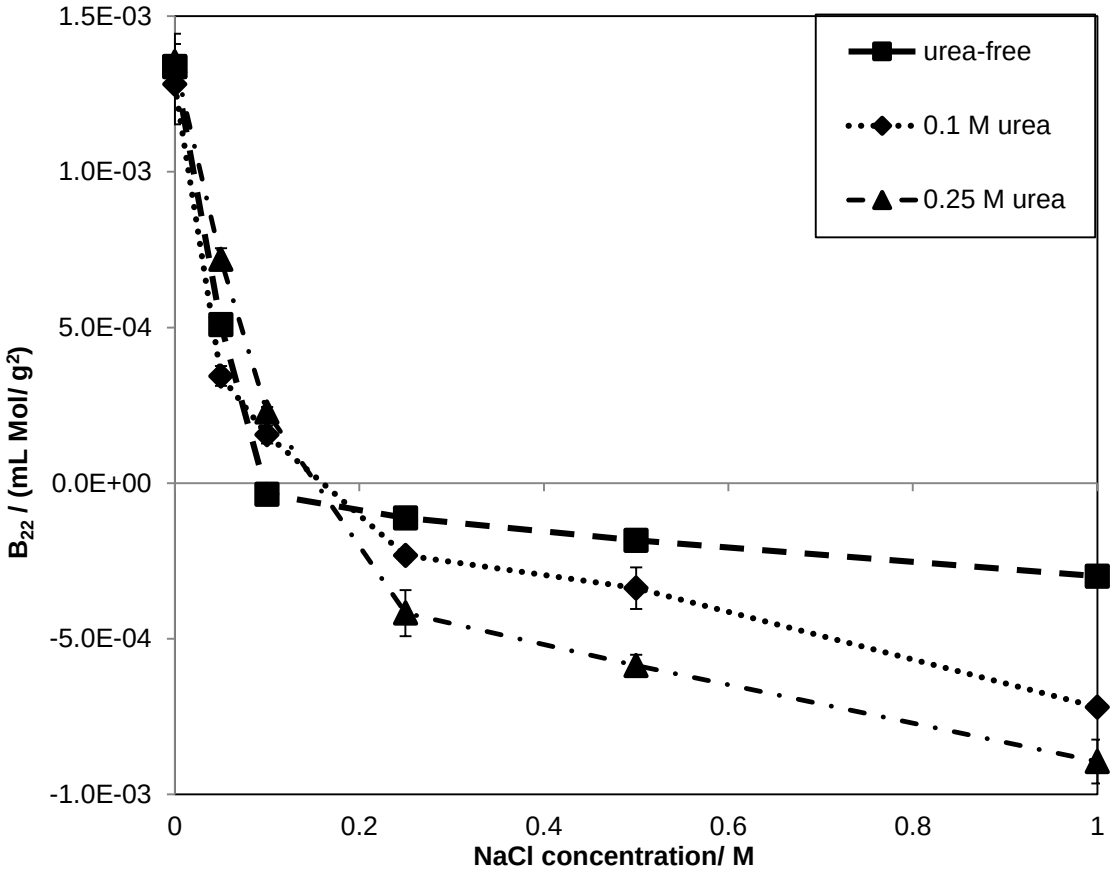


Figure 4.1. The B_{22} values of lysozyme solutions containing 50 mM sodium acetate, 0, 0.1 or 0.25 M urea and between 0 to 1 M NaCl, at pH 4.5.

For solutions of lysozyme and 0.1 M urea, increasing NaCl concentration leads initially to the same decreasing B_{22} values as observed in urea-free solutions, as electrostatic repulsion is screened with the increased ionic strength. However, in solutions at 0.25 M NaCl and 0.1 M urea, the B_{22} of lysozyme becomes more negative than the comparable NaCl-only measurement. This trend continues for solutions with higher NaCl concentrations, as the protein-protein interactions in 0.1 M urea solutions are consistently more attractive than in corresponding urea-free solutions. When the concentration of urea is increased to 0.25 M, the effect of increased attraction at higher NaCl concentration is enhanced, as B_{22} values are more negative than for the corresponding solutions at the same NaCl concentrations with either 0 or 0.1 M urea.

This observed decreased repulsion in solutions when adding urea to moderately concentrated NaCl solutions indicates that, when electrostatic repulsion is screened, low concentrations of urea induce a short-range attraction between proteins. It was hypothesised in the previous chapter that urea increases repulsion between proteins monotonically in low ionic strength solutions by reducing the

hydrophobic surface attraction by binding to hydrophobic side chains and weakening other short-range attractive forces. However, the attenuation of short-range attractive interactions does not appear to occur in solutions at 0.1 and 0.25 M urea concentrations at higher ionic strength. Conformation studies have not detected any secondary structural changes to globular proteins at urea concentrations below 1 M (Hedoux et al., 2010; Timasheff and Xie, 2003). One possibility for the increase in attraction at moderate NaCl concentrations in the presence of 0.1 and 0.25 M urea is that low concentrations of urea lead to an increase in the strength of the salting-out effect of the NaCl. Salting out of proteins from solution occurs at high kosmotropic salt concentrations due to the increased strength of hydrophobic interactions. A combination of sodium and chloride ions act synergistically with urea to enhance short-range protein-protein attraction, though an as-yet unknown mechanism.

4.2.2 High ionic strengths between 0 and 6 M urea induce a wide range of interactions in lysozyme

The second virial coefficient of lysozyme was measured in solutions at pH 4.5 with 50 mM sodium acetate buffer with either 0.1, 0.25 or 0.5 M NaCl between 0 and 4 M urea to investigate how ionic strength changes the effect of urea on protein-protein interactions, the results are shown in figure 4.2. In solutions at 100 mM NaCl, long range electrostatic repulsive interactions are predominantly screened and the protein-protein interactions are slightly attractive as the B_{22} value of -3.6×10^{-5} mL Mol/g^2 is small and negative. A similar pattern of protein-protein interactions is observed with increasing urea concentration in solutions at 100 mM NaCl to that in low ionic strength. Interactions become more repulsive with increasing urea concentration, leading to a maximum in solutions at 2 M urea, where the B_{22} value is equal to 3.6×10^{-4} mL Mol/g^2. The protein-protein repulsion observed in solutions at 100 mM NaCl is significantly weaker at all urea concentrations than that observed at low ionic strength as a result of the loss of strong electrostatic repulsive forces due to charge screening. The data indicates that, when long-range electrostatic repulsion is retained between the proteins, urea's effect of increasing protein-protein repulsion is retained. From previously discussed data regarding the zeta potential of lysozyme in low ionic strength urea, it has been shown that urea does not directly increase repulsion through increasing electrostatic charge. However, the presence of electrostatic repulsion may contribute towards protein stability and therefore the formation of partially-folded intermediates in the presence of low concentrations of urea may not occur.

In solutions at NaCl concentrations of 250 mM, significantly different protein-protein interaction behaviour with increasing urea concentration is observed when compared to solutions containing 100 mM NaCl. The B_{22} of lysozyme in solution with 250 mM NaCl is negative, equal to -1.4×10^{-4} ml Mol/ g^2, indicating mildly attractive interactions, as repulsive electrostatic interactions are screened leaving

only short-range attractive protein-protein interactions. In solutions with increasing urea concentrations between 0 and 0.5 M, lysozyme interactions become more attractive, with the value of B_{22} decreasing to a minimum of -3.8×10^{-4} ml Mol/ g^2 in solutions at 0.5 M urea. As shown previously, attractive protein-protein interactions occur in solutions with low concentrations of 0.1 and 0.25 M urea when electrostatic repulsive interactions are screened. However, above 0.5 M urea, the B_{22} values become more positive, returning to net repulsive interactions in solutions between 1 and 2 M urea. This shape of the B_{22} verses urea curve is similar to that exhibited by solutions of lysozyme as a function of Gdm HCl concentration (Liu et al, 2005), where, after an initial decrease in repulsion with increasing denaturant concentration, interactions become more repulsive above 0.5 M. In solutions at 250 mM NaCl, there is almost complete screening of long-range electrostatic repulsion such that changes in protein-protein interactions brought about by urea are due to changes in short-range interactions. Thus, it is hypothesised that net attractive protein-protein interactions are weakened in moderate urea solutions through attenuation of attractive hydrophobic interactions and a weakening of other attractive short-range interactions.

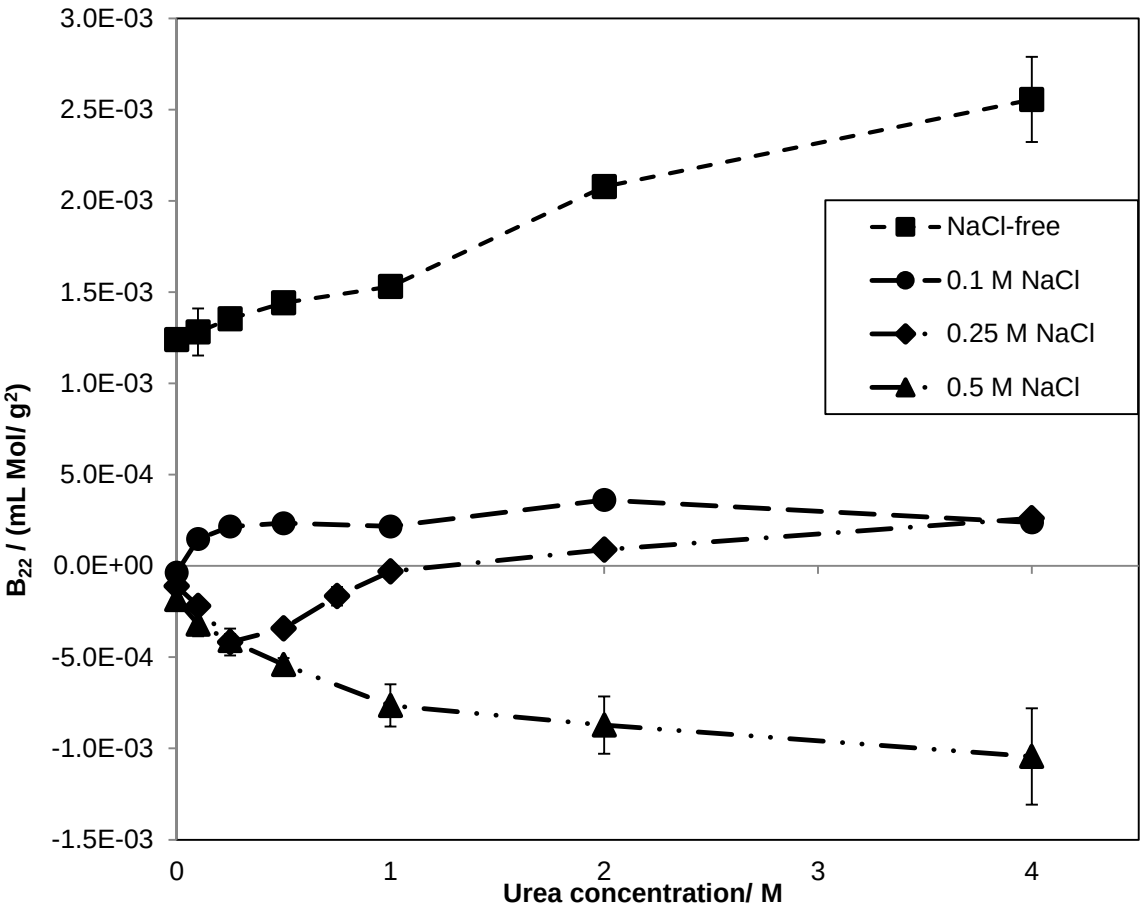


Figure 4.2. The B_{22} values of lysozyme solutions containing 0, 0.1, 0.25 or 0.5 M NaCl and between 0 and 4 M urea, with 50 mM sodium acetate buffer at pH 4.5

In solutions at 500 mM NaCl, the same initial decrease in repulsion observed at 250 mM NaCl is observed with increasing concentrations of urea. However, in solutions with 500 mM NaCl the interactions become more attractive with increasing urea concentrations above 0.5 M, unlike what occurs in 250 mM NaCl solutions. The second virial coefficient decreases continuously with increasing urea for solutions containing up to 4 M urea. Overall, this behaviour is a significant deviation from what is observed of solutions of medium to high concentrations of urea at lower NaCl concentrations. Thus in solutions of 500 mM NaCl, the short-range attraction is not being attenuated by urea and instead increasing concentrations of urea appears to increase short-range attractive protein-protein interactions. It is hypothesised that a high concentration of sodium chloride could enhance the strength of hydrophobic interactions, nullifying the attenuation of attractive short-range interactions caused by the binding of urea to the surface of the protein. An increasing concentration of urea above a certain ionic strength appears to leads to an increase, rather than a decrease, in attractive short-range interactions, suggesting that the presence of urea increases the strength of the short-range protein-protein interactions due to the ionic strength of the solution.

Second virial coefficient measurements of lysozyme in NaCl solutions show that, once electrostatic repulsive interactions are screened at 250 mM NaCl, the short-range protein-protein interactions lose sensitivity to the ionic strength of the solution, with relatively little increase in attractive interactions between 0.25 and 1 M NaCl. The data presented here suggests that the presence of moderate concentrations of urea (1-4 M) leads to an increase in the sensitivity of the short-range protein-protein interactions to the ionic strength, with relatively small changes in the ionic strength leading to significant increases in protein-protein attractive interactions that does not occur in denaturant-free solutions and is not attenuated by the solubilising effects of the denaturant.

4.2.3 Calculation of the short-range protein-protein interactive force of lysozyme in varying NaCl concentrations in urea solutions

The contribution of the short-ranged interactive force to the second virial coefficient, B_{22sr}, can be calculated from the measured B_{22} values using equation (25).

$$B_{22} = B_{22elec} + B_{22exl} + B_{22sr} \qquad \text{Equation (25)}$$

The repulsive electrostatic contribution, B_{22elec}, to B_{22} is calculated as 2.22×10^{-4} mL Mol/ g^2 for solutions of 100 mM NaCl, using equations (9) and (11). The B_{22elec} is close to zero at 250 mM NaCl due to the screening effect of the high ionic strength. The excluded volume force, B_{22ex}, is a constant

of 3.24 x 10^{-4} mL Mol/ g^2, as no change in particle size was observed for lysozyme in solutions at pH 4.5 for urea concentrations between 0 and 2 M as described in the previous chapter.

Figure 4.3 shows the short-range interactive contribution, B_{22sr}, of lysozyme in solution with 0, 100 mM, or 250 mM NaCl in solution with between 0 and 4 M urea concentrations as calculated from equation (25). Also included are the calculated B_{22sr} of lysozyme in solution with between 0 and 4 M Gdm HCl from figure 4.12 from the previous chapter. In solutions at low urea concentrations (<1 M), the short-range interactions observed at 0, 100 mM and 250 mM NaCl follow a similar pattern. The interactions are initially attractive and protein-protein repulsion increases with increasing urea concentration. However, in NaCl-free and 250 mM NaCl solutions, the net short-range interactions are more attractive than in 100mM NaCl solutions, and the same dip in B_{22} value as observed at 250 mM NaCl solutions is not observed between 0 and 0.5 M urea in the absence of NaCl. In 100 mM NaCl solutions, the B_{22sr} is the least negative observed at low urea concentrations, and increases in value more rapidly than 0 or 0.25 M NaCl solutions, following a trend similar to that of Gdm HCl solutions.

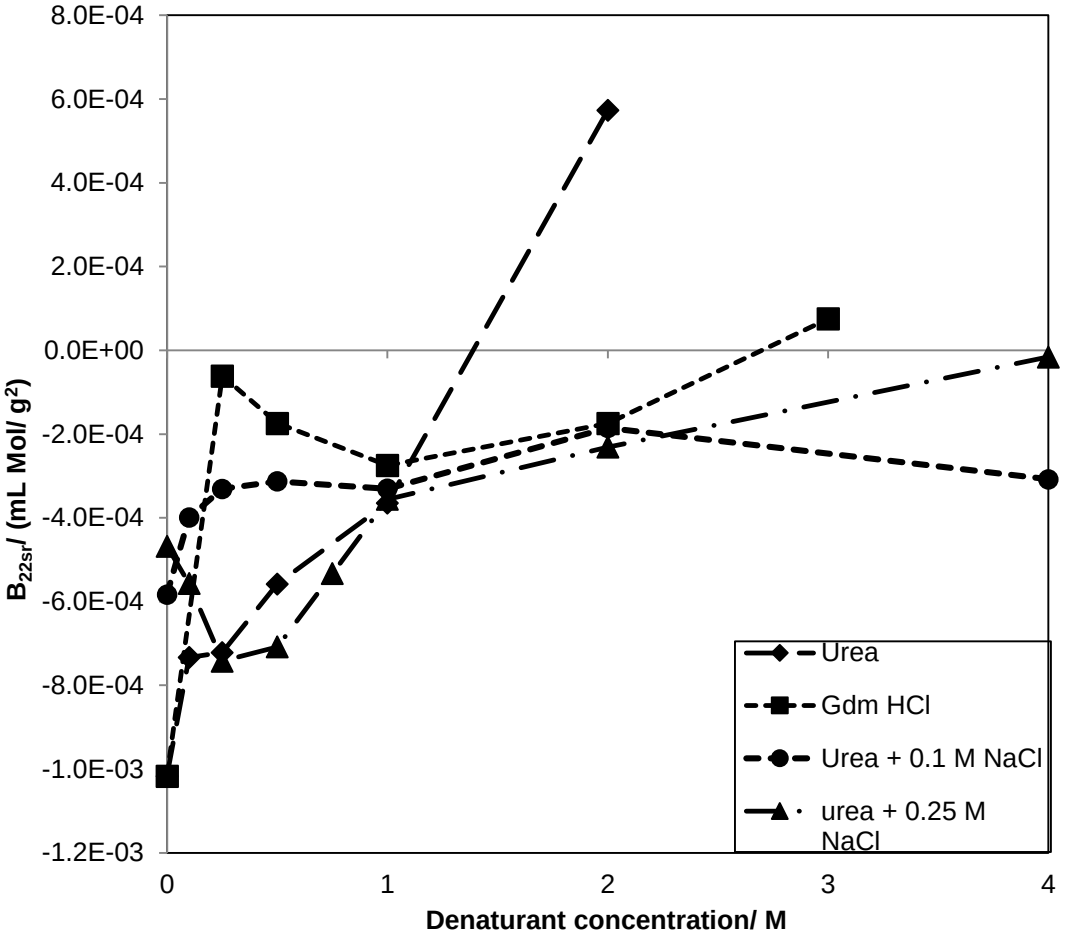


Figure 4.3. The calculated B_{22sr} values of lysozyme solutions with 50 mM sodium acetate buffer at pH 4.5 containing 0, 0.1 or 0.25 M NaCl and between 0 and 4 M urea, or between 0 and 3 M Gdm HCl

For solutions containing 1 M urea or 1 M Gdm HCl, it is observed that all calculated B_{22sr} values are roughly equal. This suggests that at this concentration of denaturant, the net effects of denaturant on short-range protein-protein interactions are equal at all ionic strengths. In 2 M urea or 2 M Gdm HCl concentration solutions, the values for B_{22sr} are also equal, with the exception of low ionic strength urea solutions, for which B_{22} is equal to 6 x 10^{-4} mL Mol/ g^2. In 100 mM and 250 mM NaCl solutions, short-range interactions remain attractive and are similar in magnitude to those measured at comparable Gdm HCl concentrations. The reduced B_{22sr} values in moderate ionic strength solutions suggests that when electrostatic interactions are screened with high ionic strength, urea and Gdm HCl have a more similar effect on short range interactions compared to low ionic strength urea solutions. Also, while repulsive short-range interactions between proteins are observed in low ionic strength urea solutions at 2 M and Gdm HCl solutions at 3 M, the increase in short-range protein-protein repulsion in high ionic strength urea solutions is slightly smaller. Interactions do not become repulsive until a urea concentration of 4 M in 250 mM NaCl solutions. This leads to several conclusions. Firstly, in all ionic strength conditions, short-range protein-protein attraction is being significantly attenuated by either urea or Gdm HCl as hydrophobic attractive interactions are reduced in strength and other short-range attractive forces are weakened. In high ionic strength urea solutions above 1 M and Gdm solutions above 1 M, the strength of short-range repulsive interactions are weaker than in low ionic strength urea solutions, suggesting that in high ionic strength solutions, urea is less effective at reducing the strength short-range attractive interactions. Alternatively, electrostatic protein-protein interactions may be significantly altered by the presence of high concentrations of urea. In the previous chapter, changes in the net charge of lysozyme at low pH and ionic strength were detected between 0 and 2 M urea, and the presence of varying concentrations of NaCl may also influence this change. This change in charge and electrostatic interactions may therefore impact the calculation of B_{22sr} interactions. Overall, however, this has significant implications for how DLVO theory is used to model protein-protein interactions in various salt concentration and solution conditions.

4.2.4 The effect of moderate NaCl concentration on preproinsulin in 0 – 4 M urea induces comparable changes in protein-protein interactions

In a previous chapter, data was presented that showed an increase in protein-protein pair repulsion for preproinsulin in urea solutions when compared to low ionic strength urea-free conditions at pH 9.0. This followed a similar pattern of increased repulsion with increasing urea concentration exhibited by both lysozyme and ribonuclease A. Static light scattering was used to determine the B_{22} value of

effect of significantly increasing the aggregation propensity of proteins in non-denaturing concentrations of urea.

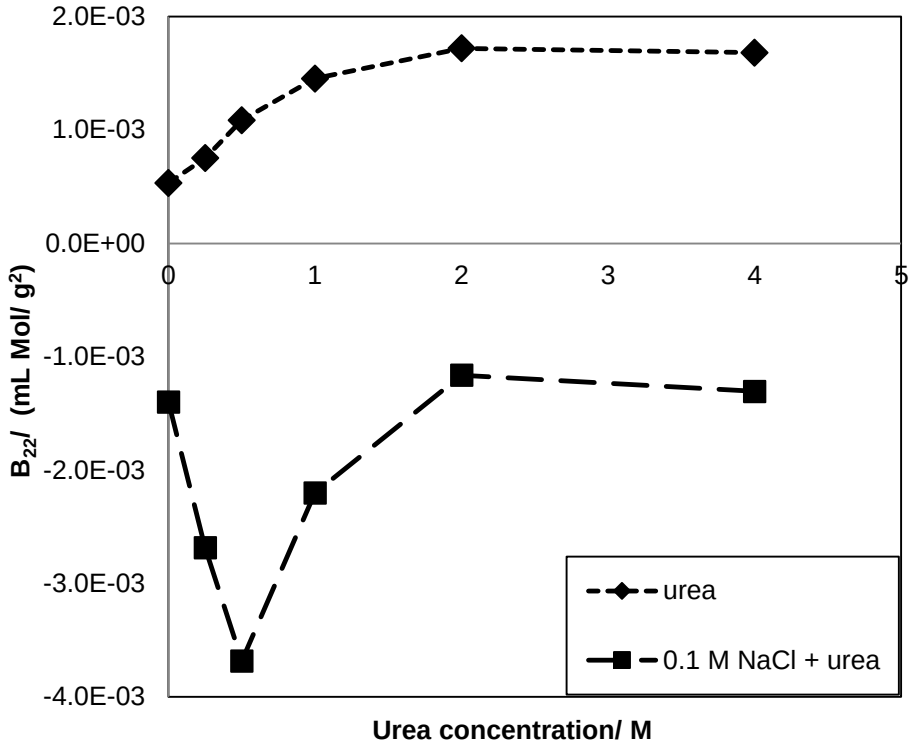


Figure 4.4. The measured B_{22} values of preproinsulin solutions with 20 mM TRIS buffer at pH 9.0 containing 0 or 0.1 M NaCl and between 0 and 4 M urea.

The results are presented in figure 4.4, and show that between 0 and 0.5 M urea, the interactions between preproinsulin molecules become more attractive. Above 0.5 M urea concentration, protein-protein interactions become less attractive, but remain net repulsive in solutions at urea concentrations up to 4 M. This pattern of interactions is similar to the pattern observed for the B_{22} measurements of lysozyme at 0.25 M NaCl with urea concentrations ranging from 0 to 4 M, with short-range attractive protein-protein interactions attenuated above 0.5 M. These results provide further evidence that the mechanisms through which urea alters interactions between proteins, both attractive and repulsive, are found in at least two proteins and may be broadly universal in nature.

4.2.5 Refolding of lysozyme into varying concentrations of NaCl leads of aggregation patterns that follow B_{22} measurements

Dialysis refolding experiments were performed to investigate the effect of ionic strength on lysozyme aggregation behaviour. Lysozyme was solubilised at a concentration of 10 g/L in 8 M urea and 10 mM DTT solution with 50 mM sodium acetate buffer at pH 4.5 by stirring the solution for 1 hour. The solution was then dialysed using 8000 Da tubing into a refolding buffer at a volume ratio of 1:30. The refolding buffer solution contained 50 mM sodium acetate, 10 mM cysteine and NaCl at a concentration ranging between 0 and 0.5 M at pH 4.5. The refolding yield is calculated from the protein concentration of the solution before and after filtering with a 0.22 µm syringe-top filter through measurement of UV absorbance at 280 nm. The pre-filtered protein concentration of the solution was calculated by re-solubilising a sample of the refolded protein solution with the 8 M urea denaturing buffer, followed by measurement of the protein concentration of the denatured protein solution. The yield is calculated by dividing the post-filtered protein concentration by the pre-filtered protein concentration and expressed as a percentage of remaining soluble protein.

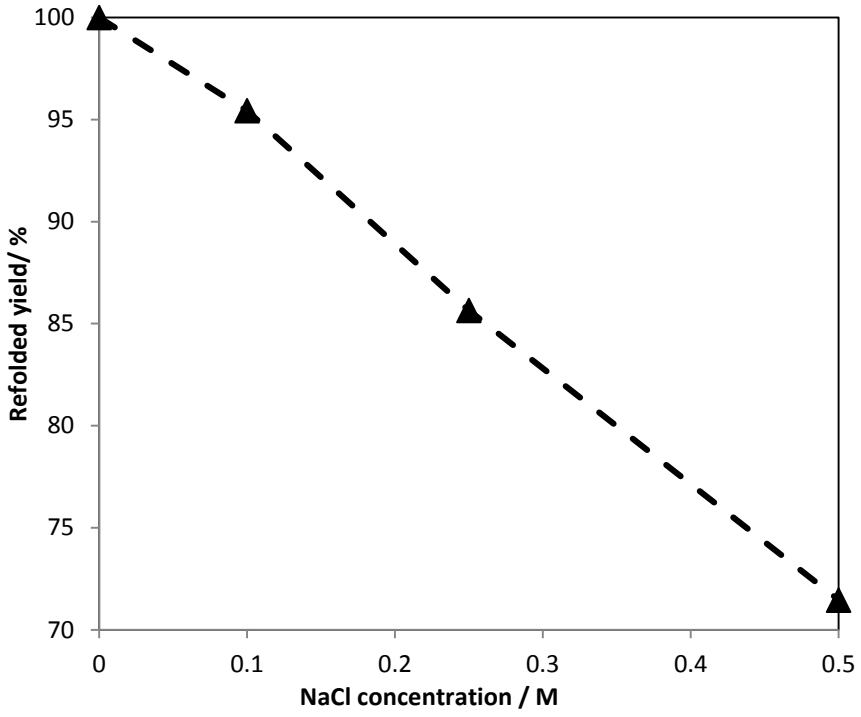


Figure 4.5. The soluble refolding yield of lysozyme solutions denatured with 8 M urea and 10 mM DTT at pH 4.5 and dilaysed into refolding solutions containing between 0 and 0.5 M NaCl

The results shown in figure 4.5 reveal a significant increase in insoluble aggregate formation when using a refolding solution containing NaCl concentrations of 0.1 M and above, with aggregation

increase monotonically with increasing NaCl concentrations. This finding correlates well with the B_{22} data described previously, as there is a significant reduction in protein-protein electrostatic repulsion in solutions with 100 mM NaCl concentrations and higher. The solutions conditions during the dialysis refolding experiment change from high to low urea concentration. The protein-protein interactions will thus follow the behaviour corresponding to the B_{22} data shown in figure 5.2. Protein-protein interactions of lysozyme in solution with 0.5 M NaCl are more attractive than in solutions at all urea concentrations for 0.1 and 0.25 M NaCl conditions. This may explain why insoluble protein aggregation is highest when the protein solution is dialysed into 0.5 M NaCl solutions. In solutions of 250 mM NaCl, protein-protein interactions are more repulsive at all urea concentrations compared to 0.5 M NaCl solutions, and therefore a lower rate of aggregation is observed. However, protein-protein interactions in solutions 0.25 M NaCl are attractive when in solution with 0.1 to 0.5 M urea, suggesting that aggregation is most likely to occur at these urea concentrations during refolding. In solutions of 0.1 M NaCl, protein-protein interactions remain repulsive at all urea concentrations, and thus little aggregation is observed compared to higher ionic strength solutions.

SEC-MALS analysis of the post-filtration soluble protein solution for each refolding experiment was used to identify the soluble aggregate distribution (data not shown). A single peak was obtained for the SEC-MALLS analysis for the soluble protein solutions obtained when refolding into solutions of 0, 0.1 or 0.25 M NaCl, indicating that all soluble protein is correctly folded and monomer under these refolding conditions. Multiple peaks were obtained when analysing the solutions after refolding at 0.5 M NaCl, reflecting the presence of different sized soluble aggregates in the refolded protein solution alongside monomer refolded protein. The refolded yield of lysozyme in 0.5 M NaCl may therefore be significantly lower than that shown in the graph above. This suggests that soluble aggregate formation is controlled by short-range attractive interactions, which are enhanced with increasing ionic strength concentrations.

4.2.6 Temperature-induced aggregation of lysozyme in urea-NaCl solutions

The increased strength of attractive protein-protein interactions by urea at high ionic strength in comparison to low ionic strength solutions has been hypothesised to derive from increased sensitivity to the ionic strength of the solution due to the presence of urea. To determine whether the presence of varying concentrations of NaCl and urea has a destabilising effect and increases aggregation propensity of lysozyme, temperature-induced unfolding experiments were performed. The Optim 1000 system was used to monitor the unfolding and aggregation behaviour as a function of temperature between 25 and 90ºC for lysozyme in solutions with 0, 0.25 or 0.5 M NaCl and urea concentrations between 0 and 4 M at pH 4.5 with 50 mM sodium acetate buffer. Florescence spectroscopy and static light scattering are used to measure the temperatures of unfolding and aggregation, respectively.

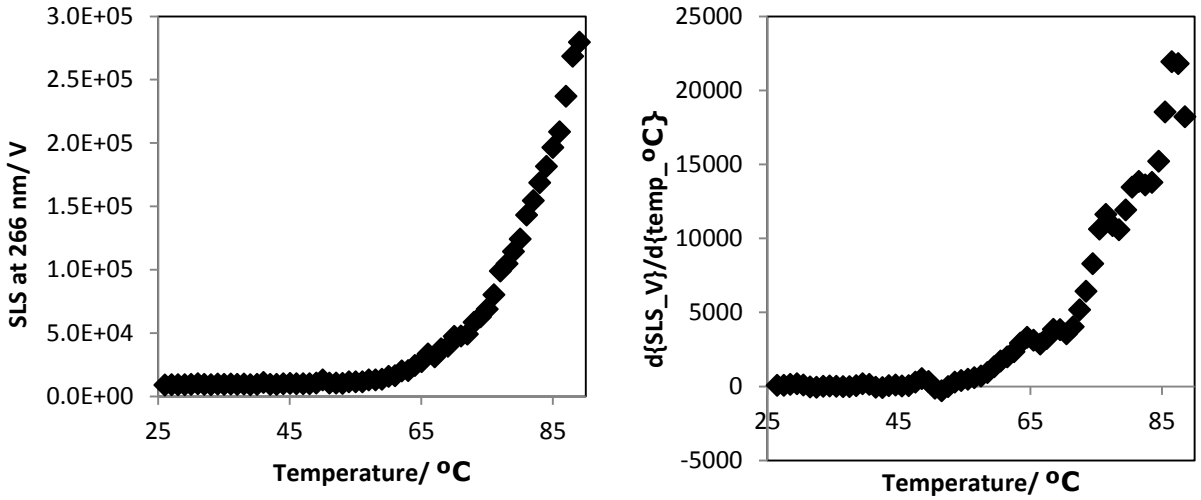


Figure 4.6. Temperature-induced aggregation monitored through SLS of lysozyme in solution with 1 M urea and 0.5 M NaCl. (a) The static light scattering reading (V) versus temperature. (b) The derivative of the SLS voltage with respect to temperature plotted versus temperature

Lysozyme in solution with 1 M urea and 0.5 M NaCl is used here as an example to demonstrate how the aggregation temperature is determined from the static light scattering measurement shown in figure 4.6.a. At low temperatures, the SLS voltage remains constant indicating the protein is stable, and in a non-aggregated state. SLS begins to steadily increase above a temperature of 55ºC. Above 65ºC, SLS readings increase more rapidly, indicating higher order aggregate formation. The derivative of the SLS voltage with respect to temperature is plotted versus temperature in figure 5.6.b for the same data set to determine the temperature at which colloidal stability changes from monomer- to aggregate-forming. The primary peak is used to estimate the aggregation temperature, T_{agg}, in this case, 63.5ºC.

The unfolding temperature, T_m, of the protein is measured from the change in intrinsic fluorescence of the protein as a function of temperature. The intrinsic fluorescence of the protein is due to the tryptophan residues within the protein. Fluorescence changes upon unfolding as the environment of the tryptophan groups changes from either hydrophobic core to being exposed to the solvent conditions. The peak fluorescence occurs at a wavelength of 300 nm for a folded protein, while the peak occurs at approximately 330 nm when all the fluorophores are exposed to the solvent. The barycentric Mean (BCM) of the fluorescence spectra measures the average wavelength (centre of mass) of the protein sample emission peak, measured in nm. The BCM is defined as:

$$BCM = \frac{\Sigma_\lambda \lambda I(\lambda)}{\Sigma_\lambda I(\lambda)} \qquad \text{Equation (26)}$$

Where $\sum_\lambda$ is all wavelengths between 300 and 450 nm, λ is the wavelength of light in nm, I(λ) is the fluorescence intensity at the wavelength.

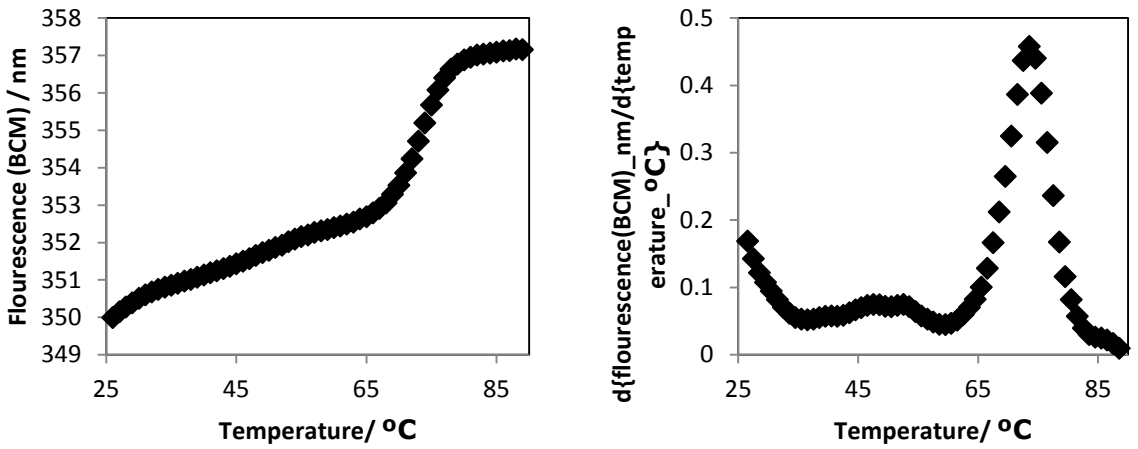


Figure 5.7. Temperature-induced unfolding monitored through fluorescence of lysozyme in solution with 1 M urea and 0.5 M NaCl. (a) BCM (nm) of sample versus temperature. (b) The derivative of the BCM with respect to temperature plotted versus temperature

Changes in the intrinsic fluorescence of the protein will change the BCM value, with an unfolded protein having a higher BCM than a folded state. An example of the fluorescence BCM data measured for lysozyme in solution with 1 M urea and 0.5 M NaCl is shown in figure 5.7.a. The BCM fluorescence of the sample begins at 350 nm at low temperatures, and increases with increasing temperature. The fluorescence reaches a plateau at 357 nm, indicating a change in the conformational equilibrium of the protein in the bulk solution. This data therefore indicates that the fluorescent BCM of the folded state of lysozyme is 350 nm and the unfolded state fluorescent BCM is 357 nm. An example plot for the differential of the peak fluorescence with respect to temperature is plotted against temperature for the same data shown in figure 5.7.b. For a two-state unfolding transition, the melting temperature corresponds to the wavelength where the differential is a maximum.

The measured T_m values for lysozyme in solutions of 50 mM sodium acetate buffer at pH 4.5 with concentrations of urea between 0 and 4 M and NaCl concentrations between 0 and 0.5 M are shown in figure 4.8. For all sodium chloride concentrations, the T_m of lysozyme reduces with increasing urea concentration as expected. Adding urea reduces the free energy of unfolding by weakening intramolecular interactions and reducing the energy required to hydrate hydrophobic residues and unfold the protein. In solutions with less than 0.5 M urea concentrations, the T_m for lysozyme with 0.25 M NaCl is greater than corresponding urea concentrations at 0 or 0.5 M NaCl. This suggests that at

moderate ionic strength and low urea concentration, the protein is structurally stabilised. At 2 and 4 M urea concentration, however, it can be seen that lysozyme unfolds at lower temperatures in the presence of 0.25 M NaCl compared to both NaCl-free and 0.5 M NaCl solutions. This shows a non-linear pattern of influence of NaCl on protein stability at both high and low urea concentrations.

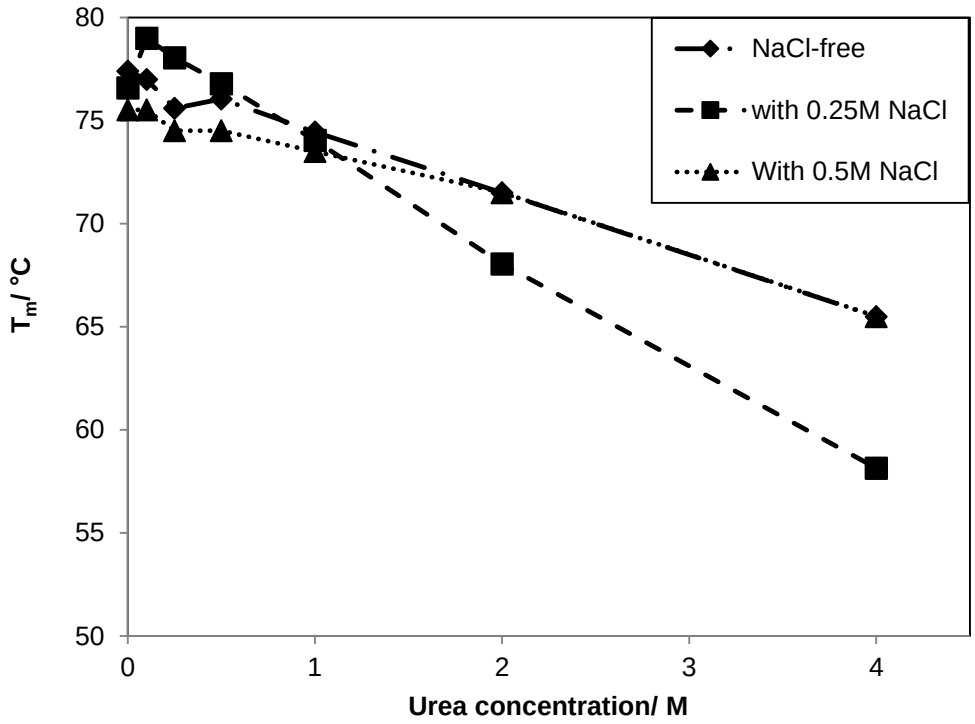


Figure 4.8. T_m values for lysozyme in solutions with between 0 and 4 M urea and 0, 0.25 or 0.5 M NaCl.

The T_{agg} values for lysozyme in solutions containing 0.25 M or 0.5 M NaCl concentrations and between 0 and 4 M urea are shown in figure 5.9. Aggregation did not occur in low ionic strength solutions at any urea concentration and aggregation was not observed in solutions of 0.5 M NaCl with 2 or 4 M urea (results not shown). In urea-free solution, lysozyme aggregates at relatively low temperatures in solutions containing either 0.25 or 0.5 M NaCl concentrations (38ºC for 0.25 M NaCl and 30ºC for 0.5 M NaCl).The increased aggregation propensity is expected, as high salt concentrations significantly reduce long-range repulsion and promote hydrophobic interactions that lead to aggregation. Under these conditions, lysozyme does not unfold until above 70ºC, aggregation is due to the association of natively folded proteins.

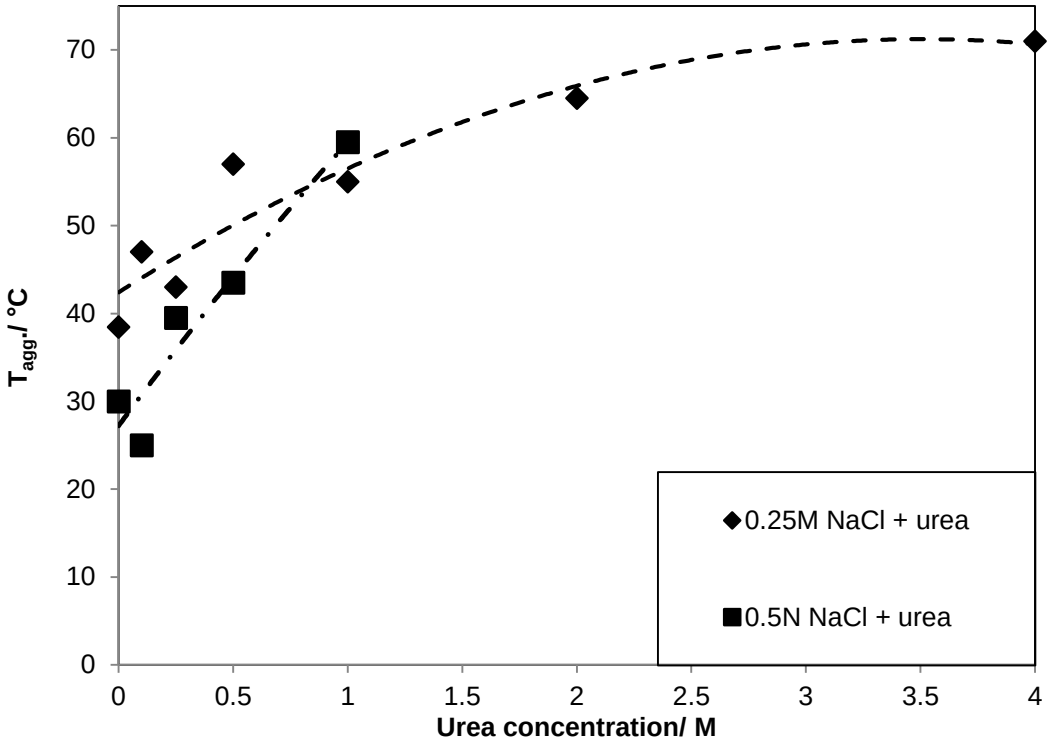


Figure 5.9. T_{agg} values for lysozyme in solutions with between 0 and 4 M urea and either 0.25 or 0.5 M NaCl.

With increasing urea concentration, the aggregation temperature of lysozyme in NaCl solutions increases to 59ºC for 0.5 M NaCl and 55ºC for 0.25 M NaCl at 1 M urea. This behaviour contradicts the second virial coefficients measured at comparable NaCl and urea concentrations, which show an initial increase in attractive protein-protein interactions with increasing urea concentration. The increasing values for aggregation temperature with increasing urea concentration suggest that urea increases protein solubility regardless of ionic strength, and thus may inhibit the formation of higher order aggregates that would lead to higher levels of detectable light scattering. Alternatively, the effects of urea on protein-protein interactions are temperature-dependent, and thus alter the second virial coefficient of lysozyme differently depending on the temperature of the solution.

In solutions below 1 M urea, the aggregation temperature for lysozyme is higher in the solution at lower ionic strength 0.25 M NaCl than that of 0.5 M NaCl. This is an expected result, as higher NaCl concentration promotes increased hydrophobic interaction strength, and thus lowers the temperature at which the protein will aggregate. Above 1 M urea, the trend is reversed, with lysozyme in 0.25 M NaCl solutions aggregating at all urea concentrations, while in 0.5 M NaCl solutions, aggregation is not observed above 1 M urea concentration. This suggests that the effect of NaCl concentration on protein aggregation in the presence of urea is non-monotonic. The lower aggregation temperatures observed for lysozyme in solutions with 0.25 M NaCl at 2 and 4 M urea compared to in 0.5 M NaCl

solutions at the same urea concentrations may be linked to the melting temperature of the protein which is lower in solutions of 0.25 M NaCl. Also, the protein-protein interactions characterised through B_{22} values cannot be correlated across the temperature range measured here. Urea has been shown to affect the hydrophobic and hydrogen bonding interactions which different dependencies on temperature. Therefore, with increasing temperature, it is unclear how urea will influence protein-protein interactions in the presence of sodium chloride.

Literature papers show that high sodium chloride concentrations can have a stabilising effect on the structure of proteins through attenuating the destabilising effects of charged side chains and increasing the strength of hydrophobic interactions, but a less clear picture is painted with the data presented here and at a smaller concentration range of NaCl (0 to 0.5 M) compared to literature values (0 – 2 M NaCl). Bye and Falconer (2013) have shown that sodium ions at high concentration can have a slight destabilising effect on lysozyme temperature-induced unfolding, as low charge density ions can weakly bind the surface of the protein and reduce the unfolding free energy (Bye and Falconer, 2013). However, this appears to be only the case for 0.25 M NaCl solutions measured here, and the effect is negated with increased NaCl concentration. The measured T_m values for lysozyme solutions between 0 and 0.5 M NaCl shows that the presence of sodium chloride does not have a linear destabilising effect on the tertiary structure of lysozyme in the presence of urea and does not appear to increase the population of aggregate-prone partially unfolded states.

4.3 Conclusions

In the data described above, it is clear that the protein-protein interactions induced by urea can be greatly affected by the ionic strength of the solution. It has been shown that at moderate to high urea concentration, repulsive interactions between proteins can change dramatically over a relatively small NaCl concentration range of between 0.1 M and 0.5 M, leading to strongly attractive short-range interactions at high ionic strength that are far greater in magnitude than are observed in denaturant-free solutions at comparable ionic strengths. It has also been shown that this change in interaction behaviour directly influences the refolding yield and aggregation behaviour of the protein of interest. Temperature-induced denaturation have shown that the influence of NaCl on protein stability in the presence of urea is non-linear to the concentration of both the salt and denaturant, suggesting a complex interplay of interactions during refolding as the protein moves from high to low denaturant concentrations. This has significant implications for the design of refolding experiments for all proteins, as it shows that the choice of denaturant and refold excipient as a pairing has a vital role in influencing protein-protein interaction behaviour.

Chapter 5: The influence of refolding excipients and denaturant combinations on protein-protein interactions and aggregation

5.1 Introduction

In previous chapters, the influence of denaturants, pH and ionic strength on protein-protein interactions and protein aggregation were discussed and investigated. However, most refolding formulations contain excipients that promote correct protein folding and reduce aggregation during the later stages of denaturant removal during refolding. Excipients are necessary for optimising refolding yield for many inclusion body-forming proteins; absence of an appropriate folding-promoter or aggregation-suppressor in a dilution or dialysis buffer can lead to near total loss of product to re-aggregation for certain proteins, such as observed for citrate synthase (Buchner et al., 1991; Mishra et al., 2005).Refolding excipients can be classified into two broad categories; stabilising agents and solubility enhancers (Timasheff, 2002; Timasheff and Arakawa, 1988). Stabilising agents increase the structural stability of proteins through promotion of correct folding. The promotion of correct folding occurs through the preferential exclusion of the excipient from the surface of the protein, leading to preferential hydration of the protein surface and promotion of hydrophobic attraction between aliphatic residues. This shifts the energetic equilibrium towards structurally compact proteins and increases the energy of unfolding. Solubility enhancers are often preferentially bound to the surface of the protein, and increase the solubility of the protein, rather than promoting structural stability. In this way, urea and Gdm HCl are considered solubility enhancers, as protein solubility is increased in solutions with these molecules. Solubility enhancers increase protein-protein repulsion, and thus the effects are sensitive to solution conditions such as pH and ionic strength, which in turn can influence the folded state, surface charge and solvent-accessible surface area properties of the protein.

A common molecule used in suppressing aggregation is arginine. Arginine is a naturally occurring amino acid, with a 3-carbon aliphatic side chain ending in a Gdm group, which is positively charged at physiological pH and lower. Arginine has been shown to slightly reduce the stability of proteins, owing to the influence of the Gdm group (Majumdar et al., 2013; Manikwar et al., 2013), but at the same time, arginine is an aggregation suppressor at high concentrations between 0.5 and 2 M in solutions at high pH (> pH 8.0) (Arakawa and Tsumoto, 2003; Tsumoto et al., 2003). This behaviour has been attributed to the preferential interactions of arginine with aromatic amino acid residues through hydrogen bonding and Van der Waals forces (Wen et al., 2015). Favourable interactions will lead to a decrease in the energy of solvation for the protein, which could be linked to the aggregation suppression ability of arginine. Arakawa produced a comprehensive review in 2007 of the major factors that contribute towards the suppression of protein-protein interactions and aggregation by arginine binding (Arakawa et al., 2007a).

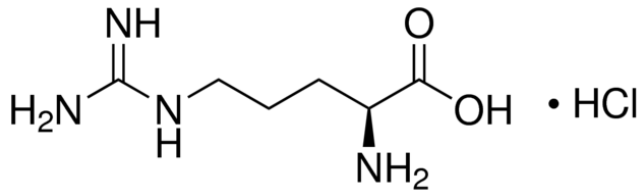


Figure 5.1. Structure of Arg HCl

Recent studies suggest that, though arginine does weakly bind to the surface of proteins, aggregation is also suppressed through the formation of arginine-arginine clusters in the solution that prevent protein-protein collisions through an excluded volume effect (Schneider et al., 2011). Arginine is also a strong base, and is paired with an anion to produce a salt such as Arg HCl when used at low pH (Arakawa and Tsumoto, 2003). However, adding high salt concentrations to a refolding buffer can enhance aggregation, as the increased ionic strength will screen electrostatic interactions between proteins and thus decrease net repulsion. The presence of high salt concentrations does however allow for the characterisation of changes in the short-range interactions induced by arginine that are otherwise masked by long-range electrostatic repulsion.

Another commonly used molecule hexylene glycol (also known as 2-methyl 2,4-pentanediol) is effective in improving refolding yield and is used to aid in protein crystallisation. Hexylene glycol is an amphipathic alcohol that is preferentially excluded from the surface of the protein, leading to increased hydration of the protein surface (Kita et al., 1994; Pittz and Timasheff, 1978). However, hexylene glycol has been shown to bind to aliphatic side chains and reduces the solvent accessible surface area significantly by displacing surface-bound water molecules (Anand et al., 2002). This surface binding is believed to suppress aggregation by reducing the fraction of solvent-exposed hydrophobic surface area and stabilises denatured states (Arakawa, 1994). In this way hexylene glycol can be both an aggregation suppressor and structurally stabilising agent, depending on the concentration at which it is used and the denatured state of the protein.

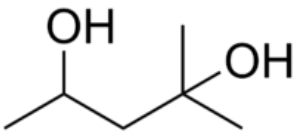


Figure 5.2. Structure of hexylene glycol

Sucrose is a disaccharide molecule that, like many poly- or monosaccharides, is used for stabilising protein solutions throughout bioprocessing, including during formulations and in refolding operations. Sucrose is preferentially excluded from the surface of proteins (Arakawa and Timasheff, 1982; Kendrick et al., 1997; Nicoud et al., 2015; Timasheff, 1991). Sucrose increases the protein structural stability and activation energy of unfolding instead of increasing repulsion between proteins through altering the surface properties of the protein through binding (Arakawa and Timasheff, 1982; Kendrick et al., 1997; Timasheff, 1998, 2002). Preferential exclusion of sucrose from the surface of the protein leads to unfavourable protein-solvent interactions, promoting a more compact protein structure and

increased protein hydration over an unfolded structure. Furthermore, it has been shown that a larger native protein surface area leads to a larger increase in the activation energy of unfolding, and thus greater stabilisation. Overall, adding sucrose depopulates aggregate-prone protein states through promotion of compact protein structures, leading to its stabilising ability during refolding. Whether this structural stabilisation by sucrose will significantly affect protein-protein interactions between unfolded states is, however, unknown.

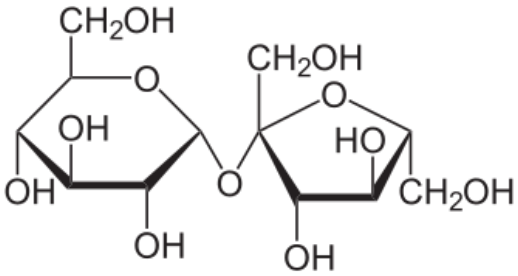


Figure 5.3 Structure of sucrose

Glycerol is a small molecule polyol, used as a protein stabilising agent in formulations, refolding, and crystallisation. Glycerol has been shown to reduce aggregation, promote protein compaction and increase the free energy of unfolding (Nicoud et al., 2015; Vagenende et al., 2009). Like sucrose, glycerol is preferentially excluded from the surface of the protein, leading to the promotion of increased protein hydration and increased structural stability to minimise protein-solvent interactions. Simulations performed by Vagenende also suggest that glycerol may bind to hydrophobic surface patches of proteins, stabilising aggregate-prone folding intermediates, in a similar fashion to hexylene glycol (Vagenende et al., 2009).

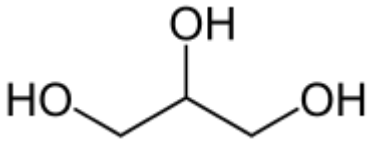


Figure 5.4. Structure of glycerol

Understanding how excipients alter refoldability requires measuring their effects on protein-protein interactions at moderate denaturant concentrations. A series of static light scattering experiments were performed, comparing the second virial coefficient of lysozyme in solutions with different excipients, including: 300 mM sucrose, 5% hexylene glycol or 5% glycerol alongside either 4 M urea or 4 M GdmHCl. The influence of Arg HCl and hexylene glycol on the protein-protein interactions and aggregation behaviour of lysozyme solutions in the presence of denaturants was determined through

static light scattering measurements, temperature-induced unfolding and dialysis refolding techniques. The second viral coefficient of lysozyme was measured at a set Arg HCl concentration of 0.5 M as a function of urea concentration between 0 and 4 M, and at a set urea concentration of 0.25 M as a function of Arg HCl concentration between 0 and 1 M. Refolding experiments were also performed to measure the effects of excipients on the levels of soluble and insoluble aggregation.

5.2 Results

5.2.1 Refolding excipients that contain no ions induce increased repulsion between lysozyme molecules

The aim of this section is to determine whether the presence of refolding excipients in protein solutions can influence protein-protein interactions. The second virial coefficient of lysozyme was measured in 50 mM sodium acetate buffer solution at pH 4.5, in solution with 300 mM sucrose, 5% hexylene glycol or 5% glycerol. The Zimm plot of each solution condition is shown in figure 5.5, with the B_{22} value presented in the figure description. In sucrose solutions, protein-protein repulsion is marginally increased, as reflected by increased second virial coefficient value compared to the excipient-free buffer measurement. 300 mM sucrose is the only excipient to increase protein-protein repulsion. The increased stabilisation by sucrose is linked to its preferential exclusion from the surface of the protein. The presence of sucrose has been shown to significantly increase the B_{22} and k_D values of lysozyme in previous studies (James and McManus, 2012; Valente et al., 2005). The exclusion causes a decrease in the flexibility of the protein and less exposed hydrophobic surface area, reducing the contribution to the B_{22} value from attractive protein-protein interactions.

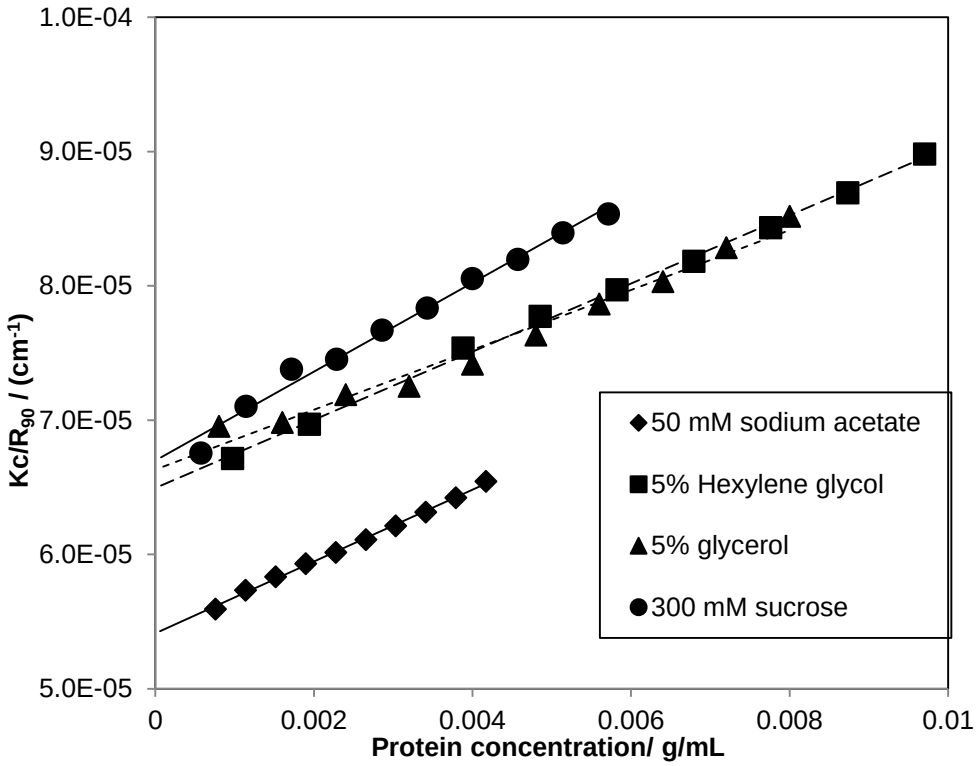


Figure 5.5. Zimm plots for lysozyme in 50 mM sodium acetate solution (B_{22} = 1.24 x 10^{-3} mL Mol/g^2), with either 5% hexylene glycol (B_{22} = 1.26 x 10^{-3} mL Mol/g^2), 5% glycerol ((B_{22} = 1.11 x 10^{-3} mL Mol/g^2), and 300 mM sucrose ((B_{22} = 1.65 x 10^{-3} mL Mol/g^2).

A small increase in repulsion is observed in hexylene glycol solutions compared to buffer conditions, while there is small decrease in repulsion observed in glycerol solutions. Hexylene glycol preferentially binds to the aliphatic surfaces of the protein, and displaces water molecules to reduce any sticky exposed hydrophobic surfaces, while glycerol is preferentially excluded from the protein surface. However, there is little or no change in the second virial coefficient of lysozyme in these excipients, suggesting that their effect does not significantly influence the protein-protein repulsion between natively folded proteins. The lack of an increase in the B_{22} value of lysozyme in glycerol solutions compared to buffer solutions is in contradiction to previously reported B_{22} values for lysozyme in glycerol solutions at pH 7.0 in 50 mM NaCl solutions, which observed an increase in protein-protein repulsion as a function of glycerol (Lui et al, 2004). The increase in protein-protein repulsion is attributed to an increase in the hydration layer of the protein and thus an increase in the excluded volume force contribution towards the B_{22} value (Gogelein et al., 2012). The values presented in figure 6.1 were measured at a lower pH and lower ionic strength than the values for Lui et al, and this may therefore explain the apparent contradictory effect of glycerol on the protein-protein interactions. A higher protein net charge at pH 4.5 compared to 7.0 will lead to a larger hydration shell, thus addition of glycerol will have a weaker effect on the hydration shell, leading to a reduced increase in the excluded volume effect.

5.2.2 The addition of excipients to urea or guandinium HCl solutions changes protein-protein interactions

To determine how non-ionic excipients influence protein-protein interactions while in the presence of denaturants, the second virial coefficient of lysozyme at pH 4.5 in solution with 300 mM sucrose, 5% hexylene glycol or 5% glycerol in solution with either 4 M urea or 4 M Gdm HCl was measured, with the results displayed in table 3.

Table 3: B_{22} values of lysozyme in denaturant and additive solutions

Excipient	B_{22} x 10^{-3}/ mL Mol/ g^2
50mM Acetate	1.24
4 M Urea	2.55
4 M Gdm HCl	0.42
300 mM Sucrose	1.65
300 mM Sucrose + 4 M urea	2.07
300 mM Sucrose + 4 M Gdm HCl	1.19
5 % Hexylene glycol	1.26
5% Hexylene glycol + 4 M urea	1.97
5% Hexylene glycol + 4 M Gdm HCl	0.49
5% Glycerol	1.11
5% Glycerol + 4 M urea	1.40
5% Glycerol + 4 M Gdm HCl	-8.40

In 4 M urea solutions, the addition of excipients reduces the strength of repulsive protein-protein interactions. The addition of 300 mM sucrose leads to a reduction in the B_{22} value of 4.8 x 10^{-4} mL Mol/ g^2, 5% hexylene glycol reduces the B_{22} value by 5.8 x 10^{-4} mL Mol/ g^2, and 5% glycerol reduces the B_{22} value by 1.4 x 10^{-3} mL Mol/ g^2. It is hypothesised that urea displaces water molecules from the surface of the protein, leading to an increase in the solubility the protein's hydrophobic residues. The preferential exclusion of sucrose or glycerol from the surface of the protein increases the chemical potential of the surface of the protein, and thus may lead to a disruption of the effect of urea on protein solubilisation.

The addition of excipients to solutions of 4 M Gdm HCl has a more complex effect on protein-protein interactions. The presence of 300 mM sucrose in 4 M Gdm HCl solutions leads to a marginal increase in protein-protein repulsion, with the value of B_{22} increasing by 7.7 x 10^{-4} mL Mol/ g^2. The addition of 5% hexylene glycol to 4 M Gdm HCl solutions does not significantly change protein-protein interactions, with a small increase of 7 x 10^{-5} mL Mol/ g^2. However, addition of 5% glycerol to 4 M

Gdm HCl solutions leads to a significant increase in protein-protein attraction, with the value of B_{22} decreasing by 8.8 x 10^{-3} mL Mol/ g^2.

Sucrose appears to have different effects on protein-protein interactions in denaturant solutions depending on the choice of denaturant. While repulsion is attenuated in the presence of sucrose in urea solutions, the opposite is true in Gdm HCl solutions to roughly the same magnitude. This suggests that, while an increase in hydration forces on the protein surface by the preferential exclusion of sucrose may reduce the efficiency of urea binding and solubilising effects, it has a positive effect on protein-protein repulsion in presence of Gdm HCl. This may be due to increased excluded volume forces through increased hydration layers.

Hexylene glycol attenuates repulsive interactions in urea solutions and does not significantly alter protein-protein interactions in Gdm HCl solutions. Hexylene glycol is preferentially excluded from the native protein surface due to the high net charge of the protein. However, in solutions of 4 M urea where populations of partially-folded intermediates are higher, exposed interior aliphatic residues provide binding sites for hexylene glycol and thus may stabilise folding intermediates that increase short-range protein-protein attractive interactions. However, protein-protein attraction does not increase in solutions of 4 M Gdm HCl and hexylene glycol compared to Gdm HCl-only solutions. This may indicate that hexylene glycol is not significantly bound or excluded from the surface of the protein in the presence of Gdm HCl, while in urea solutions the presence of hexylene glycol may disrupt surface solubilisation.

The presence of 5% glycerol to 4 M urea or 4 M Gdm HCl lysozyme solutions significantly reduces protein-protein repulsive interactions in both cases. While protein-protein repulsion is attenuated in 4 M urea solutions, the presence of glycerol in 4 M Gdm HCl solutions leads to a strongly negative B_{22} value to be observed, suggesting strongly attractive short-range interactions. This result is unexpected, and may suggest that the presence of glycerol, a known protein stabilising molecule, enhances short-range attractive interactions in partially-folded proteins through an as-yet unknown mechanism.

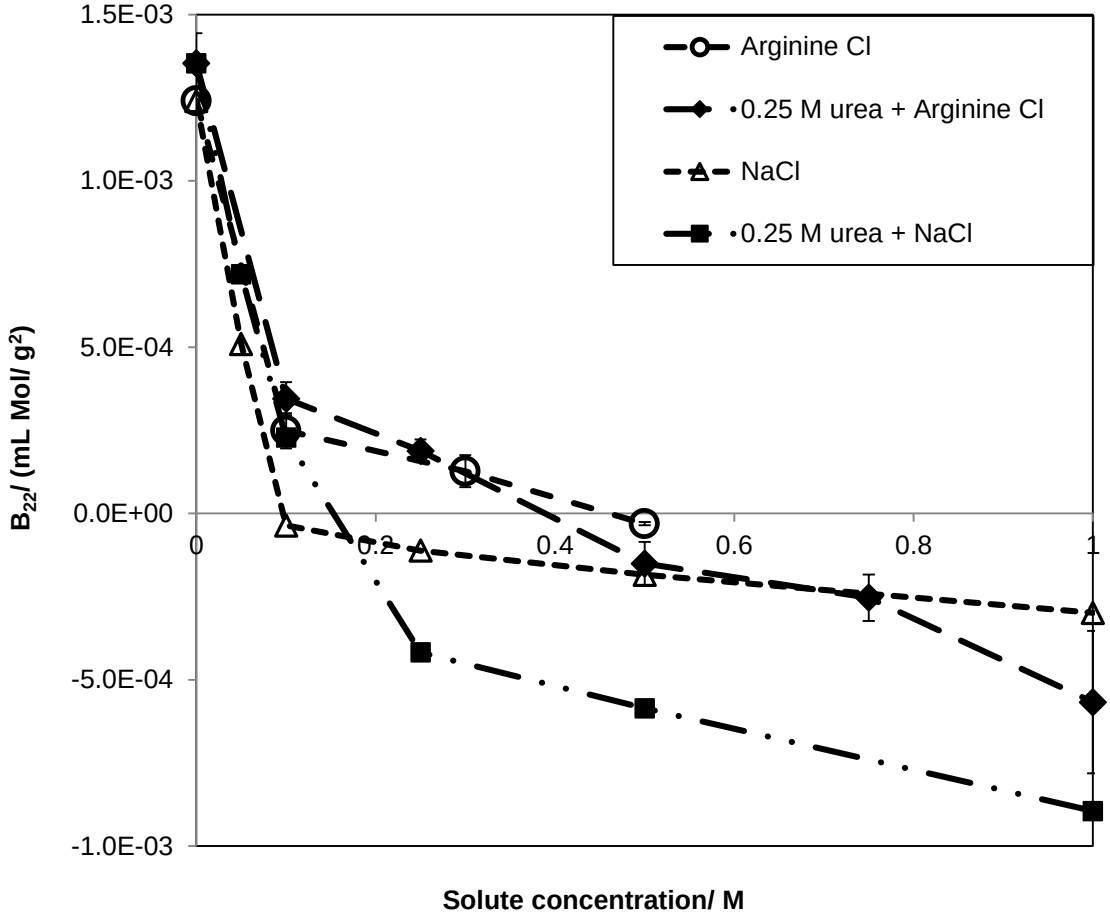


Figure 5.6. B_{22} values of lysozyme with 50 mM sodium acetate in solutions of between 0 and 0.5 M arginine chloride, between 0 and 1 M arginine chloride with 0.25 M urea, between 0 and 1 M NaCl, or between 0 and 1 M NaCl with 0.25 M urea

5.2.3 Arg HCl greatly attenuates short-range attraction between lysozyme molecules in the presence of urea compared to NaCl

The B_{22} value of lysozyme in solutions of 50 mM sodium acetate buffer at pH 4.5 at varying concentrations of Arg HCl between 0 and 0.5 M, or with 0.25 M urea and between 0 and 1 M Arg HCl was measured through SLS to investigate how the presence of the ionic refolding excipient Arg HCl influences protein-protein interactions of lysozyme in the presence of denaturants. Figure 5.6 depicts the B_{22} values for Arg HCl and urea solutions and the B_{22} values for lysozyme in solutions containing either 0 or 0.25 M urea and varying concentrations of NaCl between 0 and 1 M. The presence of Arg HCl in lysozyme solutions leads to a significant reduction in repulsion between proteins at concentrations of 100, 300 and 500 mM, with greatly reduced second virial coefficients of 2.4×10^{-4}, 1.2×10^{-4} and 1.8×10^{-5} mL Mol/ g^2, respectively, compared to excipient-free buffer solutions. This is due to the screening effect of the chloride ions within the solution, leading to significantly reduced electrostatic double-layer repulsion. However, the B_{22} value of lysozyme in Arg HCl is higher than the

B_{22} value of lysozyme at comparable NaCl concentrations. This suggests that, though electrostatic repulsion is completely screened, arginine molecules are binding to the protein surface and leading to stronger or reduced short-range attraction than is observed in comparable NaCl solutions between natively folded proteins. This result is a key finding that shows the aggregation-suppressing effects of arginine through the increase in short-range protein-protein repulsion compared to arginine-free solutions.

In solutions containing 0.25 M urea, increasing Arg HCl concentration also leads to screened electrostatic repulsive interactions, as reflected by decreasing B_{22} values. At 0.1 M Arg HCl concentration, the addition of 0.25 M urea leads to a small increase in protein-protein repulsion, which is also observed in solutions of 0.1 M NaCl. However, the increase in the B_{22} value in 0.1 M Arg HCl solutions of 9.65×10^{-5} mL Mol/g^2 is relatively small compared to the increase observed in low concentration sodium chloride solutions of 2.64×10^{-4} mL Mol/g^2, suggesting that the effect of urea on increasing protein-protein repulsion at low ionic strength is attenuated with the presence of arginine in solution. At 0.5 M Arg HCl solutions, the presence of 0.25 M urea induces a small increase in attractive protein-protein interactions, with the B_{22} value decreasing, becoming more negative compared to urea-free solutions.

An increase in attractive interactions is also observed in NaCl solutions in the presence of low concentrations of urea. The presence of 0.25 M urea in 0.25 or 0.5 M NaCl solutions leads to a significant decrease in the B_{22} value of 3.0×10^{-4} mL Mol/g^2 and 4.0×10^{-4} mL Mol/g^2 respectively. In comparison, a decrease in B_{22} value of 1.2×10^{-4} mL Mol/g^2 in the presence of urea in Arg HCl solutions is only observed at 0.5 M Arg HCl and is relatively small in comparison to NaCl solutions. At comparable ionic strengths, the addition of 0.25 M urea to NaCl solutions has a significantly greater effect on the protein-protein attractive interactions compared to Arg HCl solutions.

Comparing the values of B_{22} solutions of Arg HCl or NaCl in the presence of 0.25 M urea, it is clear that the presence of arginine significantly increases protein-protein repulsion compared to NaCl, an effect which is also observed in urea-free solutions. The effect of attenuating short-range attractive protein-protein interactions in Arg HCl solutions is more pronounced in the presence of urea, as the difference in B_{22} values at comparable additive concentrations is greater in the presence of 0.25 M urea than in urea-free solutions. This indicates that the solubilising effect of arginine is retained in solutions of urea compared to solutions at comparable NaCl concentration. It hypothesised that this is the case as Arg HCl solutions have little salting-out effect compared to the NaCl solutions, and thus there is little increase in short-range attractive interactions in Arg HCl solutions at increased ionic

strength and thus a reduction in short-range attractive interactions compared to NaCl solutions at comparable ionic strength.

At low Arg HCl concentrations, the increase in protein-protein repulsion in the presence of urea is attenuated by the presence of Arg HCl. This indicates that the mechanism through which low concentrations of urea alter short-range interactions between proteins is inhibited by the presence of arginine. Alternatively, the changes in short-range protein-protein attractive or repulsive forces induced by the presence of urea at high and low ionic strength, respectively, are being attenuated by the aggregation-suppressing effect of Arg HCl solutions. If arginine significantly reduces protein-protein contacts in solution, the effect of urea on the short-range interactions between proteins that leads to changes in the B_{22} value would be masked compared to comparable sodium chloride solutions. Alternatively, the solubilising effect of arginine on aliphatic and aromatic residues may be reducing the binding sites available for urea molecules and thus reducing the solubilising effectiveness of urea.

5.2.4 Arg HCl increases short range repulsion between lysozyme molecules over a wide urea concentration range compared to comparable NaCl concentrations

The second virial coefficient of lysozyme was measured at a set Arg HCl concentration of 0.5 M as a function of urea concentration ranging from 0 to 4 M at pH 4.5 in a 50 mM sodium acetate buffer solution to investigate how Arg HCl influences interactions between proteins over a wide range of denaturant concentrations. The results as a function of urea concentration are shown in the figure 5.7, alongside previously reported data of lysozyme in solutions containing 0.25 and 0.5 M NaCl and 50 mM acetate at pH 4.5. The second virial coefficient values of lysozyme in 0.5 M Arg HCl can be seen to follow a similar pattern between 0 and 4 M urea to that of 0.25 M NaCl solutions. The values are significantly higher than the B_{22} values for the corresponding solutions containing 0.5 M NaCl, which have the same ionic strength as 0.5 M Arg HCl solutions. For the Arg HCl solutions, there is an initial decrease in the B_{22} value, to a minimum at 0.5 M urea, before significantly increasing in value with further increasing urea concentration. B_{22} values return to positive values at 1 M urea. The B_{22} values for lysozyme in 0.5 M Arg HCl remain more positive than the corresponding values in 0.25 and 0.5 M NaCl solutions at all urea concentrations, suggesting that, when long range electrostatic interactions are screened, arginine induces weaker short range attractive interactions compared to sodium ions. It is clear that the presence of arginine in urea solutions significantly reduces attraction between proteins through attenuation of short-range attractive interactions induced by the screening effect of a high ionic strength. The increase in attractive interactions between 0 and 0.5 M urea in solutions containing 0.5 M Arg HCl suggests that, though the increased salting-out effect at high ionic strength induced by increasing urea concentration is weaker than in NaCl solutions, the presence of urea at

high concentration Arg HCl solutions does lead to an increase in short-range attractive interactions. The cause of this may be attributed to the same, as yet unknown, increased salting-out mechanism through which short-range attractive protein-protein interactions are greatly enhanced in the presence of urea and high ionic strength.

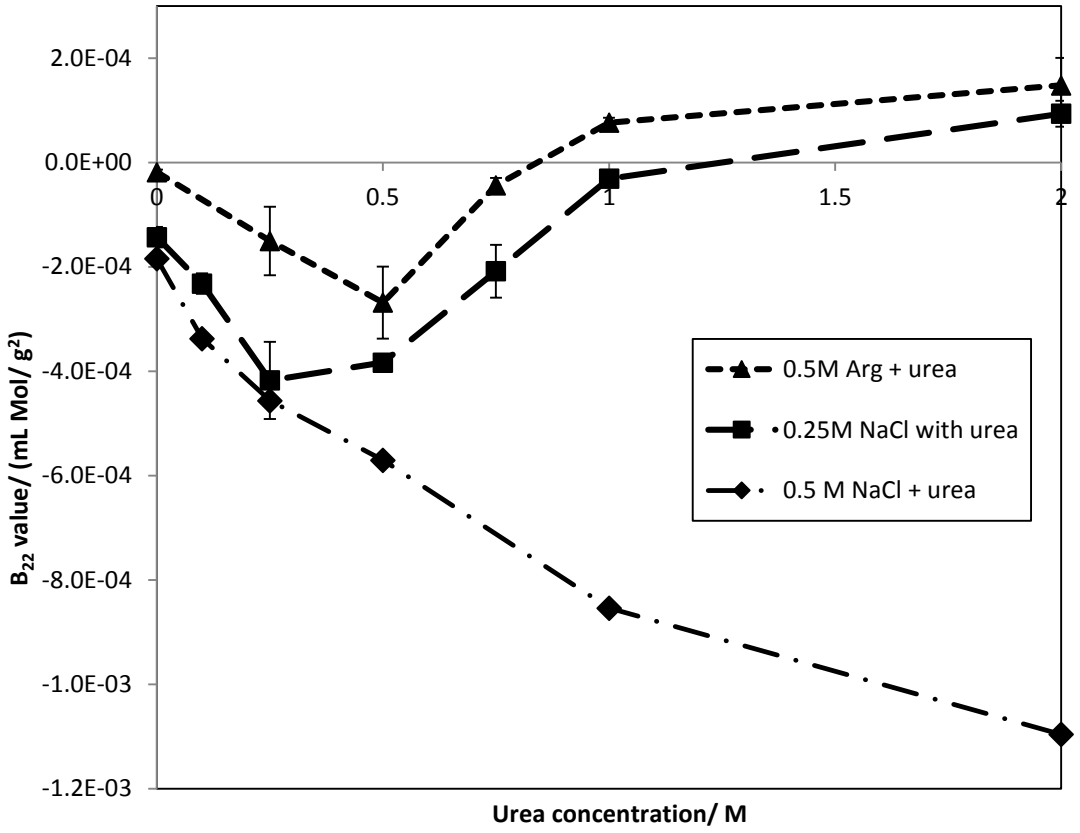


Figure 5.7. B_{22} values of lysozyme with 50 mM sodium acetate in solutions of between 0 and 4 M urea with 0.5 M arginine chloride, or between 0 and 4 M urea with 0.25 M NaCl or 0.5 M NaCl

5.2.5 The combination of 50 mM Arg HCl and 50 mM glutamic acid in the presence of urea leads to peaked repulsion between 0.25 and 0.5 M urea

Mixtures of Arg HCl and glutamic acid at low molarity have been shown to increase the solubility of proteins through co-operative binding to the surface of the protein, as positively charged arginine and negatively charged glutamic acid lead to an increased bound charge to the surface of the protein without changing the net charge of the protein (Golovanov et al., 2004). To investigate how the presence of Arg HCl and glutamic acid influence protein-protein interactions in urea solutions, the

second virial coefficient of lysozyme was measured in solutions of 50 mM Arg HCl and 50 mM glutamic acid (arg-glu) and between 0 and 4 M urea with 50 mM sodium acetate at p H 4.5. The results are shown in figure 5.8, alongside the B_{22} values of lysozyme in solutions of between 0 and 4 M urea. The B_{22} value for 50 mM Arg HCl and glutamic acid solutions is relatively low, with a value of 3.3×10^{-4} mL Mol/ g^2. This value is lower than the value for 50 mM sodium acetate buffer solutions, suggesting that the increased ionic strength of the 50 mM Arg HCl and 50 mM glutamic acid solution partially screens electrostatic repulsive interactions.

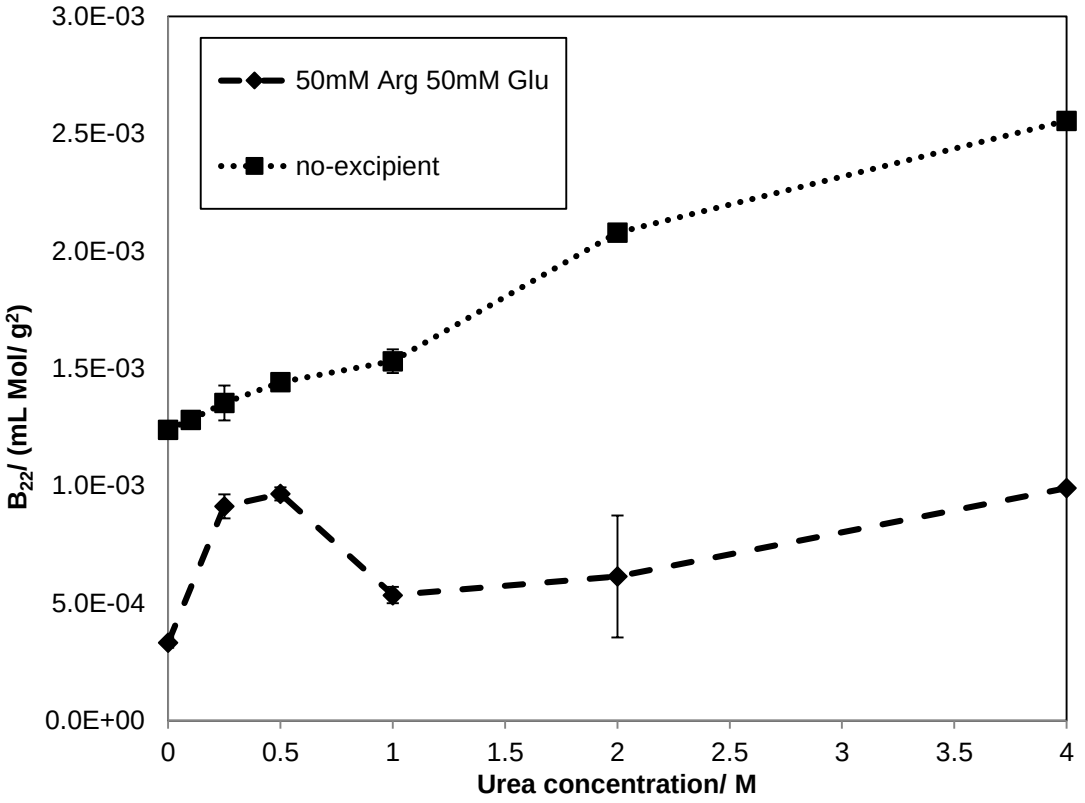


Figure 5.8. B_{22} values of lysozyme with 50 mM sodium acetate in solutions of between 0 and 4 M urea, or between 0 and 4 M urea with 50 mM arginine chloride and 50 mM glutamic acid

Increasing concentrations of urea leads to an increase in repulsive protein-protein interactions. However, unlike solutions of urea with no excipients which see a monotonic increase in repulsion with increasing urea concentration, repulsion in arg-glu solutions peaks at a concentration of 0.5 M urea. Repulsion decreases between 0.5 and 1 M urea, before increasing monotonically between 1 and 4 M urea. This pattern suggests that the presence of arg-glu in 0.5 M urea solutions leads to an additional increase in repulsion that does not occur in excipient-free solutions. It is hypothesised that repulsion is increased at 0.5 M urea solutions with arg-glu due to the binding of the charged amino acids to binding loci on the protein surface. This will lead to an increase in repulsive protein-protein interactions as the charged bound excipients lead to increased electrostatic repulsion. Repulsion decreases between 0.5 and 1 M urea due to the increased solubilisation of exposed hydrophobic

residues by urea molecules, and the weakening of the strength of attractive interactions between Arg HCl or glutamic acid and the surface of the protein. Repulsion then increases between 1 and 4 M urea in the same manner as excipient-free solutions of urea, as short-range attractive interactions are weakened by the denaturing mechanism of urea.

5.2.6 The presence of 5% hexylene glycol has a negative effect on protein-protein repulsion in the presence of urea

Hexylene glycol has been shown previously to attenuate protein-protein repulsive interactions in the presence of 4 M urea at pH 4.5 in 50 mM acetate buffer solutions. To investigate how hexylene glycol influences protein-protein interactions in solutions of denaturant, the second virial coefficient of lysozyme was measured in solution with 5% hexylene glycol as a function of urea concentration ranging from 0 to 4 M with 50 mM sodium acetate buffer at pH 4.5. The data is shown in figure 6.5, alongside B_{22} values of lysozyme in excipient-free solutions. The B_{22} values for the hexylene glycol solutions are slightly more positive than excipient-free measurement at 0 M urea, as described in a previous section of this chapter. With rising urea concentration, the B_{22} value of lysozyme in 5% hexylene glycol begins to linearly increase, to a maximum of 1.97×10^{-4} mL Mol/ g^2 at 4 M urea, following a similar pattern to that of the excipient-free solutions. Up to 1 M urea, protein-protein repulsion in the 5% hexylene glycol solutions is slightly greater than that of the excipient free solution. This increase in repulsion may be due to the preferential exclusion of hexylene glycol from the protein surface due to the high net charge of the protein, leading to the stabilisation of the protein structure due to preferential hydration of the protein surface. This will lead to small increase in repulsion due to an increase in repulsive hydration forces between the proteins.

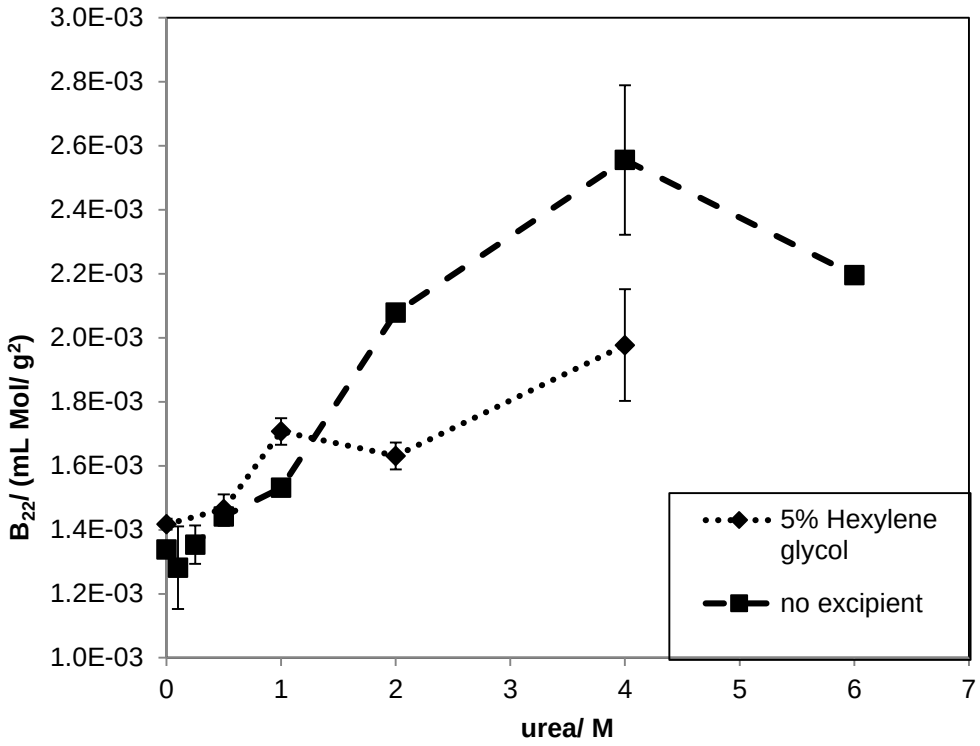


Figure 5.9. B_{22} values of lysozyme with 50 mM sodium acetate in solutions of between 0 and 4 M urea, or between 0 and 4 M urea with 5% hexylene glycol, at pH 4.5

However, the B_{22} values for 2 M and 4 M urea in 5% hexylene glycol are significantly less positive than those of their equivalent excipient-free urea concentrations, and the gradient of the increased repulsion with increasing urea concentration is less than that of excipient-free solutions. This suggests that the presence of hexylene glycol reduces protein-protein repulsion while in solution with intermediate concentrations of urea. The solubilisation of the protein surface by urea will lead to altered hydration forces between the proteins and thus may alter the binding of hexylene glycol to the surface of the protein, leading to increased short-range attractive hydrophobic protein-protein interactions (Anand et al., 2002; Kita et al., 1994).This in turn will lead to a less positive B_{22} value.

5.2.7 Dialysis refolding of lysozyme into refolding excipients with a range of ionic strengths

The second virial coefficient measurements described above reflect the effects of common refolding excipients on protein-protein interactions in solutions at low and high ionic strength. Dialysis refolding experiments were carried out to determine whether the effects of excipients on protein-protein interactions control their ability to effect aggregation propensity of lysozyme in varying ionic strength. Lysozyme was denatured with 8 M urea and 10 mM DTT at 10 mg/mL followed by dialysis into a refolding buffer containing 50 mM acetate, 10 mM cysteine and a refolding excipient mixture. The

refolding conditions investigated were solutions containing between 0 and 0.5 M Arg HCl, solutions containing 5% hexylene glycol with 0 to 0.5 M NaCl, and solutions containing only NaCl between 0 to 0.5 M. NaCl was used to regulate the ionic strength of the non-ionic refolding excipients. Increasing ionic strength is needed to increase the aggregation-propensity of the protein as high refolding yields are always obtained at low ionic strength due to the strong long range electrostatic repulsion that exists between lysozyme molecules. This allows a more clear differentiation for the excipient effects on protein aggregation-behaviour. Refolding yield was calculated through measurement of the soluble protein concentration before and after filtering with 0.22 µm filter through measurement of absorbance at 280 nm. The pre-filtered protein concentration of the solution was calculated by re-solubilising a sample of the refolded protein solution with the 8 M urea denaturing buffer, followed by measurement of the protein concentration of the denatured protein solution. Soluble refolding yield was calculated by dividing the concentration of protein remaining soluble after filtering by the pre-filtration protein concentration. The results are presented in the figure 6.10.

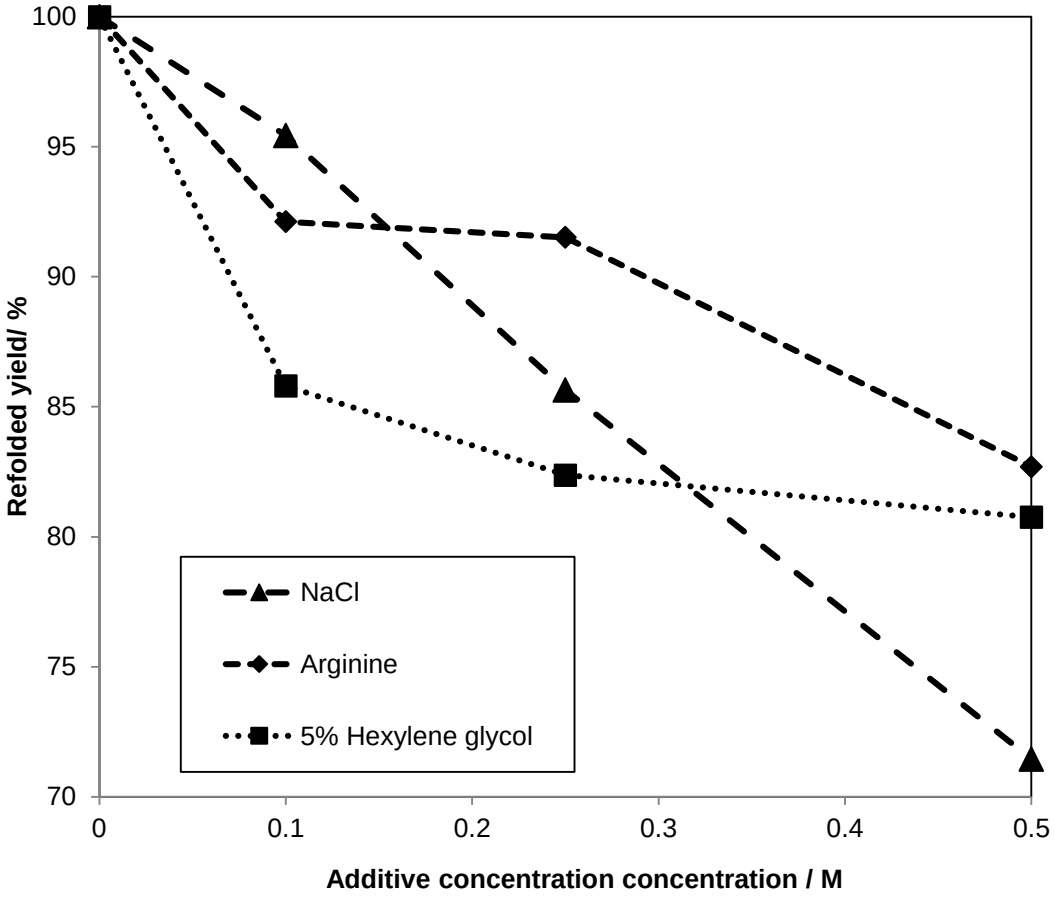


Figure 6.10. The soluble refolding yield of lysozyme solutions denatured in 8 M urea and 10 mM DTT with 50 mM sodium acetate solutions at pH 4.5, and dialysed at a ratio of 1:30 into a range of refolding additives, each containing 10 mM cysteine and 50 mM sodium acetate at pH 4.5. The additives used were: between 0 and 0.5 M NaCl, 5% hexylene glycol with between 0 and 0.5 M NaCl, and between 0 and 0.5 M arginine chloride.

The refolding yield where the refolding buffer contains no excipients and only buffer was 100%. This result is to be expected, as lysozyme is a stable protein that refolds well in low ionic strength solutions and at low protein concentration. Increasing NaCl concentration to 0.1 M leads to an 8% reduction in soluble yield in NaCl only solutions. Increasing NaCl concentration above 0.1 M NaCl increases the level of insoluble aggregate formation monotonically, leading to a 30% loss of protein in 0.5 M NaCl solutions. The increase in aggregation observed with increasing NaCl concentration is due to the screening of electrostatic repulsion between the proteins, leading to more dominant short-range attractive interactions. This leads to an increase in protein-protein attractive interactions that promote aggregate formation as the solubility of the protein decreases with decreasing urea concentration during refolding. As the sodium and chloride ions are excluded from the protein surface, a higher NaCl concentration also leads to a stronger salting out force that adds to the electrostatic screening effect, to significantly increase the level of aggregation during refolding at higher salt concentration. This aggregation behaviour follows a similar pattern as the B_{22} values calculated for lysozyme in NaCl and urea solutions, which show that in urea solutions increasing ionic strength leads to significantly increased short-range protein-protein attractive interactions, which have been shown here to lead to an increase propensity for aggregation.

The aggregation-behaviour of lysozyme in Arg HCl solutions exhibits a similar pattern to that of NaCl, with a reduction in soluble yield at 0.1 M Arg HCl and an increase in aggregation propensity with increasing ionic strength. However, the negative effect of Arg HCl on refolding yield is less than that for NaCl solutions at higher additive concentrations. For instance, at a concentration of 0.5 M Arg HCl, aggregate formation is significantly less than that observed in 0.5 M NaCl solutions. The reduced aggregation in 0.25 M and 0. 5 M Arg HCl solutions compared to the same ionic strength NaCl solutions is due to the aggregation-suppressing effect of the arginine molecules. Aggregation propensity is increased with increasing Arg HCl concentration due to the increased ionic strength of the solution, leading to screened electrostatic repulsive interactions between refolding proteins. This can be seen in the second virial coefficient measurements of lysozyme in solution with Arg HCl that show a reduction in repulsion between proteins with increasing Arg HCl concentration, but remaining more positive in value than comparable NaCl concentrations. This clearly corresponds with the aggregation behaviour that is observed in figure 6.6.

stabilising aggregation-prone folding intermediates. At high ionic strength (0.5 M NaCl), hexylene glycol increases the refolding yield compared to NaCl-only solutions. It is hypothesised that this may be due to the solubilising effect of hexylene glycol reducing the “salting-out” effect of the increased ionic strength, and suppressing aggregation. It is clear, however, that hexylene glycol has a weaker aggregation-suppressing effect compared to Arg HCl at comparable ionic strengths.

5.3 Conclusions

In the data presented above, measurements of the second virial coefficient of lysozyme in solution with a range of non-ionic refolding excipients solutions show that sucrose and hexylene glycol slightly increase repulsion between natively folded proteins, while glycerol slightly decreases repulsion. However, the addition of these excipients to 4 M urea solutions leads to a universal decrease in protein-protein repulsion compared to excipient-free urea solutions. This indicates that the non-ionic excipients may have a disruptive effect on the protein-protein repulsion that is induced in solutions of urea between proteins. The mechanism through which sucrose, hexylene glycol and glycerol affect protein stability is primarily through preferentially exclusion from the protein surface, while urea is preferentially bound to the surface of the protein. The opposed nature of how the excipients influence protein-protein interactions compared to urea may lead to a disruption of the mechanisms through which urea increases repulsion, primarily through solubilisation of hydrophobic surface residues. Dialysis refolding of denatured lysozyme solutions from high urea concentrations into 5% hexylene glycol at a range of ionic strengths shows that at low to moderate ionic strength, the presence of hexylene glycol leads to an increase in attractive protein-protein interactions, leading to increased aggregation compared to hexylene-glycol free solutions at comparable ionic strength. This correlates with measured B_{22} values of lysozyme in solution with hexylene glycol over a range of urea concentrations, showing that the value of B_{22} for solutions of hexylene glycol and between 2 and 4 M urea is lower than comparable excipient-free urea solutions. Hexylene glycol has also been shown to stabilise denatured proteins. This evidence suggests that excipients which are preferentially excluded from the protein surface may disrupt the solubilising effect of urea that decreases short-range attractive interactions, leading to more attractive protein-protein interactions.

A comprehensive series of second virial coefficient measurements show Arg HCl screens repulsive electrostatic protein-protein interactions due to the increased ionic strength, but attenuates short-range attractive interactions at all urea concentrations in comparison to 0.25 M NaCl solutions, which has similarly screened electrostatic repulsion. Dialysis refolding experiments show the aggregation-inhibiting effect of arginine molecules with the significant reduction in aggregation compared to NaCl

concentrations at the same ionic strength, suggesting a broad correlation between second virial coefficients and aggregation propensity. It is clear that arginine attenuates short-range attraction interactions between proteins compared to solutions of NaCl at the same ionic strength. The increase in short-range attraction in Arg HCl solutions compared to NaCl solutions is less in solutions containing 0.25 M urea, indicating that the solubilising effect of arginine is greater than NaCl solutions while in solution with urea, where short-range attractive interactions are strengthened at high ionic strength.

Second virial coefficient measurements of solutions of lysozyme with 50 mM Arg HCl and 50 mM glutamic acid and between 0 and 4 M urea revealed that protein-protein repulsion was enhanced in solutions of low concentrations of urea (>1 M) in comparison to excipient-free solutions. This effect of enhanced repulsion is potentially due to the binding of the amino acids to the hydrophobic surface residues of the protein, which have increased solubility at low urea concentrations. The binding of the amino acids is attenuated at higher urea concentrations due to the weakening of attractive amino acid-protein interactions by the denaturing mechanism of urea. This could potentially be a highly effective refolding solution additive combination, as it is theorised that aggregation during refolding occurs when partially folded intermediates are highly populated, and increasing protein-protein repulsion during that period may significantly inhibit the formation of aggregates.

To conclude, measurements taken here show that there is a clear interplay between denaturants, salts and refolding excipients that has a strong impact on protein interactions and, in turn, protein aggregation. The pairing of refolding excipient and denaturant is also important, as one can attenuate or promote the effects of the other.

Chapter 6: Conclusions

The purpose of this project was to ascertain and characterise the factors that influence protein aggregation and folding during inclusion body refolding, and how the composition of solubilising and refolding solutions impacts upon protein-protein interactions and aggregate formation. Previous literature has suggested that the second virial coefficient of a protein is sensitive to the pH, ionic strength, and excipient composition of the solution and can be used to predict the propensity at which a protein will aggregate in refolding experiments, with a more positive B_{22} value in denaturing conditions predicting a higher refolding yield. It was hypothesised that more repulsive interactions between partially folded and denatured states will lead to reduced levels of aggregation during refolding. Static and dynamic light scattering experiments have measured the interaction parameters of lysozyme, ribonuclease A and preproinsulin in a range of denaturants, ionic strengths and excipient concentrations to determine how each of these factors influences protein-protein interactions. Refolding experiments have been used to determine whether the aggregation propensity of the protein in varying denaturation and refolding conditions is correlated with the measured B_{22} values in the same solution conditions.

Our findings have shown that, with all proteins investigated, the presence of urea in protein solutions leads to increased protein-protein repulsion compared to denaturant-free solutions, reaching a peak at concentrations of urea in which there is a significant population of unfolded proteins. Calculation of the contribution of different interaction forces towards the B_{22} value revealed that both urea and Gdm HCl solutions significantly attenuate the short-range attractive interactions between proteins, becoming net repulsive at denaturant concentrations that promote unfolding. The evidence suggests that the denaturing mechanism of both urea and Gdm HCl has a significant impact on the protein-protein interactions exhibited by proteins, even at the non-denaturing concentrations (0.1 – 1 M) at which refolding occurs, predominantly through the solubilisation of hydrophobic residues on the surface of the protein. This also reveals that urea and Gdm HCl share a common effect on short-range protein-protein interactions, potentially suggesting a comparable refolding mechanism. However, comparing the B_{22sr} of lysozyme in low ionic strength urea solutions to that of Gdm HCl shows that Gdm HCl has a stronger effect on reducing short-range attraction that urea at low denaturant concentrations, indicating that there are important differences in the effects of the denaturants at non-denaturing concentrations. Due to the nature of the SLS measurements, the origin of the differences in B_{22sr} cannot be directly characterised, as only the averaged interactions are collected and not specific molecular contributions towards the B_{22} value. To further characterise how Gdm and urea molecules influence protein-protein interactions at non-denaturing concentrations, further experiment studies measuring the strength of hydrophobic interactions, alongside molecular simulations focusing on the binding of denaturant molecules to the surface of proteins are required.

Refolding experiments revealed that, though both urea and Gdm HCl significantly attenuate short-range protein-protein attraction, the high ionic strength of Gdm HCl solutions leads to a significantly increased propensity for aggregation in both lysozyme and preproinsulin solutions. Increased levels of protein aggregation in Gdm HCl solutions correlates with the measured B_{22} values of lysozyme, that

shows significantly more repulsive protein-protein interactions in urea solutions compared to Gdm HCl solutions at all concentrations. This was our first evidence of a correlation between B_{22} values and aggregation propensity during refolding, and can be clearly explained by the screening effects of the high ionic strength in Gdm HCl solutions reducing long-range electrostatic protein-protein repulsion, leading to the dominance of short-range attractive interactions and the formation of aggregates. It is clear that, though Gdm HCl significantly attenuates short-range attraction, at low denaturant concentration this does not stop aggregation from occurring and strong repulsive electrostatic interactions are key to protein solubility during refolding.

To further understand the influence of ionic strength on the protein-protein interactions exhibited by proteins at moderate denaturant concentrations, the B_{22} values of lysozyme solutions were measured over a range of urea concentrations and ionic strengths, with the ionic strength regulated by sodium chloride. What was discovered was that interactions between proteins in solution with urea are highly sensitive to the ionic strength of the solution in a way that is not replicated in denaturant-free solutions. At moderate urea concentrations between 1 and 4 M, the addition of a relatively small range of NaCl concentrations leads to a dramatic reduction in protein-protein repulsion and a significant increase in net attractive interactions. At low urea concentrations (0.1 – 0.75 M), at which refolding of proteins occurs, a reduction in repulsion and an increase in attractive interactions is observed with increasing urea concentration at higher NaCl concentrations, suggesting that the presence of urea at higher ionic strength may increase the strength of short-range protein-protein interactions. Repulsion increases with increasing urea concentration in 0.25 M NaCl solutions, but not in 0.5 M NaCl solutions, suggesting an increased sensitivity to ionic strength above 1 M urea. Refolding experiments where lysozyme, denatured in high concentration urea solutions, was dialysed into varying concentrations of NaCl found that aggregation significantly increased between 0.25 M and 0.5 M NaCl refolding solutions. Since electrostatic repulsion is screened at both 0.25 M and 0.5 M NaCl, and thus does not contribute towards the protein-protein interactions in both solutions, the increase in aggregation in 0.5 M NaCl solutions must be the result of changes in the short-range attractive interactions between the proteins. This indicates that the presence of NaCl has a significant impact on the short-range, as well as the long-range, protein-protein interactions between unfolded and partially folded proteins in urea solutions. Further, it should be noted that the aggregation observed during refolding of lysozyme denatured in urea solutions is significantly higher in solutions where a negative B_{22} value is recorded at any point between 0 and 4 M urea. This is true for 0.25 M and 0.5 M NaCl solutions, which both have negative B_{22} values between 0 and 4 M urea, and both exhibit markedly higher levels of aggregation than solutions containing 0.1 M or zero NaCl. Interestingly, the sensitivity of short-range interactions to ionic strength is not replicated in Gdm HCl solutions, which, though reaching ionic strengths equal to the concentration of the denaturant, do not exhibit as dramatic an increase in attractive interactions proportional to ionic strength. This may due the presence of the Gdm ions, which have a solubilising effect due to their preferential interaction on the protein surface, compared to the preferential exclusion of sodium ions. The mechanism through which short-range protein-protein attraction significantly increases over such a small range of ionic strengths at moderate urea concentrations may be due to an enhanced salting out effect, as the

increased exposure of hydrophobic residues in the molten globule and partially folded structures of the protein increase the net hydrophobic surface area of the proteins. However, this mechanism cannot be confirmed from the T_m, T_{agg} and B_{22} measurements presented, and may require further investigation into the influence of salts on the structural stability and phase behaviour of partially-folded proteins at room temperature.

A series of SLS and refolding measurements of lysozyme solutions with denaturants and ionic and non-ionic refolding excipients were performed to determine how excipients influence protein-protein interactions in solutions of either urea or Gdm HCl. Understanding how excipients influence protein-protein interactions in the presence of denaturants is essential if a clear understanding of how aggregation is suppressed or promoted by the competing effects of protein solubilisation by denaturants and protein stabilisation by folding promoters such as sucrose and hexylene glycol can be determined. Sucrose, hexylene glycol and glycerol were all found to attenuate protein-protein repulsion in urea solutions, suggesting that the preferential exclusion of the excipients from the surface of the protein may have a disruptive effect on the surface solubilisation of the protein and thus lead to a reduction in the repulsion-inducing effects of urea solutions. However, interactions remain repulsive in all non-ionic excipient and urea solutions. In solutions of Gdm HCl solutions, the addition of the same excipients leads to an increase in repulsion with 300 mM sucrose, no change in interactions with 5% hexylene glycol and an increase in attraction in 5% glycerol solutions. This suggests that the presence of excipients in Gdm HCl solutions has a far more unpredictable effect on protein-protein interactions that cannot be simply explained by the preferential exclusion of the excipients from the surface of the protein, and may be related to how the excipients may interact with the aromatic and aliphatic residues on the protein surface instead, and how Gdm HCl may influence this binding in comparison to urea. The measurement of the B_{22} value of lysozyme in solutions of 5% hexylene glycol with between 0 and 4 M urea revealed that whether the presence of hexylene glycol leads to increased protein-protein repulsion or attraction induced by the excipient are dependent on the concentration of urea in which the measurements are taken. This may well be the case for glycerol, sucrose and many other non-ionic refolding excipients, and a more comprehensive investigation of the influence of each excipient on the protein-protein interactions in a range of urea or Gdm HCl concentrations is required before clear and precise theories can be hypothesised as to how each influences protein-protein interactions and aggregation in the presence of each denaturant.

The influence of Arg HCl on protein-protein interactions in urea solutions was also investigated. It was found that arginine significantly increases short-range protein-protein repulsion compared to - NaCl solutions of equal ionic strength, and this effect is enhanced in the presence of urea. This result is expected from the well characterised effects of arginine on protein solubility, and the fact that the solubilising effect is enhanced in urea solutions, where partially-folded intermediate populations are increased compared to denaturant-free solutions, contributes to the theory that arginine preferentially binds to aromatic and aliphatic residues on the protein surface which screens short-range attractive interactions.

To conclude, the prediction of whether an inclusion body protein will refold or aggregate based upon the second virial coefficient of the protein in high concentration denaturant solutions is far more complicated than has been previously theorised. Interactions between proteins change from attractive to repulsive depending on the concentration of the denaturant and the denaturant being used. Ionic strength has a far greater influence on the aggregation propensity of a partially-folded protein than previously characterised, primarily due to how the ionic environment can influence the repulsive electrostatic interactions between proteins, and also the short-range attractive interactions which can dominant aggregation if electrostatic repulsion is weak. Further, more detailed investigation of how excipients can be used to increase, rather than decrease, protein-protein repulsion in the presence of denaturants is a clear avenue of research, as it is clear from the data presented here that the influence of excipients and denaturants on the protein surface chemistry and thus the protein-protein interactions do not occur in isolation, but strongly influence each other.

Chapter 7: References

Ami, D., A. Natalello, P. Gatti-Lafranconi, M. Lotti, and S. M. Doglia, 2005, Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy: Febs Letters, v. 579, p. 3433-3436.

Anand, K., D. Pal, and R. Hilgenfeld, 2002, An overview on 2-methyl-2,4-pentanediol in crystallization and in crystals of biological macromolecules: Acta Crystallographica Section D-Biological Crystallography, v. 58, p. 1722-1728.

Andrews, J. M., and C. J. Roberts, 2007, Non-native aggregation of alpha-chymotrypsinogen occurs through nucleation and growth with competing nucleus sizes and negative activation energies: Biochemistry, v. 46, p. 7558-7571.

Anfinsen, C. B., 1973, PRINCIPLES THAT GOVERN FOLDING OF PROTEIN CHAINS: Science, v. 181, p. 223-230.

Anfinsen, C. B., E. Haber, M. Sela, and F. H. White, Jr., 1961, The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain: Proceedings of the National Academy of Sciences of the United States of America, v. 47, p. 1309-14.

Arakawa, T., D. Ejima, K. Tsumoto, N. Obeyama, Y. Tanaka, Y. Kita, and S. N. Timasheff, 2007a, Suppression of protein interactions by Arg HCl: A proposed mechanism of the arginine effects: Biophysical Chemistry, v. 127, p. 1-8.

Arakawa, T., Y. Kita, and S. N. Timasheff, 2007b, Protein precipitation and denaturation by dimethyl sulfoxide: Biophysical Chemistry, v. 131, p. 62-70.

Arakawa, T., and S. N. Timasheff, 1982, STABILIZATION OF PROTEIN-STRUCTURE BY SUGARS: Biochemistry, v. 21, p. 6536-6544.

Arakawa, T., and S. N. Timasheff, 1985, THE STABILIZATION OF PROTEINS BY OSMOLYTES: Biophysical Journal, v. 47, p. 411-414.

Arakawa, T., and K. Tsumoto, 2003, The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation: Biochemical and Biophysical Research Communications, v. 304, p. 148-152.

Arakawa, T., K. Tsumoto, Y. Kita, B. Chang, and D. Ejima, 2007c, Biotechnology applications of amino acids in protein purification and formulations: Amino Acids, v. 33, p. 587-605.

Bardwell, J. C. A., K. McGovern, and J. Beckwith, 1991, IDENTIFICATION OF A PROTEIN REQUIRED FOR DISULFIDE BOND FORMATION INVIVO: Cell, v. 67, p. 581-589.

Basu, A., X. Li, and S. S. J. Leong, 2011, Refolding of proteins from inclusion bodies: rational design and recipes: Applied Microbiology and Biotechnology, v. 92.

Bennion, B. J., and V. Daggett, 2003, The molecular basis for the chemical denaturation of proteins by urea: Proceedings of the National Academy of Sciences of the United States of America, v. 100.

Bowden, G. A., A. M. Paredes, and G. Georgiou, 1991, STRUCTURE AND MORPHOLOGY OF PROTEIN INCLUSION-BODIES IN ESCHERICHIA-COLI: Bio-Technology, v. 9, p. 725-730.

Buchner, J., M. Schmidt, M. Fuchs, R. Jaenicke, R. Rudolph, F. X. Schmid, and T. Kiefhaber, 1991, GROE FACILITATES REFOLDING OF CITRATE SYNTHASE BY SUPPRESSING AGGREGATION: Biochemistry, v. 30, p. 1586-1591.

Buswell, A. M., and A. P. J. Middelberg, 2003, A new kinetic scheme for lysozyme refolding and aggregation: Biotechnology and Bioengineering, v. 83.

Bye, J. W., and R. J. Falconer, 2013, Thermal stability of lysozyme as a function of ion concentration: A reappraisal of the relationship between the Hofmeister series and protein stability: Protein Science, v. 22, p. 1563-1570.

Chatterjee, A., A. Kumar, J. Chugh, S. Srivastava, N. S. Bhavesh, and R. V. Hosur, 2005, NMR of unfolded proteins: Journal of Chemical Sciences, v. 117, p. 3-21.

Chen, B., H. J. Zhang, W. Y. Xi, L. Q. Zhao, L. Liang, and Y. T. Chen, 2014, Unfolding mechanism of lysozyme in various urea solutions: Insights from fluorescence spectroscopy: Journal of Molecular Structure, v. 1076, p. 524-528.

Chi, E. Y., S. Krishnan, B. S. Kendrick, B. S. Chang, J. F. Carpenter, and T. W. Randolph, 2003, Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor: Protein Science, v. 12, p. 903-913.

Chiti, F., and C. M. Dobson, 2006, Protein misfolding, functional amyloid, and human disease, Annual Review of Biochemistry: Annual Review of Biochemistry, v. 75: Palo Alto, Annual Reviews, p. 333-366.

Clark, E. D., 2001, Protein refolding for industrial processes: Current Opinion in Biotechnology, v. 12, p. 202-207.

Clark, E. D., E. Schwarz, and R. Rudolph, 1999, Inhibition of aggregation side reactions during in vitro protein folding: Amyloid, Prions, and Other Protein Aggregates, v. 309, p. 217-236.

Conchillo-Sole, O., N. S. de Groot, F. X. Aviles, J. Vendrell, X. Daura, and S. Ventura, 2007, AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides: Bmc Bioinformatics, v. 8.

Curtis, R. A., and L. Lue, 2006, A molecular approach to bioseparations: Protein-protein and protein-salt interactions: Chemical Engineering Science, v. 61.

Curtis, R. A., J. Newman, H. W. Blanch, and J. M. Prausnitz, 2001, McMillan-Mayer solution thermodynamics for a protein in a mixed solvent: Fluid Phase Equilibria, v. 192.

Curtis, R. A., J. Ulrich, A. Montaser, J. M. Prausnitz, and H. W. Blanch, 2002, Protein-protein interactions in concentrated electrolyte solutions - Hofmeister-series effects: Biotechnology and Bioengineering, v. 79, p. 367-380.

Dempsey, C. E., T. J. Piggot, and P. E. Mason, 2005, Dissecting contributions to the denaturant sensitivities of proteins: Biochemistry, v. 44, p. 775-781.

Deyoung, L. R., K. A. Dill, and A. L. Fink, 1993, AGGREGATION AND DENATURATION OF APOMYOGLOBIN IN AQUEOUS UREA SOLUTIONS: Biochemistry, v. 32, p. 3877-3886.

Dong, X. Y., Y. Huang, and Y. Sun, 2004, Refolding kinetics of denatured-reduced lysozyme in the presence of folding aids: Journal of Biotechnology, v. 114.

Dumetz, A. C., A. M. Snellinger-O'Brien, E. W. Kaler, and A. M. Lenhoff, 2007, Patterns of protein - protein interactions in salt solutions and implications for protein crystallization: Protein Science, v. 16, p. 1867-1877.

Elcock, A. H., and J. A. McCammon, 2001, Calculation of weak protein-protein interactions: The pH dependence of the second virial coefficient: Biophysical Journal, v. 80.

Fang, Y., and L. A. Yi, 2003, Unfolding of lysozyme induced by urea and guanidine hydrochloride studied by "phase diagram" method of fluorescence: Acta Chimica Sinica, v. 61, p. 803-807.

Finke, J. M., M. Roy, B. H. Zimm, and P. A. Jennings, 2000, Aggregation events occur prior to stable intermediate formation during refolding of interleukin 1 beta: Biochemistry, v. 39, p. 575-583.

Flocco, M. M., and S. L. Mowbray, 1994, PLANAR STACKING INTERACTIONS OF ARGININE AND AROMATIC SIDE-CHAINS IN PROTEINS: Journal of Molecular Biology, v. 235, p. 709-717.

Freydell, E. J., L. van der Wielen, M. Eppink, and M. Ottens, 2010, Ion-exchange chromatographic protein refolding: Journal of Chromatography A, v. 1217, p. 7265-7274.

George, A., and W. W. Wilson, 1994, PREDICTING PROTEIN CRYSTALLIZATION FROM A DILUTE-SOLUTION PROPERTY: Acta Crystallographica Section D-Biological Crystallography, v. 50.

Gillis, S., M. M. Ferm, W. Ou, and K. A. Smith, 1978, T-CELL GROWTH-FACTOR - PARAMETERS OF PRODUCTION AND A QUANTITATIVE MICROASSAY FOR ACTIVITY: Journal of Immunology, v. 120, p. 2027-2032.

Gogelein, C., D. Wagner, F. Cardinaux, G. Nagele, and S. U. Egelhaaf, 2012, Effect of glycerol and dimethyl sulfoxide on the phase behavior of lysozyme: Theory and experiments: Journal of Chemical Physics, v. 136, p. 12.

Golovanov, A. P., G. M. Hautbergue, S. A. Wilson, and L. Y. Lian, 2004, A simple method for improving protein solubility and long-term stability: Journal of the American Chemical Society, v. 126, p. 8933-8939.
Greene, R. F., and C. N. Pace, 1974, UREA AND GUANIDINE-HYDROCHLORIDE DENATURATION OF RIBONUCLEASE, LYSOZYME, ALPHA-CHYMOTRYPSIN, AND BETA-LACTOGLOBULIN: Journal of Biological Chemistry, v. 249, p. 5388-5393.
Greenfie.N, and G. D. Fasman, 1969, COMPUTED CIRCULAR DICHROISM SPECTRA FOR EVALUATION OF PROTEIN CONFORMATION: Biochemistry, v. 8, p. 4108-&.
Grigsby, J. J., H. W. Blanch, and J. M. Prausnitz, 2001, Cloud-point temperatures for lysozyme in electrolyte solutions: effect of salt type, salt concentration and pH: Biophysical Chemistry, v. 91, p. 231-243.
Hayakawa, S., and S. Nakai, 1985a, CONTRIBUTION OF HYDROPHOBICITY, NET CHARGE AND SULFHYDRYL-GROUPS TO THERMAL-PROPERTIES OF OVALBUMIN: Canadian Institute of Food Science and Technology Journal-Journal De L Institut Canadien De Science Et Technologie Alimentaires, v. 18, p. 290-295.
Hayakawa, S., and S. Nakai, 1985b, RELATIONSHIPS OF HYDROPHOBICITY AND NET CHARGE TO THE SOLUBILITY OF MILK AND SOY PROTEINS: Journal of Food Science, v. 50, p. 486-491.
Hedoux, A., S. Krenzlin, L. Paccou, Y. Guinet, M. P. Flament, and J. Siepmann, 2010, Influence of urea and guanidine hydrochloride on lysozyme stability and thermal denaturation; a correlation between activity, protein dynamics and conformational changes: Physical Chemistry Chemical Physics, v. 12, p. 13189-13196.
Ho, J. G. S., and A. P. J. Middelberg, 2004, Estimating the potential refolding yield of recombinant proteins expressed as inclusion bodies: Biotechnology and Bioengineering, v. 87, p. 584-592.
Ho, J. G. S., A. P. J. Middelberg, P. Ramage, and H. P. Kocher, 2003, The likelihood of aggregation during protein renaturation can be assessed using the second virial coefficient: Protein Science, v. 12, p. 708-716.
Hua, L., R. H. Zhou, D. Thirumalai, and B. J. Berne, 2008, Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding: Proceedings of the National Academy of Sciences of the United States of America, v. 105, p. 16928-16933.
Huang, Q. L., J. Zhao, Y. H. Tang, S. Q. Shao, G. J. Xu, and Y. M. Feng, 2007, The sequence determinant causing different folding behaviors of insulin and insulin-like growth factor-1: Biochemistry, v. 46, p. 218-224.
Jahn, T. R., and S. E. Radford, 2008, Folding versus aggregation: Polypeptide conformations on competing pathways: Archives of Biochemistry and Biophysics, v. 469, p. 100-117.
James, S., and J. J. McManus, 2012, Thermal and Solution Stability of Lysozynne in the Presence of Sucrose, Glucose, and Trehalose: Journal of Physical Chemistry B, v. 116, p. 10182-10188.
Javid, N., K. Vogtt, C. Krywka, M. Tolan, and R. Winter, 2007, Protein-protein interactions in complex cosolvent solutions: Chemphyschem, v. 8, p. 679-689.
Kawahara, K., and C. Tanford, 1966, VISCOSITY AND DENSITY OF AQUEOUS SOLUTIONS OF UREA AND GUANIDINE HYDROCHLORIDE: Journal of Biological Chemistry, v. 241, p. 3228-&.
Kelly, S. M., T. J. Jess, and N. C. Price, 2005, How to study proteins by circular dichroism: Biochimica Et Biophysica Acta-Proteins and Proteomics, v. 1751, p. 119-139.
Kendrick, B. S., B. S. Chang, T. Arakawa, B. Peterson, T. W. Randolph, M. C. Manning, and J. F. Carpenter, 1997, Preferential exclusion of sucrose from recombinant interleukin-1 receptor antagonist: Role in restricted conformational mobility and compaction of native state: Proceedings of the National Academy of Sciences of the United States of America, v. 94, p. 11917-11922.
Kita, Y., T. Arakawa, T. Y. Lin, and S. N. Timasheff, 1994, CONTRIBUTION OF THE SURFACE FREE-ENERGY PERTURBATION TO PROTEIN SOLVENT INTERACTIONS: Biochemistry, v. 33, p. 15178-15189.

Koppel, D. E., 1972, ANALYSIS OF MACROMOLECULAR POLYDISPERSITY IN INTENSITY CORRELATION SPECTROSCOPY - METHOD OF CUMULANTS: Journal of Chemical Physics, v. 57.

Kuehner, D. E., J. Engmann, F. Fergg, M. Wernick, H. W. Blanch, and J. M. Prausnitz, 1999, Lysozyme net charge and ion binding in concentrated aqueous electrolyte solutions: Journal of Physical Chemistry B, v. 103, p. 1368-1374.

Lefebvre, B. G., and A. S. Robinson, 2003, Pressure treatment of tailspike aggregates rapidly produces on-pathway folding intermediates: Biotechnology and Bioengineering, v. 82, p. 595-604.

Lim, W. K., J. Rosgen, and S. W. Englander, 2009, Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group: Proceedings of the National Academy of Sciences of the United States of America, v. 106, p. 2595-2600.

Liu, W., D. Bratko, J. M. Prausnitz, and H. W. Blanch, 2004, Effect of alcohols on aqueous lysozyme-lysozyme interactions from static light-scattering measurements: Biophysical Chemistry, v. 107, p. 289-298.

Liu, W., T. Cellmer, D. Keerl, J. M. Prausnitz, and H. W. Blanch, 2005, Interactions of lysozyme in Gdmchloride solutions from static and dynamic light-scattering measurements: Biotechnology and Bioengineering, v. 90, p. 482-490.

Louis-Jeune, C., M. A. Andrade-Navarro, and C. Perez-Iratxeta, 2012, Prediction of protein secondary structure from circular dichroism using theoretically derived spectra: Proteins-Structure Function and Bioinformatics, v. 80, p. 374-381.

Majumdar, R., P. Manikwar, J. M. Hickey, H. S. Samra, H. A. Sathish, S. M. Bishop, C. R. Middaugh, D. B. Volkin, and D. D. Weis, 2013, Effects of Salts from the Hofmeister Series on the Conformational Stability, Aggregation Propensity, and Local Flexibility of an IgG1 Monoclonal Antibody: Biochemistry, v. 52, p. 3376-3389.

Manikwar, P., R. Majumdar, J. M. Hickey, S. V. Thakkar, H. S. Samra, H. A. Sathish, S. M. Bishop, C. R. Middaugh, D. D. Weis, and D. B. Volkin, 2013, Correlating excipient effects on conformational and storage stability of an IgG1 monoclonal antibody with local dynamics as measured by hydrogen/deuterium-exchange mass spectrometry: Journal of Pharmaceutical Sciences, v. 102, p. 2136-2151.

Mason, P. E., J. W. Brady, G. W. Neilson, and C. E. Dempsey, 2007, The interaction of Gdm ions with a model peptide: Biophysical Journal, v. 93, p. L4-L6.

Middelberg, A. P. J., 2002, Preparative protein refolding: Trends in Biotechnology, v. 20, p. 437-443.

Mishra, R., R. Seckler, and R. Bhat, 2005, Efficient refolding of aggregation-prone citrate synthase by polyol osmolytes - How well are protein folding and stability aspects coupled?: Journal of Biological Chemistry, v. 280, p. 15553-15560.

Moglich, A., F. Krieger, and T. Kiefhaber, 2005, Molecular basis for the effect of urea and Gdm chloride on the dynamics of unfolded polypeptide chains: Journal of Molecular Biology, v. 345, p. 153-162.

Mossuto, M. E., B. Bolognesi, B. Guixer, A. Dhulesia, F. Agostini, J. R. Kumita, G. G. Tartaglia, M. Dumoulin, C. M. Dobson, and X. Salvatella, 2011, Disulfide Bonds Reduce the Toxicity of the Amyloid Fibrils Formed by an Extracellular Protein: Angewandte Chemie-International Edition, v. 50, p. 7048-7051.

Muschol, M., and F. Rosenberger, 1995, Interactions in undersaturated and supersaturated lysozyme solutions: Static and dynamic light scattering results: Journal of Chemical Physics, v. 103, p. 10424-10432.

Narayanan, J., and X. Y. Liu, 2003, Protein interactions in undersaturated and supersaturated solutions: A study using light and x-ray scattering: Biophysical Journal, v. 84, p. 523-532.

Neira, J. L., and M. Rico, 1997, Folding studies on ribonuclease A, a model protein: Folding & Design, v. 2, p. R1-R11.

Nicoud, L., N. Cohrs, P. Arosio, E. Norrant, and M. Morbidelli, 2015, Effect of polyol sugars on the stabilization of monoclonal antibodies: Biophysical Chemistry, v. 197, p. 40-46.

Noppert, A., K. Gast, M. MullerFrohne, D. Zirwer, and G. Damaschun, 1996, Reduced-denatured ribonuclease A is not in a compact state: Febs Letters, v. 380, p. 179-182.
O'Brien, E. P., R. I. Dima, B. Brooks, and D. Thirumalai, 2007, Interactions between hydrophobic and ionic solutes in aqueous Gdm chloride and urea solutions: Lessons for protein denaturation mechanism: Journal of the American Chemical Society, v. 129, p. 7346-7353.
Oh, M., K. H. Mok, and Y. Lim, 1998, A comparison of three dimensional structures of biosynthesized preproinsulin and insulin using NMR: Hanguk Nongwhahak Hoechi, v. 41, p. 572-577.
Ortore, M. G., R. Sinibaldi, F. Spinozzi, F. Carsughi, D. Clemens, A. Bonincontro, and P. Mariani, 2008, New Insights into Urea Action on Proteins: A SANS Study of the Lysozyme Case: Journal of Physical Chemistry B, v. 112, p. 12881-12887.
Pawar, A. P., K. F. DuBay, J. Zurdo, F. Chiti, M. Vendruscolo, and C. M. Dobson, 2005, Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases: Journal of Molecular Biology, v. 350, p. 379-392.
Pittz, E. P., and S. N. Timasheff, 1978, INTERACTION OF RIBONUCLEASE-A WITH AQUEOUS 2-METHYL-2,4-PENTANEDIOL AT PH 5.8: Biochemistry, v. 17, p. 615-623.
Provencher, S. W., 1982, CONTIN - A GENERAL-PURPOSE CONSTRAINED REGULARIZATION PROGRAM FOR INVERTING NOISY LINEAR ALGEBRAIC AND INTEGRAL-EQUATIONS: Computer Physics Communications, v. 27.
Qu, Y. X., C. L. Bolen, and D. W. Bolen, 1998, Osmolyte-driven contraction of a random coil protein: Proceedings of the National Academy of Sciences of the United States of America, v. 95, p. 9268-9273.
Raman, B., T. Ramakrishna, and C. M. Rao, 1996, Refolding of denatured and denatured/reduced lysozyme at high concentrations: Journal of Biological Chemistry, v. 271, p. 17067-17072.
Roberts, C. J., 2007, Non-native protein aggregation kinetics: Biotechnology and Bioengineering, v. 98, p. 927-938.
Roberts, D., R. Keeling, M. Tracka, C. F. van der Walle, S. Uddin, J. Warwicker, and R. Curtis, 2015, Specific Ion and Buffer Effects on Protein-Protein Interactions of a Monoclonal Antibody: Molecular Pharmaceutics, v. 12, p. 179-193.
Rossky, P. J., 2008, Protein denaturation by urea: Slash and bond: Proceedings of the National Academy of Sciences of the United States of America, v. 105, p. 16825-16826.
Rudolph, R., and H. Lilie, 1996, In vitro folding of inclusion body proteins: Faseb Journal, v. 10, p. 49-56.
Samuel, D., T. K. S. Kumar, G. Ganesh, G. Jayaraman, P. W. Yang, M. M. Chang, V. D. Trivedi, S. L. Wang, K. C. Hwang, D. K. Chang, and C. Yu, 2000, Proline inhibits aggregation during protein refolding: Protein Science, v. 9, p. 344-352.
Schneider, C. P., D. Shukla, and B. L. Trout, 2011, Arginine and the Hofmeister Series: The Role of Ion-Ion Interactions in Protein Aggregation Suppression: Journal of Physical Chemistry B, v. 115, p. 7447-7458.
Schoemaker, J. M., A. H. Brasnett, and F. A. O. Marston, 1985, EXAMINATION OF CALF PROCHYMOSIN ACCUMULATION IN ESCHERICHIA-COLI - DISULFIDE LINKAGES ARE A STRUCTURAL COMPONENT OF PROCHYMOSIN-CONTAINING INCLUSION-BODIES: Embo Journal, v. 4, p. 775-780.
Shao, Q., Y. B. Fan, L. J. Yang, and Y. Q. Gao, 2012, Counterion Effects on the Denaturing Activity of Gdm Cation to Protein: Journal of Chemical Theory and Computation, v. 8, p. 4364-4373.
Shaw, K. L., G. R. Grimsley, G. I. Yakovlev, A. A. Makarov, and C. N. Pace, 2001, The effect of net charge on the solubility, activity, and stability of ribonuclease Sa: Protein Science, v. 10, p. 1206-1215.
Shire, S. J., Z. Shahrokh, and J. Liu, 2004, Challenges in the development of high protein concentration formulations: Journal of Pharmaceutical Sciences, v. 93, p. 1390-1402.
Shugar, D., 1952, THE MEASUREMENT OF LYSOZYME ACTIVITY AND THE ULTRA-VIOLET INACTIVATION OF LYSOZYME: Biochimica Et Biophysica Acta, v. 8, p. 302-309.

Silvers, R., F. Sziegat, H. Tachibana, S. Segawa, S. Whittaker, U. L. Gunther, F. Gabel, J. R. Huang, M. Blackledge, J. Wirmer-Bartoschek, and H. Schwalbe, 2012, Modulation of Structure and Dynamics by Disulfide Bond Formation in Unfolded States: Journal of the American Chemical Society, v. 134, p. 6846-6854.

Sinibaldi, R., M. G. Ortore, F. Spinozzi, S. D. Funari, J. Teixeira, and P. Mariani, 2008, SANS/SAXS study of the BSA solvation properties in aqueous urea solutions via a global fit approach: European Biophysics Journal with Biophysics Letters, v. 37, p. 673-681.

Street, T. O., D. W. Bolen, and G. D. Rose, 2006, A molecular mechanism for osmolyte-induced protein stability: Proceedings of the National Academy of Sciences of the United States of America, v. 103, p. 13997-14002.

Strickler, S. S., A. V. Gribenko, T. R. Keiffer, J. Tomlinson, T. Reihle, V. V. Loladze, and G. I. Makhatadze, 2006, Protein stability and surface electrostatics: A charged relationship: Biochemistry, v. 45, p. 2761-2766.

Stumpe, M. C., and H. Grubmuller, 2007, Interaction of urea with amino acids: Implications for urea-induced protein denaturation: Journal of the American Chemical Society, v. 129, p. 16126-16131.

Tartaglia, G. G., A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti, and M. Vendruscolo, 2008, Prediction of aggregation-prone regions in structured proteins: Journal of Molecular Biology, v. 380, p. 425-436.

Tartaglia, G. G., and M. Vendruscolo, 2008, The Zyggregator method for predicting protein aggregation propensities: Chemical Society Reviews, v. 37, p. 1395-1401.

Tessier, P. M., H. R. Johnson, R. Pazhianur, B. W. Berger, J. L. Prentice, B. J. Bahnson, S. I. Sandler, and A. M. Lenhoff, 2003, Predictive crystallization of ribonuclease A via rapid screening of osmotic second virial coefficients: Proteins-Structure Function and Genetics, v. 50, p. 303-311.

Thomson, J. A., B. A. Shirley, G. R. Grimsley, and C. N. Pace, 1989, CONFORMATIONAL STABILITY AND MECHANISM OF FOLDING OF RIBONUCLEASE-T1: Journal of Biological Chemistry, v. 264, p. 11614-11620.

Timasheff, S. N., 1991, PROTEIN STABILIZATION BY LOW MOLECULAR WEIGHT COSOLVENTS: Cryobiology, v. 28, p. 521.

Timasheff, S. N., 1998, Control of protein stability and reactions by weakly interacting cosolvents: The simplicity of the complicated: Advances in Protein Chemistry, Vol 51, v. 51, p. 355-432.

Timasheff, S. N., 2002, Protein hydration, thermodynamic binding, and preferential hydration: Biochemistry, v. 41, p. 13473-13482.

Timasheff, S. N., and T. Arakawa, 1988, MECHANISM OF PROTEIN PRECIPITATION AND STABILIZATION BY CO-SOLVENTS: Journal of Crystal Growth, v. 90, p. 39-46.

Timasheff, S. N., and G. F. Xie, 2003, Preferential interactions of urea with lysozyme and their linkage to protein denaturation: Biophysical Chemistry, v. 105, p. 421-448.

Tsumoto, K., T. Arakawa, and L. Chen, 2010, Step-Wise Refolding of Recombinant Proteins: Current Pharmaceutical Biotechnology, v. 11.

Tsumoto, K., D. Ejima, I. Kumagai, and T. Arakawa, 2003, Practical considerations in refolding proteins from inclusion bodies: Protein Expression and Purification, v. 28, p. 1-8.

Uversky, V. N., 1993, USE OF FAST PROTEIN SIZE-EXCLUSION LIQUID-CHROMATOGRAPHY TO STUDY THE UNFOLDING OF PROTEINS WHICH DENATURE THROUGH THE MOLTEN GLOBULE: Biochemistry, v. 32, p. 13288-13298.

Vagenende, V., M. G. S. Yap, and B. L. Trout, 2009, Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol: Biochemistry, v. 48, p. 11084-11096.

Valente, J. J., K. S. Verma, M. C. Manning, W. W. Wilson, and C. S. Henry, 2005, Second virial coeffcient studies of cosolvent-induced protein self-interaction: Biophysical Journal, v. 89, p. 4211-4218.

Valle-Delgado, J. J., J. A. Molina-Bolivar, F. Galisteo-Gonzalez, and M. J. Galvez-Ruiz, 2011, Evidence of hydration forces between proteins: Current Opinion in Colloid & Interface Science, v. 16, p. 572-578.

Vallejo, L. F., and U. Rinas, 2004, Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins: Microbial Cell Factories, v. 3.

Velev, O. D., E. W. Kaler, and A. M. Lenhoff, 1998, Protein interactions in solution characterized by light and neutron scattering: Comparison of lysozyme and chymotrypsinogen: Biophysical Journal, v. 75, p. 2682-2697.

Vernaglia, B. A., J. Huang, and E. D. Clark, 2004, Guanidine hydrochloride can induce amyloid fibril formation from hen egg-white lysozyme: Biomacromolecules, v. 5, p. 1362-1370.

Wang, G. Z., X. Y. Dong, and Y. Sun, 2011, Peptide disulfides CGC and RKCGC facilitate oxidative protein refolding: Biochemical Engineering Journal, v. 55, p. 169-175.

Wang, L., S. K. Maji, M. R. Sawaya, D. Eisenberg, and R. Riek, 2008, Bacterial inclusion bodies contain amyloid-like structure: Plos Biology, v. 6, p. 1791-1801.

Weiss, M. S., G. J. Palm, and R. Hilgenfeld, 2000, Crystallization, structure solution and refinement of hen egg-white lysozyme at pH 8.0 in the presence of MPD: Acta Crystallographica Section D-Biological Crystallography, v. 56, p. 952-958.

Wen, L. L., Y. Chen, J. Liao, X. X. Zheng, and Z. N. Yin, 2015, Preferential interactions between protein and arginine: Effects of arginine on tertiary conformational and colloidal stability of protein solution: International Journal of Pharmaceutics, v. 478, p. 753-761.

Werner, M. H., G. M. Clore, A. M. Gronenborn, A. Kondoh, and R. J. Fisher, 1994, REFOLDING PROTEINS BY GEL-FILTRATION CHROMATOGRAPHY: Febs Letters, v. 345, p. 125-130.

Winter, J., H. Lilie, and R. Rudolph, 2002, Renaturation of human proinsulin - a study on refolding and conversion to insulin: Analytical Biochemistry, v. 310.

Woody, R. W., and A. Koslowski, 2002, Recent developments in the electronic spectroscopy of amides and alpha-helical polypeptides: Biophysical Chemistry, v. 101, p. 535-551.

Zangi, R., R. H. Zhou, and B. J. Berne, 2009, Urea's Action on Hydrophobic Interactions: Journal of the American Chemical Society, v. 131, p. 1535-1541.

Zhang, F., M. W. A. Skoda, R. M. J. Jacobs, R. A. Martin, C. M. Martin, and F. Schreiber, 2007, Protein interactions studied by SAXS: Effect of ionic strength and protein concentration for BSA in aqueous solutions: Journal of Physical Chemistry B, v. 111, p. 251-259.

Zhang, J., and X. Y. Liu, 2003, Effect of protein-protein interactions on protein aggregation kinetics: Journal of Chemical Physics, v. 119.

Zhang, Y. J., and P. S. Cremer, 2010, Chemistry of Hofmeister Anions and Osmolytes, *in* S. R. Leone, P. S. Cremer, J. T. Groves, M. A. Johnson, and G. Richmond, eds., Annual Review of Physical Chemistry, Vol 61: Annual Review of Physical Chemistry, v. 61: Palo Alto, Annual Reviews, p. 63-83.

Zou, Q., B. J. Bennion, V. Daggett, and K. P. Murphy, 2002, The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea: Journal of the American Chemical Society, v. 124, p. 1192-1202.