TOWARDS A MOLECULAR UNDERSTANDING OF PROTEIN SOLUBILITY

A Dissertation

by

RYAN MAHNKEN KRAMER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2011

Major Subject: Biochemistry

Towards a Molecular Understanding of Protein Solubility

TOWARDS A MOLECULAR UNDERSTANDING OF PROTEIN SOLUBILITY

A Dissertation

by

RYAN MAHNKEN KRAMER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	J. Martin Scholtz
Committee Members,	C. Nick Pace
	David Brems
	Michael Polymenis
	Sarah Bondos
Head of Department,	Gregory Reinhart

August 2011

Major Subject: Biochemistry

ABSTRACT

Towards a Molecular Understanding of Protein Solubility.

(August 2011)

Ryan Mahnken Kramer, B.S., Washington State University

Chair of Advisory Committee: Dr. J. Martin Scholtz

Protein solubility is a problem for many protein chemists including structural biologists and those developing protein pharmaceuticals. Knowledge of how intrinsic factors influence solubility is limited due to the difficulty in obtaining quantitative solubility measurements. Solubility measurements in buffer alone are difficult to reproduce, as gels or supersaturated solutions often form, making the determination of solubility values impossible for many proteins. Protein precipitants can be used to obtain comparative solubility measurements, and they fall into three broad classes: salts, long-chain polymers, and organic solvents.

Our group has used a model protein, RNase Sa, to create 20 variants that differ by the residues at a single surface-exposed position. We have measured the protein solubility of these variants and have generated an amino acid solubility scale, in the context of a protein, measured in ammonium sulfate. Here, we present solubility scales for these variants using PEG-8000 and isopropanol as precipitants. We find that amino acids can be divided into three groups based on their contribution to protein solubility: those that increase protein solubility, those that decrease protein solubility, and those

that show little change in protein solubility as compared to our wild-type protein which has a threonine at the variable position. Of the 20 variants used here, the aspartic acid, glutamic acid, and serine variants show the greatest increases in protein solubility. Based on our results, we propose a strategy for increasing protein solubility: substitute exposed hydrophobic, asparagine, glutamine, and threonine residues with aspartic acid, glutamic acid or serine. To test this hypothesis, we utilize this strategy on a low solubility variant of RNase Sa.

Here, we compare the use of representatives from two classes of precipitants, ammonium sulfate and polyethylene glycol 8000, by measuring the solubility of seven proteins. We find that increased negative surface charge correlates strongly with increased protein solubility and may be due to strong binding of water by the acidic amino acids. We also find that the solubility results obtained in the two different precipitants closely agree with each other, suggesting that the two precipitants probe similar properties that are relevant to solubility in buffer alone.

DEDICATION

To my grandfather, Carl C. Mahnken

ACKNOWLEDGEMENTS

I would like to thank the advisors in our lab, Dr. Marty Scholtz and Dr. Nick Pace for their help, support, patience, encouragement, and advice. I thank all my committee members for their council throughout the course of this project. I thank Dr. Dave Brems at AMGEN for his expert advice on protein solubility and for his personal career advisement. I would also like to thank Dr. David Geidroc for extending to me the opportunity to benefit from the Molecular Biophysics Training Program and associated NIH training grant. I would also like the thank all of the past members of the Pace/Scholtz labs whom I had the opportunity to learn from, particularly Dr. Jared Trefethen, Dr. Hailong Fu, Dr. Yun Wei, and Dr. Katherine Ridinger. I would like to individually thank Dr. Saul Trevino for his help from the very beginning of my graduate career during my rotation in the lab, for allowing me to inherit this project from him, and for his advice with this difficult project along the way. I thank Dr. Jerry Grimsley for all the support that he has offered including with editing parts of this written work. I thank the more than dozen rotation, undergraduate, and summer intern students that I have had the opportunity to mentor over the years at Texas A&M University; of these Nicole Motl and Varad Shende contributed a tremendous amount of time and energy to this project.

NOMENCLATURE

WT	Wild type
PEG	Polyethylene glycol
RNase Sa	Ribonuclease Sa
PDB	Protein Data Bank

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

NOMENCLATURE vii

TABLE OF CONTENTS viii

LIST OF FIGURES xi

LIST OF TABLES xiii

CHAPTER

I INTRODUCTION TO PROTEIN SOLUBILITY 1

Amorphous versus crystalline solubility 2
In vivo versus *in vitro* protein solubility 3
The protein solubility problem 3
Extrinsic factors that influence protein solubility 5
Measuring protein solubility 7
Protein solubility in precipitant solutions 8
Protein stability in the presence of precipitants 13
Initial protein concentration dependence of solubility measurements 15
Increasing protein solubility 18

II MATERIALS AND METHODS 21

Thermal denaturation experiments 21
Construction of RNase Sa variants 21
Expression and purification of RNase Sa variants 22
Preparation of protein stock solutions 22
Preparation of precipitant stock solutions 23

CHAPTER Page

Solubility measurements 23
Solubility as a function of pH 24
Accessible surface area calculations 24

III THE AMINO ACID CONTRIBUTION TO PROTEIN SOLUBILITY IN POLYETHYLENE GLYCOL AND ISOPROPANOL 26

Stability of RNase Sa variants 27
Effect of polyethylene glycol and isopropanol on RNase Sa stability 28
Amino acid solubility scale in polyethylene glycol 30
Amino acid solubility scale in isopropanol 33
Comparison of solubility scales in three types of precipitants 35
Strategy for increasing protein solubility 40
Charge and protein solubility 41
Summary 45

IV INCREASING PROTEIN SOLUBILITY BY REPLACING POLAR SURFACE RESIDUES WITH ASPARTIC ACID AND SERINE 47

Selecting candidate positions and designing mutations to increase the solubility of T76W RNase Sa 48
Stability of T76W RNase Sa variants 53
Solubility measurements of the serine variants of T76W RNase Sa 53
Solubility measurements of the aspartic acid variants of T76W RNase Sa (3K) 56
The effectiveness of serine surface mutations in increasing the solubility of T76W RNase Sa. 61
The effectiveness of aspartic acid surface mutations in increasing the solubility of T76W RNase Sa. 62
Making hydrophobic to hydrophilic surface mutations to increase protein solubility compared with this approach 64
Summary 64

CHAPTER Page

V TOWARDS A MOLECULAR UNDERSTANDING OF PROTEIN SOLUBILITY: INCREASED NEGATIVE SURFACE CHARGE CORRELATES WITH INCREASED SOLUBILITY. 66

Proteins are folded under experimental conditions 66
Solubility measurements rapidly reach equilibrium 71
Solubility curves in polyethylene glycol..... 71
Solubility curves in ammonium sulfate 74
Comparing protein solubility in PEG-8000 and ammonium sulfate 76
Correlation of molecular weight and net charge with solubility measurements..... 80
Correlation of the intrinsic properties of the accessible surface area with protein solubility 83
Conclusions 88

VI SUMMARY 90

REFERENCES 94

VITA 109

LIST OF FIGURES

FIGURE		Page
1	Theoretical protein solubility curves in ammonium sulfate and PEG.....	11
2	An equilibrium diagram showing the unfolding of a protein from natively folded protein, through a folding intermediate, to unfolded protein.....	14
3	Change in measured solubility as a function of initial protein concentration for α-lactalbumin in 27.5 % PEG-8000 and 1.7 M ammonium sulfate.....	16
4	The melting temperature of RNase Sa as a function of PEG-8000 and isopropanol concentration.....	29
5	The solubility of RNase Sa as a function of PEG-8000 and isopropanol.....	31
6	Normalized solubility values for the RNase Sa position 76 variants.....	37
7	The absolute value of the calculated net charge for RNase Sa variants T76D, WT, and T76R.....	42
8	Three views of RNase Sa showing candidate positions for substitution....	51
9	Relative solubility of T76W RNase Sa and serine variants at pH 4.25.....	55
10	Relative solubility of T76W RNase Sa (3K) aspartic acid variants at negative net charge.....	58
11	Relative solubility of T76W RNase Sa (3K) aspartic acid variants at positive net charge.....	60
12	Change in melting temperature as a function of ammonium sulfate concentration for RNase Sa, α-chymotrypsin, lysozyme, human serum albumin, ovalbumin, and α-lactalbumin.....	68

FIGURE Page

13 Change in melting temperature as a function of polyethylene glycol concentration for RNase Sa, α-chymotrypsin, lysozyme, human serum albumin, ovalbumin, and α-lactalbumin 69

14 The solubility of several proteins in polyethylene glycol-8000..... 72

15 The solubility of several proteins in ammonium sulfate 75

16 A comparison of the solubility data obtained in ammonium sulfate and PEG-8000 77

17 The correlation of molecular weight and net charge with PEG-8000 and ammonium sulfate solubility measurements..... 81

18 The correlation of fraction polar and nonpolar ASA with PEG-8000 and ammonium sulfate solubility measurements..... 84

19 The correlation of the fraction of ASA that is charged, positively charged, and negatively charged at pH 7.0 with PEG-8000 and ammonium sulfate solubility measurements..... 86

LIST OF TABLES

TABLE		Page
1	RNase Sa solubility measurements in 5%(w/v) PEG-8000 and 50 mM sodium acetate at pH 4.25	32
2	RNase Sa solubility measurements in 10% (v/v) isopropanol and 50 mM sodium acetate at pH 4.25	34
3	RNase Sa solubility measurements in 50 mM sodium acetate at pH 4.25	36
4	Solubility and charge properties of RNase Sa variants.	39
5	RNase Sa residues that have surface exposed side chains that contribute unfavorably to solubility	49
6	Melting temperatures for the serine and aspartic acid variants of T76W RNase Sa in 30 mM MOPS at pH 7.0	52
7	Solubility measurements of RNase Sa T76W serine variants in 50 mM sodium acetate pH 4.25.	54
8	Solubility measurements and net charge calculations for RNase Sa T76W (3K) apartic acid variants.	57
9	Fit values for PEG-8000 and ammonium sulfate solubility curves	73
10	Protein properties and surface properties used for correlations	79

CHAPTER I

INTRODUCTION TO PROTEIN SOLUBILITY

Protein solubility is an important consideration for structural biologists,[1] the pharmaceutical industry,[2] and all scientists that work with proteins in solution. Structural studies[1,3,4] and pharmaceutical applications[2,5,6] often require protein samples at very high concentration. Low protein solubility also plays a role in several human diseases.[7-12] The solubility of a protein in aqueous solution ranges from hundreds of milligrams per milliliter to completely insoluble. For example, many serum albumins have solubilities greater than 500 milligrams per milliliter.[13] Conversely, crambin, a hydrophobic protein and member of the plant toxin family thionin, is reported to be completely insoluble in water; though, many other members or this protein family are water soluble.[14,15]

Operationally, protein solubility is the concentration of protein in a saturated solution that is in equilibrium with a solid phase, either crystalline or amorphous, under a given set of conditions.[16,17] Arakawa and Timasheff rigorously defined protein solubility as a thermodynamic parameter.[17] They noted that at equilibrium the chemical potential of the protein in the solid and solution phases must be equal,

---

This dissertation follows the style of the *Journal of Molecular Biology*.

$$\mu_{p,l} = \mu_{p,s} \qquad (1)$$

where $\mu_{p,l}$ and $\mu_{p,s}$ are the chemical potential of the protein in the solution (liquid) and solid phases, respectively. Assuming that the chemical potential of the solid phase remains a constant, solubility is defined by the following equation:

$$a_{p,l} = \gamma_{p,l} S_{p,l} = \text{constant} \qquad (2)$$

where $\gamma_{p,l}$ and $S_{p,l}$ are the activity coefficient and the solubility (concentration) of the protein in the liquid phase, and $a_{p,l}$ is the activity of the protein in the solution phase as defined by:

$$\mu_{p,l} = \mu_0 + RT \ln a_{p,l} \qquad (3)$$

The assumption that the chemical potential of the solid phase is a constant generally holds for two component systems (water and protein); however, as will be discussed later, this does not necessarily hold for three component systems when the third component is a precipitant.

Amorphous versus crystalline solubility

The solubility of a two phase system at equilibrium can be defined for both amorphous and crystalline solid phases.[17] The formation of a highly ordered crystalline solid phase, especially if one wishes to obtain large crystals that produce high quality x-ray diffraction patterns, often requires the formation of a meta-stable supersaturated protein solution that slowly comes to equilibrium. In contrast, the formation of an amorphous solid phase does not require a supersaturated solution, is not highly ordered, and upon precipitation can reach equilibrium almost immediately.[18] Also, as we have

noted in a recently published review, amorphous protein is often more soluble and is of more pharmacological and general experimental significance.[19,20]

In vitro versus *in vivo* protein solubility

The term low protein solubility is sometimes used in the context of proteins that express poorly upon recombinant expression in *E. coli*, and several reviews have been written on this topic.[4,21-24] In this case, the use of the term low protein solubility is usually inaccurate, because what is often happening is that, upon cell lysis and subsequent centrifugation, the protein is found in the pellet, perhaps in inclusion bodies, not solubilized in the aqueous fraction. This is more likely caused by issues relating to protein stability than to poor *in vitro* protein solubility as defined in the thermodynamic sense.[21] While the study of poor recombinant protein expression *in vivo* is an important topic, this study focuses on the thermodynamic quantity that is *in vitro* protein solubility, which is more relevant to structural studies and pharmaceutical applications.

The protein solubility problem

Structural biology

Structural studies often require protein samples at very high concentration. For example, it is ideal for proteins to have a minimum solubility of approximately 1 mM for characterization by NMR experiments.[1] Often target proteins are not soluble enough for structural determination, and steps must be taken in order to increase the solubility of the protein or else the experiment cannot proceed. Bagby[1] discusses how one could go about screening solutions conditions for maximizing protein solubility in NMR

experiments, but this approach is equally applicable to other experiments where it is desirable to increase protein solubility.

High concentration protein formulations

Low protein solubility poses a significant problem for the formulation of proteins for use as pharmaceuticals that need to be stable during the shipping, storage, and administration processes.[2] Of particular interest are monoclonal antibodies (MAbs) which have been used to treat a wide variety of conditions including cardiovascular disease, several different types of cancer, asthma and other autoimmune disorders, and allergic reactions.[25-33] The most convenient route of administration is subcutaneous injection via a syringe that the patient can use at home.[34] However, the required dosages are often on the order of milligrams of protein per kilogram of body weight, so this quickly leads to the desire to create formulations that are in the hundreds of milligrams per milliliter, and for most MAbs this is difficult to achieve.

Low protein solubility in human disease

Low protein solubility has been implicated in a number of human diseases including sickle cell anemia, cataracts, and amyloid formation (a type of low protein solubility) diseases including Alzheimer's disease and dialysis related amyloidosis.[7-12] Disease causing mutations in human γD-crystalin (HGD) show a markedly lowered solubility compared with WT protein.[8] The *in vitro* solubility of HGD is over 400 mg/ml; however, the solubility of the P23T mutant of HGD is only 1-2 mg/ml and leads to childhood onset of cataracts.[8] This drastic decrease in solubility does not appear to coincide with any significant structural changes.

Extrinsic factors that influence protein solubility

Many intrinsic and extrinsic factors affect protein solubility. Extrinsic factors that influence solubility include pH,[35] temperature,[8] ionic strength,[17,36,37] and the presence of other molecules that either increase (solubilizers)[1,38] or decrease (precipitants)[39] protein solubility. Intrinsic factors that affect protein solubility include the properties of the surface-exposed amino acids,[37] protein net charge,[17] and the conformational state of the protein.[40] There have been several studies on the effects of extrinsic factors on protein solubility;[4,20,41] but a detailed understanding of how to alter the intrinsic properties of a protein in order to increase protein solubility is lacking,[1,4,42] and this is the major focus of this dissertation.

Temperature

Protein solubility is very sensitive to changes in temperature, but it is complicated because changes in thermal energy can influence both the dielectric constant and molecular motions of the protein in both the solid and solution phases.[20] For most proteins, the solubility increases with temperature until thermal unfolding occurs and solubility decreases at lower temperatures. However, many exceptions to this behavior have been noted.[43] For example, Pande et al. studied the solubility of HGD as a function of temperature and found that, while the solubility of native HGD increased with temperature, the solubility of several cataract forming variants of HGD exhibited a decrease in solubility with increasing temperature and a significant increase in solubility at low temperatures. This inverted solubility behavior is also displayed by hemoglobin S in sickle-cell anemia.[12]

pH

How the solubility of a protein changes as a function of pH is has been reasonably well studied.[16,44-46] In general, the solubility displays a minimum near the pI and increases proportionally to the square of the net charge of the protein.[44] This is due to the fact that the interactions between protein molecules should be the strongest when the net charge on the protein is zero. However, this situation can be complicated by pH-dependent conformational changes, binding of counter-ions, and the inability to accurately calculate the net charge of a protein.[47,48]

Solution additives

Small molecules added to a solution can have various effects on protein solubility. Many denaturing agents, such as guanidine and urea,[49] and sugars[17] increase protein solubility. Organic solvents,[39] and long chain polymers[50] usually decrease protein solubility. Many other classes of small molecules, such as amino acids,[38,43] salts,[17,51] and other surfactants, can have a variable effect on protein solubility depending on the identity of the molecule and the specific interactions with the protein . The effect on solubility of these various types of molecules is complex and they often influence several solution properties including surface tension, direct binding to polar and non-polar regions on the protein surface, preferential hydration, and excluded volume. A unique example are the protic ionic liquids (PILs) including ethylammonium nitrate and trimethylammonium methanesulfonate. Byrne and Angell investigated protein solubility as a function of PIL concentration and found that the solubility profile was complex with

a maximum and multiple minima.[52] They used this information to grow high diffraction quality crystals using a unique rehydrating method.

Measuring protein solubility

Measuring protein solubility requires a saturated protein solution that contains a solid phase. If this condition can be met, solubility measurements can be made by simply removing the solid phase by centrifugation and measuring the concentration of protein in the supernatant. A saturated protein solution is often difficult to achieve, however, due to the high solubility of many proteins and the non-ideal behavior exhibited by protein solutions at high concentrations.

Concentration by ultrafiltration

Ultrafiltration, with an appropriate molecular weight cutoff, can be used to attain a more concentrated protein solution,[53,54] but this can result in the formation of a gel or suspension, a supersaturated solution, or protein crystallization.[20] In the case where gel formation or suspension occurs, the protein may become kinetically trapped in this poorly understood state, and determination of solubility may be impossible. In the case of supersaturation, the protein concentration goes above the solubility limit and may become somewhat stable at a high concentration for a period of time. However, the solution is not at equilibrium and the protein will eventually begin to precipitate over time, but this may take hours, days, weeks, or longer. When crystallization occurs, protein solubility can be determined for the crystalline state; though, this is often very different from the amorphous solubility measurement.[19]

Addition of lyophilized powder to solvent

For proteins with low solubility, saturated solutions can sometimes be achieved by the addition of lyophilized protein to solvent;[4,8,53] however, for proteins with higher solubility, this is often not the case. One problem inherent with this method is that the water and buffer content of the lyophilized powder is a difficult and important variable to be able to control, and extensively freeze-dried samples become very difficult to dissolve.[20] As with ultrafiltration, many of the same non-ideal solutions often result including gel formation, suspensions and supersaturated solutions. In cases where this approach is successful in obtaining a saturated solution, dynamic light scattering may be useful in determining the exact concentration at which precipitation starts to occur.[52,55] Both ultrafiltration and addition of lyophilized powder to solvent may require a prohibitively large quantity of protein, as protein solubilities can be in the hundreds of milligrams per milliliter.

Protein solubility in precipitant solutions

One way to avoid the difficulties of measuring protein solubility seen with the methods discussed above is to make use of an extraneous agent that lowers the solubility of a protein called a precipitant. [16,39,50,51,56] These precipitants are the same ones used by crystallographers to achieve slow precipitation and crystal formation; however, they can also be used to induce amorphous precipitation by direct mixing with protein solutions.[19,37,51,57-62] Protein precipitants can be divided into three main classes: salts, organic solvents, and polymers.[39] Common examples from each of these three classes of precipitants are ammonium sulfate, isopropanol, and polyethylene glycol. Each class of

precipitants reduces protein solubility by a different mechanism and will be discussed below. The relationship between precipitant concentration and protein solubility is described by an equation of the following form:[16,57]

$$\text{Log } S = \text{Log } a_o - \beta[\text{Precipitant}] \qquad (4)$$

where S is the measured solubility at a given concentration of precipitant and β is the dependence of solubility on precipitant concentration for a given protein. Log a_0 represents the y-intercept of the solubility plot and is a constant that may or may not be related to solubility or activity in the absence of precipitant depending on the identity of the precipitant and the behavior of the protein.

One benefit of this method is that a much lower quantity of protein is needed in order to make solubility measurements. Precipitants are useful for comparing protein solubility, but solubility values obtained with a given precipitant are only relevant for the solid state produced by that precipitant.[20,42] Solid states produced using a precipitant may be different from those in buffer alone. Therefore, solubility results obtained by the use of precipitants are best used in a comparative manner and not as accurate predictions of solubility in the absence of precipitant.

Salts

Salts are the most commonly used class of protein precipitant.[37,56,58] At low concentrations salts increase the solubility of a protein by ion-screening as described by the Debye-Hückel theory.[20,63] This behavior at low ionic strength is known as salting-in. At high concentrations, chaotropic ions (such as chloride ions) increase the protein solubility due to decreased surface tension or may decrease protein solubility if direct

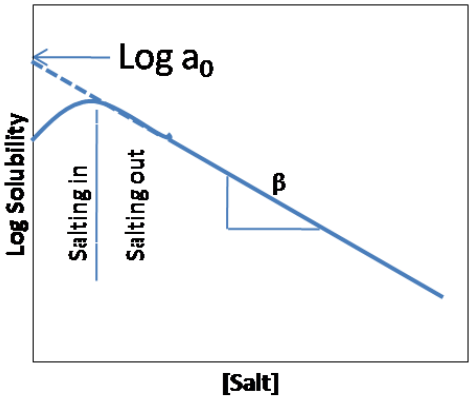
binding and charge abolishment occurs.[64] At high concentrations, kosmotropic ions bind water tighter than water binds itself,[64] and the surface tension of the solution increases, effectively competing with the surface of the protein for water molecules for hydration. As less water becomes available to hydrate the protein surface, the protein molecules self associate and precipitate.[16]

Cohn and Edsall described the relationship between salt concentration and protein solubility for salting-out, and it follows the form of equation 4.[16] Figure 1(a) shows a theoretical solubility curve for a protein in the presence of a kosmotropic salt. Upon addition of low concentrations of salt, salting-in is observed as the solubility increases. Once a certain concentration of salt is reached, salting-out occurs, and the logarithm of protein solubility begins to decrease linearly with respect to salt concentration as described by equation 4. The change in solubility as a function of salt is described by β and is a constant for a given salt and protein pair. Log a_0 is a constant and represents the projected y-intercept of the salting-out region; however, a relationship between log a_o and protein solubility in the absence of precipitant has not been established.

Long-chain polymers

Another common class of protein precipitants is long-chain polymers, and the most commonly used long-chain polymer is polyethylene glycol,[20,51,62,65-67] but other examples include polyamines and Jeffamines.[39] Long-chain polymers occupy more space in solution than a protein of similar molecular weight and lower the solubility of a protein through an excluded volume mechanism in that the proteins are sterically

(a)



(b)

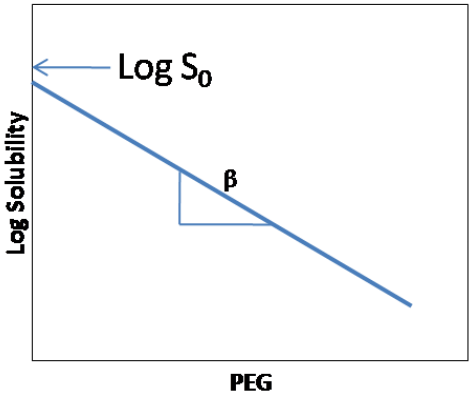


Figure 1. Theoretical protein solubility curves in ammonium sulfate (a), and PEG (b). At low concentrations of ammonium sulfate, protein solubility increases in a phenomenon known as salting-in. At higher concentrations, salting-out occurs and protein solubility decreases with a change of β. Protein solubility as a function of PEG concentration (usually measured in mg/ml) decreases linearly at high and low concentrations of PEG by β. The y-intercept of PEG precipitation curves can be used to estimate protein solubility in the absence of precipitant, whereas the y-intercept of an ammonium sulfate precipitation curve cannot (see text).

excluded from the volume of solvent occupied by the polymers,[50,51] effectively crowding the proteins out of solution. Middaugh et al.[57] showed that for polyethylene glycol precipitations, the linearity of equation 4 extends to zero precipitant for proteins whose solubility could be accurately measured in buffer alone. In this case:

$$a_0=\gamma_0 S_0 \qquad (5)$$

where a_0 is the constant from equation 4 and is the activity in the absence of precipitant, and γ_0 and S_0 are the activity coefficient and solubility, respectively, in the absence of precipitant. Note that as γ_0 approaches 1, equation 4 becomes:

$$\text{Log } S = \text{Log } S_o - \beta[\text{Precipitant}] \quad (6)$$

If $\gamma_0 < 1$, equation 6 will yield an underestimate of solubility in the absence of precipitant. Figure 1(b) shows a theoretical solubility curve for a protein in the presence of PEG. The solubility is described by equation 5 and, in contrast to salting-out, is linear at low PEG concentrations. The change in solubility with respect to PEG concentration is described by β, and the solubility in the absence of precipitant can be estimated from the y-intercept (Log S_0). It should be stressed that even S_0 obtained from PEG experiments should only be used qualitatively and comparatively, as the solid state may be different from in the absence of precipitant. It should also be pointed out that in cases where there is homotypic associate between protein molecules that the linearity of the plot of Log solubility versus PEG concentration is not linear.[57]

Organic solvents

Organic solvents, such as alcohols, are used for crystal formation in crystallography and for precipitation of proteins during purification. Organic solvents lower the dielectric constant of the solution. As the dielectric constant decreases, the solution becomes a poorer solvent for the protein. Consequently, the relative favorability of protein-protein interactions increase and the protein precipitates.[39] Alcohols do not exhibit "salting-in" at low concentrations, so solubility curves are expected to look like PEG precipitations curves shown in Figure 2(a).

Protein stability in the presence of precipitants

Solubility is an equilibrium measurement of the concentration of protein in solution in the presence of a solid phase; however, this equilibrium becomes very complex (see Figure 2) if we cannot confine our experiments to examining only native protein in solution with native protein as the solid phase. As a protein begins to unfold, the unfolded protein and any folding intermediates become populated. The solubility of the unfolded protein, folding intermediates, or even different conformers of the protein may be significantly different from the natively folded protein.[40] Therefore, in order to make meaningful solubility measurements that we can interpret in terms of native protein solubility, we need to make certain that proteins remain folded under experimental conditions. A simple way to do this is by performing thermal unfolding experiments as a function of precipitant to make sure that our protein of interest remains folded over the range of precipitant concentrations used.

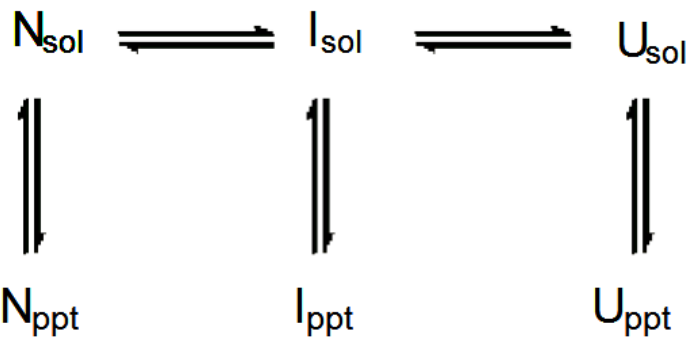


Figure 2. An equilibrium diagram showing the unfolding of a protein from natively folded protein (N_{sol}), through a folding intermediate (I_{sol}), to unfolded protein (U_{sol}). The equilibrium diagram also depicts the precipitation of the different folded species to their respective solid phases. The solubility of a protein can be significantly different between folding conformations.

The three classes of protein precipitants have different influences on protein stability. Kosmotropic salts increase protein stability by preferential hydration and by increasing the surface tension of bulk water.[63] Long-chain polymers, such as polyethylene glycol, have little effect on protein stability, [50,51] but in some cases low molecular weight PEG molecules decrease protein stability.[68] In the case of ammonium sulfate, sulfate ions stabilize proteins by salting-out the hydrophobic groups of the protein interior.[69] Alcohols and other organic solvents are known to denature proteins. Thomas and Dill[70] investigated the mechanism by which alcohols destabilize proteins and found that it was complex and dependent upon protein sequence and structure. They concluded that alcohols destabilize proteins mainly by weakening hydrophobic interactions. Special attention must be taken to insure that proteins remain folded under experimental conditions if organic solvents are used as precipitants, and lower temperatures may need to be used to achieve this.[39]

Initial protein concentration dependence of solubility measurements

For a well behaved protein system, the concentration of protein in solution increases linearly, with a slope of one, with respect to total protein in the system until the solubility value is reached. At this point, the protein in solution remains constant as total protein increases, and the remaining protein becomes part of the solid phase. When a protein is mixed with a precipitant, the total protein in the system is determined by the initial concentration of protein in the sample, and the measured solubility is the amount of protein remaining in solution after precipitation. If the system is well behaved, the solubility measurement will be independent of the initial protein concentration; however,

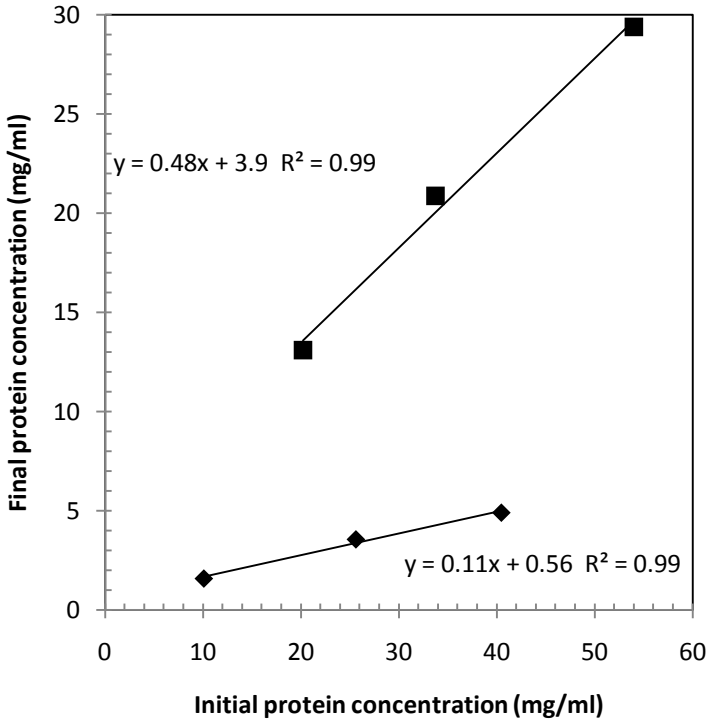


Figure 3. Change in measured solubility (final protein concentration) as a function of initial protein concentration for α-lactalbumin in 27.5 % PEG-8000 (diamonds) and 1.7 M ammonium sulface (squares). The slope of the line is $\Delta S/\Delta C_i$ and describes how strongly the measured solubility depends on the total protein. On average, we find that polyethylene glycol measurements have a smaller $\Delta S/C_i$ than ammonium sulfate.

the measured protein solubility frequently does depend on initial protein concentration when protein precipitation is induced by a precipitant. If the solubility of a protein is highly dependent upon initial protein concentration, this can make the determination and comparison of solubility values complicated. This phenomenon has been reported for protein precipitations by salts[56,58] and by polyethylene glycols.[57] The explanation for initial protein concentration dependence is unknown, and the phenomenon is poorly understood and impossible to predict for a given protein. Shih et al.[56] has studied this phenomenon in detail using salts as precipitants. They find that for proteins that show initial protein concentration dependence, there is a change in the composition of the solid phase as a function of total protein. They propose a distribution coefficient to explain the relationship between the amount of protein in solution and the solid phase. Protein solubility is a constant assuming that the chemical potential of the solid phase is also a constant.[17] If the activity of the solid phase of a protein changes with initial protein concentration, then the solubility will also be variable.

Here we propose a way to quantify initial protein concentration dependence by measuring the solubility of a protein as a function of initial protein concentration, with a fixed precipitant concentration. Measured solubility versus initial protein concentration displays a strongly linear result. The slope is the dependence of solubility on initial protein concentration ($\Delta S/\Delta C_i$). $\Delta S/\Delta C_i$ can vary between 0 and 1 with 0 indicating no observed initial protein concentration dependence and 1 indicating a maximal dependence on initial protein concentration (ie an increase of 1 mg/ml of initial protein concentration correlates to an increase in measured solubility of 1 mg/ml). We consider

a $\Delta S/\Delta C_i$ of less than 0.1 to indicate that the initial protein concentration dependence is insignificant and within the 5-10 % error typically observed for solubility measurements.

As an example, Figure 3 shows the initial protein concentration dependence for α-lactalbumin in 1.7 M ammonium sulfate and 27.5 % (w/v) PEG-8000. At 1.7 M ammonium sulfate $\Delta S/\Delta C_i = 0.48$ indicating that solubility measurements made under these conditions are strongly dependent on initial protein concentration. However, at 27.5 % (w/v) PEG-8000, the results are much different. A $\Delta S/\Delta C_i$ of 0.11 is obtained which indicates that solubility only slightly depends on initial protein concentration. Therefore, for α-lactalbumin solubility measurements determined in PEG-8000 could be used comparatively with much more confidence than solubility determined in ammonium sulfate. On average, we have observed a higher level of initial protein concentration dependence in ammonium sulfate than in PEG-8000.

Increasing protein solubility

Varying extrinsic factors that influence protein solubility such as pH, ionic strength, temperature, and the presence of different solvent additives can lead to increased solubility.[1,4] One particularly effective approach has been to add small quantities of arginine or glutamic acid to the solution.[38] The mechanism of how this increases solubility is not fully understood, but it may involve interactions between charges on the free amino acids and oppositely charged groups on the protein surface and interactions between hydrophobic portions of the free amino acids and the hydrophobic patches on the protein surface.

Altering these solution conditions, however, is not always appropriate, and it is often insufficient to increase protein solubility to the extent required. It then becomes necessary to make mutations in order to alter the intrinsic properties of a protein. In our group's recent review or protein solubility,[19] we noted that the commonly used approach of making surface hydrophobic to hydrophilic mutations[71-81] may not be the most effective strategy, even if the structure is known, because of the wide range of contributions to protein solubility of hydrophilic amino acid residues.[37] Also, if the structure of the protein is not known, hydrophobicamino acids are more likely to be buried than on the surface of the protein.[82] Consequently, a detailed understanding of how to alter the intrinsic properties of a protein in order to increase protein solubility is lacking. [1,4,42] In a recent study, our lab has taken steps towards understanding the intrinsic factors that influence protein solubility with the goal of developing a strategy for making mutations that increase protein solubility.[37] In this study, an amino acid solubility scale was created for a surface exposed position in a protein using ammonium sulfate. It was found that aspartic acid, glutamic acid and serine contribute the most favorably to protein solubility while other polar amino acids including threonine, asparagine and glutamine contribute unfavorably to solubility. Our proposed strategy for making mutations that increase protein solubility will be discussed in more detail in subsequent chapters; in short we suggest targeting exposed hydrophobic residues when available and threonine, aspargine and glutamine residues and making mutations to aspartic acid, glutamic acid and serine. In the proceeding chapters, data will be presented and discussed that further develops and tests our strategy for increasing

protein solubility including amino acid solubility scales in other precipitants, a test of our strategy to increase protein solubility of a low solubility protein, and a study of how the surface properties of a group of proteins correlate with solubility.

CHAPTER II

MATERIALS AND METHODS

Thermal denaturation experiments

Thermal denaturation experiments were performed on AVIV spectrophotometers with either a 62DS or a 202SF model using a 3 minute equilibration time, a 30 second averaging time, and a 1 cm pathlength cuvette. The solution was at pH 7.0 in 50 mM sodium phosphate for all proteins except RNase Sa which was at pH 7.0 in 30 mM MOPS due to the fact RNase Sa binds phosphate ions. Elipticity was monitored at 222 nm for ovalbumin, α-chymotrypsin, lysozyme, and human serum albumin with a protein concentration of 0.025 mg/ml. Elipticity was monitored at 234 nm for RNase Sa and 270 nm for α-lactalbumin with protein concentrations of 0.1 mg/ml and 0.5 mg/ml respectively. The data were analyzed using KaleidaGraph version 3.52, Using methods that have been described elsewhere.[83]

Construction of RNase Sa variants

The position 76 variants of RNase Sa were obtained from and constructed by Trevino *et al.*[37] The following variants were constructed using a QuikChange Site Directed Mutagenesis Kit from Stratagene or Agilent starting with the T76W construct: N20S + T76W, T46S + T76W, Q77S + T76W, N20D + T76W (3K), T46D + T76W (3K), and Q77D + T76W (3K), were 3K is a basic variant of RNase Sa containing the D1K, D17K, and E41K mutations.

Expression and purification of RNase Sa variants

RNase Sa was purified as described Hebert *et al.*[84] In summary, plasmid DNA was transformed into the MQ *E. coli* cell line and a single colony was used to inoculate terrific broth (TB) media. The cultures were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at an optical density of approximately 1.0 and allowed to grow overnight at 30 ˚C. Cultures were then centrifuged, cell pellets were resuspended in a 20% sucrose solution, centrifuged, resuspended in a non-sucrose solution, centrifuged, and supernatants were pooled from both the sucrose and the non-sucrose washes. A 50 mM succinic acid pH 3.25 acid precipitation step followed, and the supernatant was applied to an SP Sephadex C25 cation exchange column and eluted with a pH 3.25 to pH 8.0 gradient. The samples were pooled by checking fractions for RNase activity and comparing to the elution profile. Pooled samples were lyophilized, subsequently resuspended in ammonium bicarbonate, and applied to a Sephadex G50 size exclusion column. Fractions containing the protein were selected and lyophilized. Yields varied between 15 mg/L and 50 mg/L. Purity and identity were determined by SDS PAGE and mass spectrometry.

Preparation of protein stock solutions

Human fibrinogen, human serum albumin, bovine α-chymotrypsin, and bovine α-lactalbumin were obtained from Calbiochem, chicken lysozyme and ovalbumin were obtained from Sigma, and RNase Sa was purified as described above. Protein samples were dialyzed overnight into the corresponding buffer, which had been pH adjusted using NaOH or HCl, in Slide-A-Lyzer dialysis cassettes from Thermo Scientific.

Samples were then concentrated using Amicon Ultra centrifugal filters from Millipore. Concentrations were determined using the following extinction coefficients (ε_{280}= L cm^{-1} mg^{-1}) 1.55, 0.53, 2.04, 2.09, 2.60, 1.16, 0.71 for fibrinogen, human serum albumin, α-chymotrypsin, α-lactalbumin, lysozyme, and ovalbumin, respectively.[51,85] An extinction coefficient (ε_{278}= L cm^{-1} mg^{-1}) of 1.16 was used for all RNase Sa variants except for variants which contained the T76W or T76Y mutations for which 1.65 and 1.3 were used, respectively.

Preparation of precipitant stock solutions

The following precipitant stock solutions were prepared: 3.5 M ammonium sulfate (obtained from Sigma), 40% (w/v) PEG-8000 (obtained from Hamilton), and 40% (v/v) isopropanol (obtained from EMD). All precipitant solutions contained the buffer used in a given experiment and were adjusted to the pH and concentration of the experiment.

Solubility measurements

Solubility measurements were performed as described previously,[37] with some minor modifications. Three samples were prepared: precipitant and protein solutions as described above, and the corresponding buffer. These solutions were mixed together to achieve the desired protein and precipitant concentrations with final volumes between 10 and 100 μL, and they were allowed to equilibrate for approximately 10 minutes at room temperature (25 ˚C) for precipitations using ammonium sulfate and PEG-8000 and at 4˚C for precipitations using isopropanol. Samples were then centrifuged in an Eppendorf 5417R microcentrifuge at 16,000 RPM for 10 minutes. For samples where

visible precipitation and pelleting occurred, aliquots from the supernatant were taken, diluted as necessary with water, and the concentration was determined using an Agilent 8453 UV-Visible spectrophotometer.

Solubility as a function of pH

For the solubility measurements as a function of pH discussed in chapter III, experiments were performed as described above, but a fourth solution of acid with buffer was included. A multi-component buffer was used that was composed of 5 mM borate, 5 mM citrate, and 5 mM MOPS, and the proteins were dialyzed into this buffer at pH 7.0. The samples were prepared as described above, and varying amounts of acid were added to the individual precipitation experiments in order to achieve the desired pH. The pH was measured after mixing with a Thermo Scientific Micro Combination pH electrode and a Mettler Toledo pH meter prior to centrifugation.

Accessible surface area calculations

Accessibility data and hydrogen bonding information were determined using pfis (PDF file information software)[86] for all proteins except fibrinogen for which the program Naccess[87] was used. Both programs are based on the algorithm by Lee and Richards[88] and provide accessible surface area data for all atoms in the protein; however, pfis is only able to accept single peptide chains as input. The PDB files 3GHG,[89] 1E78,[90] 1YPH,[91] 1F6R,[92] 2VB1,[93] 1OVA,[94] and 1RGG[95] were used for fibrinogen, human serum albumin, α-chymotrypsin, α-lactalbumin, lysozyme, ovalbumin, and RNase Sa respectively. For the protein surface properties determined in Chapter V, fraction polar and nonpolar ASA, carbon, and sulfur atoms were defined as nonpolar and

nitrogen and oxygen atoms as polar. For fraction charged, positively charged, and negatively charged ASA calculations, the nitrogen atoms of the lysine and arginine side chains and the N-terminus were defined as positively charged and the oxygen atoms of the aspartic acid and glutamic acid side chains and the C-terminus were defined as negatively charged.

CHAPTER III

THE AMINO ACID CONTRIBUTION TO PROTEIN SOLUBILITY IN POLYETHYLENE GLYCOL AND ISOPROPANOL

Protein solubility is influenced by a number of intrinsic and extrinsic factors. Extrinsic factors that influence protein solubility have been well studied and include pH, ionic strength, temperature, and solvent additives.[3,4,17,20,41] Intrinsic factors that influence protein solubility are less well understood and are influenced by the amino acids on the surface of the protein which determine protein-protein and protein-solvent interactions. Knowledge of how the surface amino acid composition influences protein solubility is still incomplete,[1,4] and the goal of this study is to better understand the amino acid contribution to protein solubility in order to develop a mutational strategy for increasing protein solubility.

Recently, our lab has published an amino acid solubility scale for proteins in ammonium sulfate.[37] We used RNase Sa, a small ribonuclease, as a model system to study solubility. Position 76 of RNase Sa is a threonine residue that is entirely solvent exposed. We have replaced threonine 76 with all other 19 amino acids in order to study the effects of mutating a single surface residue on protein solubility. We found that in addition to surface-exposed hydrophobic residues, glutamine, threonine, and asparagine also contribute unfavorably to protein solubility. We also found that aspartic acid,

glutamic acid, and serine mutations cause the most favorable increases in solubility. Our data suggest that the solubility of a protein can be modulated through rational mutation of surface residues, and we are developing a strategy to increase protein solubility.

In addition to salts such as ammonium sulfate, alcohols and long chain polymers can be used as protein precipitants.[39] The mechanisms by which each class of precipitants lowers solubility is different, so results obtained with one precipitant may or may not be different from another precipitant. For example, long-chain polymers operate through an excluded volume mechanism,[50,51] kosmotropic salts operate through a salting-out mechanism,[16,64] and alcohols lower the dielectric constant of the solvent.[39] In order to gain a better understanding of the amino acid contribution to protein solubility in the absence of precipitant, it would be valuable to compare our results in ammonium sulfate to results obtained with the other two classes of precipitant: polymers and alcohols. Common trends are more likely to be applicable in the absence of precipitant. Here we have again used the position 76 variants to compose solubility scales in PEG-8000 (long-chain polymer) and isopropanol (alcohol) and compared them with our previous results in ammonium sulfate.

Stability of RNase SA variants

To determine if any of the mutations at position 76 significantly destabilized RNase Sa, thermal unfolding was monitored by circular dichroism. These experiments were performed and discussed in a previous study.[37] In summary, 19 of the 20 proteins showed a stabilization of 0.0 to 1.0 kcal mol^{-1}. One variant (T76C) showed a decrease in stability of 0.3 kcal mol^{-1} and a decrease in melting temperature from 47.8 °C to 46.9 °C.

These results show that the stability of RNase Sa has not been significantly changed by the mutations at position 76 and, importantly, that all the proteins are still folded at room temperature.

Effect of polyethylene glycol and isopropanol on RNase Sa stability

To confirm that the solubility measurements of RNase Sa observed are due to precipitation of the native state and not due to unfolding induced by precipitant, thermal unfolding experiments were performed in the presence of precipitants (Figure 4). As has been observed for other proteins,[50,51] polyethylene glycol 8000 does not have a significant effect on the stability of RNase Sa. Over the range of concentrations used (0 to 20% (w/v)), the melting temperature of RNase Sa increases linearly from 47.7 °C in the absence of PEG to 48.7 °C at 20% (w/v) PEG. For some proteins, however, low molecular weight polyethylene glycols[68] or high concentrations of polyethylene glycols[17] have been shown to decrease thermal stability.

Isopropanol and other alcohols destabilize proteins by decreasing the effect of hydrophobic burial.[70] As expected, isopropanol significantly decreases the stability of RNase Sa. Over the range of concentrations used (0 to 30% v/v), the melting temperature decreases linearly. At 30% (v/v) isopropanol the melting temperature is 24.5 °C, and approximately half of the protein would be unfolded at room temperature. To make certain that RNase Sa is folded in the presence of isopropanol, experimental conditions have to take into account the destabilizing effect of isopropanol. For

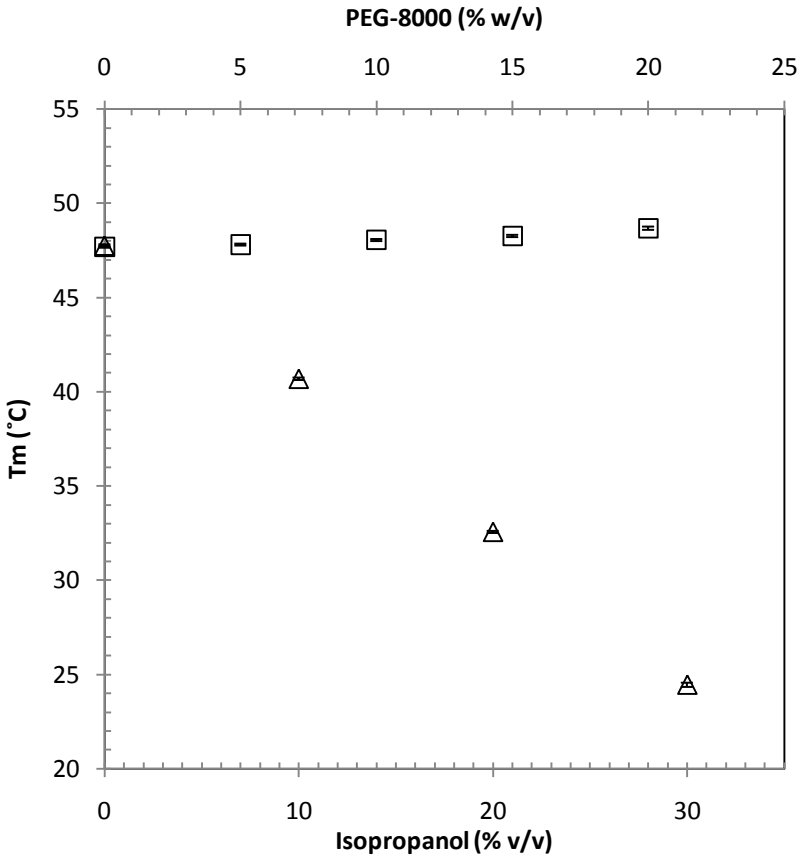


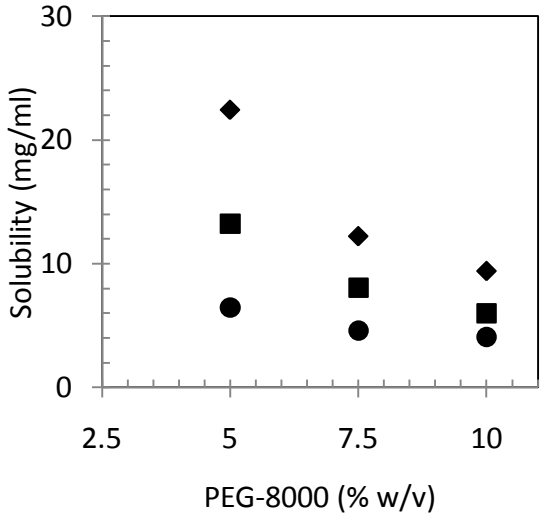
Figure 4. The melting temperature of RNase Sa as a function of PEG-8000 (squares) and isopropanol (triangles) concentration. The melting temperature was monitored by circular dichroism at 234 nm. Error bars are shown and are smaller than the data points.

example, at concentrations of isopropanol greater than 20% (v/v), a fraction of the protein may be unfolded at room temperature. Since we are interested in the solubility of folded RNase Sa, the experiment must be performed either below 20% (v/v) isopropanol or below room temperature. Experiments in this study were carried out at concentrations of isopropanol less than or equal to 20% (v/v) and at 4 °C.

Amino acid solubility scale in polyethylene glycol

Solubility curves for the 20 variants of RNase Sa were measured in PEG-8000 at pH 4.25. Solubility curves as a function of PEG-8000 are shown in Figure 5(a) for the aspartic acid, WT (theonine), and tryptophan variants of RNase Sa. The experimentally determined pI of RNase Sa is 3.5, so at pH 4.25 the net charge is expected to be negative (approximately -1 for WT and variants that do not add a charge).[84] The log of solubility is linear with respect to the concentration of PEG-8000 over the range of concentrations used; this has been reported for other proteins.[57] As similarly seen with ammonium sulfate,[37] the samples reach equilibrium almost immediately and are independent on initial protein concentration (data not shown). Table 1 displays the solubility of the 20 variants in 5% (w/v) PEG-8000. Due to the logarithmic relationship between protein solubility and precipitant concentration, the difference between two variants with similar slopes at a given precipitant concentration is a constant on the log scale but the difference decreases with increasing precipitant concentration. Therefore, it is best to compare variants at the lowest possible precipitant concentration. 5% (w/v) was chosen in this case because at concentrations lower than 5% (w/v) PEG-8000, it was

(a)



(b)

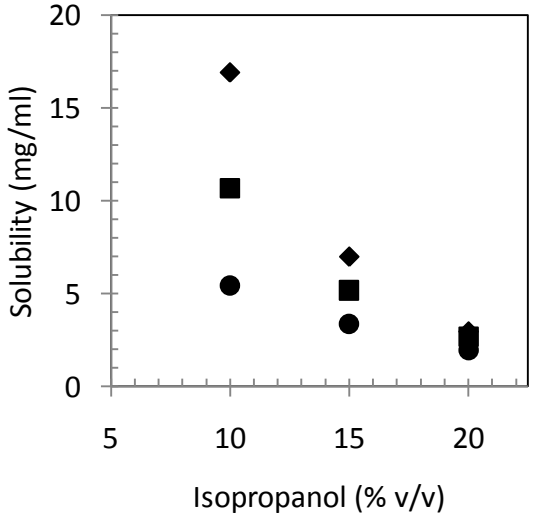


Figure 5. The solubility of RNase Sa as a function of PEG-8000 (a) and isopropanol (b). T76D (diamonds), WT (squares), and T76W (circles) are shown. Solubility varies logarithmically with precipitant concentration. The error is +/- 5%.

Table 1. RNase Sa solubility measurements in 5% (w/v) PEG-8000 and 50 mM sodium acetate at pH 4.25.

Amino acid at position 76	Solubility (mg/ml)[a]
Asp	22
Glu	21
Gly	18
Lys	17
His	17
Arg	15
Gln	15
Ser	15
Asn	14
Pro	14
Ile	14
Leu	14
Thr	13
Met	13
Val	12
Ala	11
Cys	11
Tyr	7
Phe	7
Trp	6

a- Error is +/- 5%

difficult to obtain concentrated enough samples in order to achieve precipitation for some of the variants. The solubility values vary over an approximately 4-fold range with the aspartic acid variant being the most soluble at 22 mg/ml and the tryptophan mutant being the least soluble at 6 mg/ml. The variants with negatively charged residues at the substitution site have the highest solubility and the variants with aromatic residues at the substitution site have the lowest solubility. The solubilities of most of the other variants falls into a somewhat narrow range. The solubility of WT RNase Sa (threonine at position 76) is 13 mg/ml and 12 out of 20 variants have similar solubility values in the range of 11-15 mg/ml. Only five variants have a significant increase in solubility (T76D, T76E,T76G, T76K, and T76H), and the three variants with aromatic substitutions (T76Y, T76F, and T76W) are the proteins only proteins that have a significant decrease in solubility. There does not appear to be a significant difference in the solubility of the proteins with non-aromatic, non-polar residues, and many of the polar residues at the substitution site.

Amino acid solubility scale in isopropanol

Solubility curves for the 20 variants of RNase Sa were measured in isopropanol at pH 4.25. Solubility curves as a function of isopropanol are shown in Figure 5(b) for the aspartic acid, WT (threonine), and tryptophan variants of RNase Sa. The log of solubility is linear with respect to the concentration of isopropanol over the range of concentrations used and is independent of initial protein concentration. Table 2 displays the solubility of the 20 variants RNase Sa in 10% (v/v) isopropanol. The solubility

Table 2. RNase Sa solubility measurements in 10% (v/v) isopropanol and 50 mM sodium acetate at pH 4.25.

Amino acid at position 76	Solubility (mg/ml)[a]
Asp	18
Glu	17
Gly	16
Pro	16
Lys	15
Ser	14
Asn	14
Gln	12
His	12
Ala	11
Thr	11
Met	10
Ile	9
Arg	9
Val	9
Leu	7
Tyr	6
Phe	6
Trp	5
Cys	3

a- Error is +/- 5%

values vary over an approximately 6-fold range with the aspartic acid variant being the most soluble at 18 mg/ml and the cysteine variant being the least soluble at 3 mg/ml. Similar to the PEG-8000 scale, the proteins with negatively charged residues at the variable site are the most soluble and the proteins with the aromatic residues at the variable site are some of the least soluble; however, the cysteine variant is the least soluble on the isopropanol scale. Though the cysteine variant has a higher solubility on the PEG-8000 and ammonium sulfate scales, it is still one of the least soluble in each of the scales. Unlike the PEG-8000 scale, the solubility of the 20 variants on the isopropanol scale is relatively continuous. The solubility of WT RNase Sa (threonine) is 11 mg/ml and is in the middle of the scale, both in terms of rank order and average solubility (also 11 mg/ml). T76D, T76E, T76G, T76P, T76K, T76S, and T76N have solubility values higher than WT, T76I, T76R, T76V, T76L, T76Y, T76F, T76W, and T76C have solubility values lower than WT, and T76Q, T76H, T76A, and T76M have solubility values approximately the same as WT.

Comparison of solubility scales in three types of precipitants

Table 3 shows the previously published solubility scale in 1.1 M ammonium sulfate[37] along with the solubility scales presented in this study. Figure 6 shows the normalized solubility data using ammonium sulfate, PEG-8000, and isopropanol as precipitants. The data are normalized to the solubility of WT protein (threonine at position 76) for each respective scale and are ordered by average normalized solubility

Table 3. RNase Sa solubility measurements in 50 mM sodium acetate at pH 4.25.

Amino acid at position 76	Solubility (mg/ml)[a]		
	1.1 M Ammonium Sulfate[b]	5 % PEG-8000	10 % Isopropanol
Asp	43	22	18
Arg	42	15	9
Glu	42	21	17
Ser	39	15	14
Lys	31	17	15
Gly	27	18	16
Ala	27	11	11
His	24	17	12
Asn	21	14	14
Thr	20	13	11
Gln	20	15	12
Pro	15	14	16
Cys	12	11	3
Met	11	13	10
Val	10	12	9
Leu	9	14	7
Ile	8	14	9
Tyr	6	7	6
Phe	4	7	6
Trp	4	6	5

[a]Error is +/- 5%
[b]Data from Trevino et al.[37]

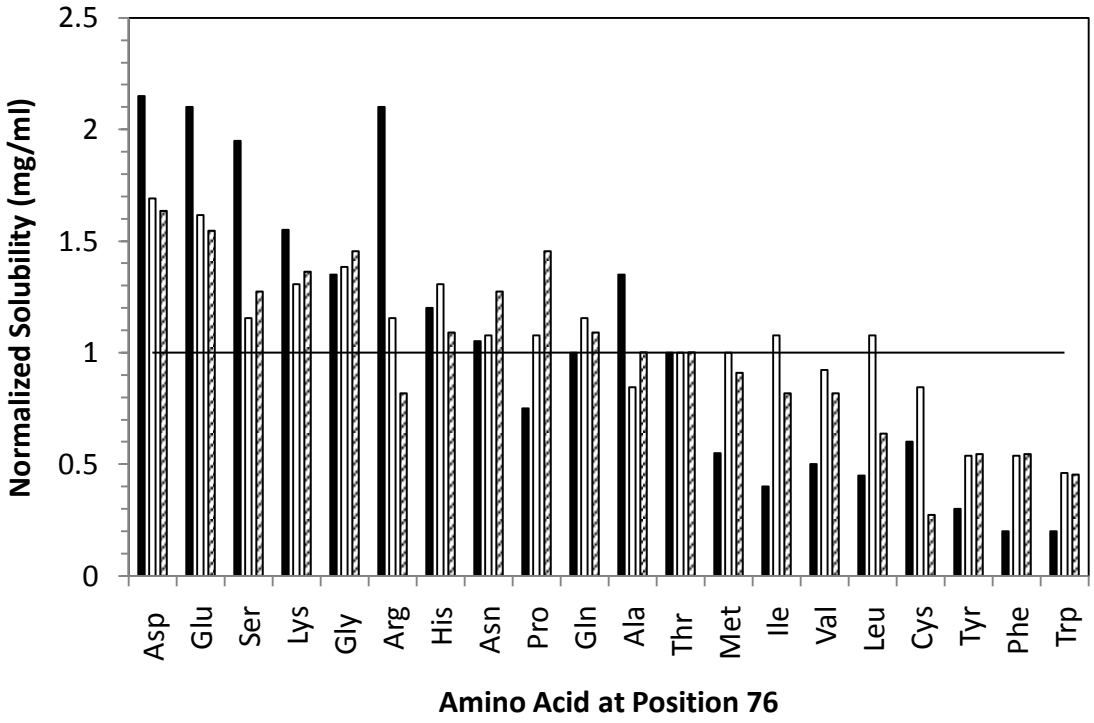


Figure 6. Normalized solubility values for the RNase Sa position 76 variants. Data were normalized to the WT protein from the respective scale. Ammonium sulfate is the black bar, PEG-8000 is the white bar, and isopropanol is the striped bar. The horizontal line is unity, and is shown to guide the eye.

from the most soluble (aspartic acid) to the least soluble (tryptophan) protein variant. There are differences between the three scales, but many of the observed trends are the same. In all three scales the proteins with aspartic acid and glutamic acid at the variable site are the most soluble. This may be partially explained by the increase in negative net charge observed for the acidic residues at pH 4.25 (see Table 4); however, the high solubility of the proteins with negatively charged residues relative to proteins with positively charged and neutrally charged residues at the variable site is also seen under conditions of positive net charge in ammonium sulfate, and the relative solubility of the variants with positively charged residues decreases at higher net charge.[37] The role of charge in protein solubility will be discussed further in Chapter V.

In all three scales the proteins with aromatic residues and cysteine show a significant decrease in solubility relative to the wild type protein with threonine at positions 76. This decrease in solubility can be observed qualitatively in buffer alone; solutions of WT RNase Sa can be concentrated to over 100 mg/ml, but the T76W variant cannot be concentrated to even 10 mg/ml at pH 4.25 (data not shown). In ammonium sulfate and isopropanol, the variants with methionine, isoleucine, valine, and leucine exhibit a moderate decrease in solubility; in PEG-8000 there is little change in solubility relative to threonine for these variants. In general, the proteins with histidine, asparagine, proline, glutamine and alanine at the variable site show similar solubility to wild type (threonine). In addition to the acidic residues mentioned earlier, the proteins with serine, lysine and glycine at position 76 exhibit moderate to significant increases in solubility in all three scales. The solubility of the arginine variant is variable and shows

Table 4. Solubility and charge properties of RNase Sa variants.

Variant	Solubility (mg/ml)			pI_{calc}[c]	Z_{net} (pH 4.25)[c]
	1.1 M $(NH_4)_2SO_4$[a]	10% Isopropanol (v/v)[a,b]	5% PEG 8000 (w/v)[a]		
T76D	43	18	22	3.8	-1.8
T76S	39	14	15	3.9	-1.0
WT	20	11	13	3.9	-1.0
T76R	42	9	15	4.2	-0.1

[a]Solubility values were measured in 50 mM sodium acetate at pH 4.25 and 25 ˚C. Error is +/- 5%.
[b]Solubility values were measured at 4˚C.
[c]pI and net charge were calculated using published pK values for RNase Sa.[96]

a significant increase in solubility on the ammonium sulfate scale and a slight increase and decrease in solubility on the PEG-8000 and isopropanol scales respectively.

Strategy for increasing protein solubility

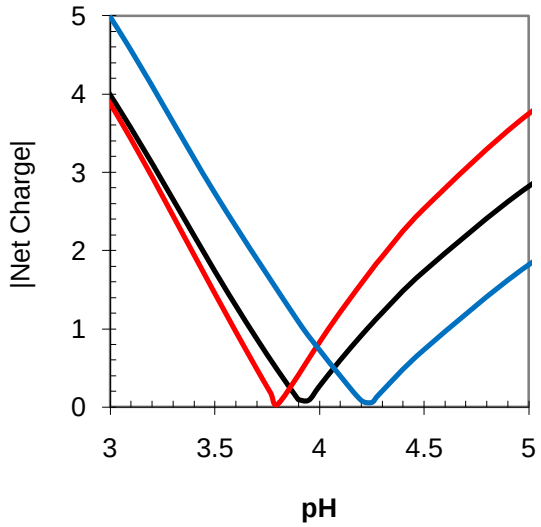
In using these data to rationally design mutations to increase protein solubility, one might assume that an effective strategy would be to mutate hydrophobic amino acids that decrease protein solubility to hydrophilic amino acids that increase protein solubility.[71-81] The first problem with this strategy is that if the structure of the protein is not known, the chance of mutating a hydrophobic residue that is on the surface of the protein is small;[82] therefore, it is more likely that a buried hydrophobic residue will be selected and will result in a decrease in conformational stability and no change to protein solubility. The second problem with this strategy is that, as seen in the three solubility scales, not all hydrophobic amino acids significantly decrease protein solubility, and not all hydrophilic amino acids significantly increase solubility. For example, methionine, alanine, and proline are hydrophobic amino acids, but they do not have a significantly unfavorable effect on protein solubility. Similarly, asparagine, glutamine, and threonine are polar amino acids, but they do not have a significantly favorable effect on protein solubility. In contrast, aspartic acid, glutamic acid and serine contribute more favorably to protein solubility than other hydrophilic amino acids. Therefore, we suggest the following strategy: if the PDB structure is available, use molecular modeling software, such as Naccess,[87] to identify if solvent accessible hydrophobic amino acids are present, and make mutations to aspartic acid or glutamic acid if you don't mind changing the net charge or to serine if you want to conserve charge. It is also ideal when selecting target

residues to avoid making mutations that may disrupt conformational stability, for example residues which participate in side-chain hydrogen bonds. If the PDB structure is not known or if hydrophobic surface residues are not available, we suggest mutating asparagine, glutamine, or threonine residues to aspartic acid, glutamic acid, or serine. This strategy will likely be more successful than mutating traditionally targeted hydrophobic residues which are on average 14% exposed to solvent in folded proteins as compared to asparagine, glutamine, and threonine which are on average 39% exposed to solvent.[82] It may also be helpful to use β-turn prediction programs to select asparagine, glutamine, and threonine residues that are in β-turns as these residues are often highly exposed to solvent.[19,97,98]

Charge and protein solubility

The data show that charged amino acids can play an important role in determining the solubility of a protein. In general, the solubility of a protein is lowest at a pH near the isoelectric point (pI) and increases as the pH becomes more basic or acidic and the charge on the protein becomes more negative or positive respectively.[16,44] The solubility would be expected to follow the absolute value of the net charge as shown in Figure 7(a). Table 4 shows the solubility, isoelectric point, and calculated net charge at pH 4.25 of WT RNase Sa along with an acidic and basic variant. The serine mutant is also included for comparison because the solubility increases for this mutant in all three precipitants. If only net charge is considered, at pH 4.25 one would expect that the solubility of these proteins would vary by the net charge show in Table 4 with the

(a)



(b)

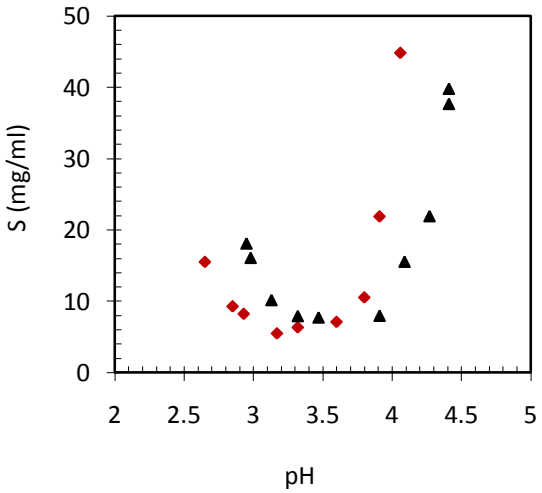


Figure 7. (a) The absolute value of the calculated net charge or RNase Sa variants T76D (red), WT (black), and T76R (blue). Curves were calculated using the Henderson-Hasselbalch equation and published pKa values.[96] (b) Measured solubility as a function of pH for WT (black) and T76D (red) RNase Sa.

greater net charge having the highest solubility. In other words, the solubility of the aspartic acid variant would be the highest (net charge is -1.8), the solubility of WT and charge neutral variants including serine would be intermediate (net charge is -1.0), and the solubility of the arginine variant would be the lowest (net charge is -0.1). However, the solubility behavior that we observe cannot be fully explained by net charge. In the ammonium sulfate scale, the solubility values of the aspartic acid, serine, and arginine variants are approximately the same as each other and significantly higher than WT, even though the net charge values are different. In the PEG-8000 scale, the aspartic acid variant is the most soluble, and the solubility of the serine and arginine variants increase somewhat with respect to WT. The isopropanol scale is the only scale that follows the behavior one would expect based on net charge. In this scale the aspartic acid variant is the most soluble, and the serine variant and WT have an intermediate solubility (though the serine variant is still more soluble than WT), and the arginine variant is the least soluble.

Clearly, the difference in solubility between the acidic and basic variants cannot be explained by net charge alone. To better explore the role of charged amino acids in determining protein solubility, it would be helpful to be able to compare proteins not at the same pH but at the same net charge. Because the calculated net charge can vary significantly from the effective net charge,[48] it is difficult to accurately predict the pH values necessary to achieve the same net charge for acidic, basic, and neutral mutations. Tom Laue's group has designed an experimental approach to measure effective charge

on a protein in solution,[47] and efforts to measure the charge of these variants are underway.

It would be nearly as informative if we could measure the solubility of different variants as a function of pH in the different precipitants, and a significant amount of effort has been put towards this, though with limited success. Experimental conditions requiring molar quantities of ammonium sulfate are not ideal for studying solubility at a wide range of pH values. In short, at low pH the kosmotropic sulfate ion becomes the chaotropic hydrogen sulfate ion and no longer acts as a precipitant and may even bind the positive charges on the surface of the protein.[64,99-101] At high pH, the ammonium ion (pKa=9.3) becomes ammonia and significant quantities of salt would be introduced to achieve pH values greater than approximately pH 8 due to the addition of base. This effectively limits the pH range for using ammonium sulfate from about pH 4 to pH 8. PEG-8000 is more amendable to this type of study, and an example of a solubility versus pH curve is shown in Figure 7(b) for WT RNase Sa and T76D. The results approximately follow the net charge as shown in Figure 7(a). Both curves display a minimum in the range of their respective pIs, and increase as the pH becomes more acidic or basic. WT RNase Sa has a minimum solubility in the range of pH 3.3 to pH 3.9, and T76D has a minimum solubility in the range of pH 3.0 to pH 3.6. Interestingly, both curves are nearly identical but are offset by approximately 0.3 pH units. The aspartic acid variant is more soluble than WT above pH 3.8. WT is more soluble than the aspartic acid variant below pH 3.3; however, the carboxyl group is likely predominantly protonated and uncharged in this pH range. To probe the role of charged

acidic and basic residues under conditions of positive and negative net charge, it would be necessary to perform experiments on a protein that has an isoelectric point in the neutral pH range far enough away from the pKas of the acidic and basic residues. The net charge and possibly the solubility of a protein is unlikely to change significantly in this region due to only histidine residues having a native pKa in this range.

Summary

Of the three solubility scales, the ammonium sulfate scale shows a more significant difference between proteins, varying over a nearly 11-fold range. The difference between proteins in the isopropanol and PEG-8000 scales is a 6-fold and a 4-fold range respectively. While the distribution of amino acids is continuous over the range of the ammonium sulfate and isopropanol scales, 12 out of 20 amino acids occupy a narrow range in the middle of the PEG-8000 scale. WT protein (threonine at position 76) was towards the center of all three scales both in terms of rank order and average solubility. When examined together, the amino acids can roughly be divided into three groups based upon their contribution to protein solubility, those that significantly increased solubility (aspartic acid, glutamic acid, serine lysine, and glycine), those that result in a solubility similar to threonine (histidine, asparagine, proline, glutamine, and alanine), and those that decreased solubility (isoleucine, valine, leucine, cysteine, tyrosine, phenyalanine, and tryptophan). Interestingly, arginine had a variable contribution to solubility ranging from a significant increase in solubility in the ammonim sulfate to a moderate decrease in the isopropanol scale. Of the 20 amino acids, the negatively charged aspartic acid and glutamic acid residues showed the

greatest increase in solubility in each of the three scales, and showed a greater increase in solubility than the basic residues. This may be somewhat due to the solubility measurements being made at negative net charge, but, at least for ammonium sulfate, this also pertains to conditions of positive net charge.[37] A likely explanation for this is that the carboxylates of aspartic acid and glutamic acid are kosmotropic, bind water tighter than water binds itself, and are significantly hydrated in solution, whereas the amino and guanidino groups of lysine and arginine are chaotropic, bind water weaker than water binds itself, and are mostly unhydrated in solution.[64,99-101] These data have led us to propose the following strategy for increasing protein solubility: mutate exposed hydrophobics and asparagine, glutamine and threonine to aspartic acid or glutamic acid if charge conservation is not crucial or to serine if it is.

CHAPTER IV

INCREASING PROTEIN SOLUBILITY BY REPLACING POLAR SURFACE RESIDUES WITH ASPARTIC ACID AND SERINE

Our strategy for increasing protein solubility is to substitute exposed hydrophobic residues, if available, and polar amino acids that contribute unfavorably to protein solubility, including asparagine, glutamine, and threonine residues, to aspartic acid and glutamic acid, if you don't mind change the net charge on the protein, or to serine if you need to keep the net charge the same.[19,37] Here, we test this strategy on a low solubility variant of RNase Sa, T76W. The solubility of WT RNase Sa has a solubility over 300 mg/ml, but this variant of RNase Sa has a solubility that is less than 10 mg/ml at pH 4.25 (unpublished data). Developing this strategy is important because there is currently not a clearly defined or effective strategy for making mutations that increase protein solubility.[4,19,20]

Selecting candidate positions and designing mutations to increase the solubility of T76W RNase Sa

To select the candidate positions to test our strategy for increasing protein solubility on T76W RNase Sa, we made use of *pfis* (PDF file information software),[86] a molecular modeling program that calculates the assessable surface area of a protein from a PDB file and provides hydrogen bonding information. Table 5 summarizes these data for amino acids on the surface of RNase Sa that contribute unfavorably to protein solubility as defined by Trevino *et al.* and chapter III of this dissertation. Six residues were found that meet these criteria and that had side chains that were less than 40% buried. The 40% side chain burial was an arbitrary cut off used for RNase Sa, and this parameter can be adjusted for other proteins as necessary to generate a sufficient number of candidate positions. In general, the more exposed the residue, the more likely that this strategy will be successful. Conversely, the greater the burial of the residue the less effective this strategy may be as it is primarily interactions of the surface residues that influence the intrinsic factors that effect protein solubility.[20] The six residues identified include asparagine 20, threonine 46, tyrosine 49, glutamine 77, threonine 88 and threonine 95. Interestingly, of the six residues identified, only tyrosine is hydrophobic, and the other five are polar. We will avoid residues whose substitutions are likely to cause a decrease in stability, specifically residues which form side-chain hydrogen bonds. Two out of the six residues form side-chain hydrogen bonds including threonine 88 and threonine 95, so these residues will not be targeted. Also, tyrosine is in the i +1

Table 5. RNase Sa residues that have surface exposed side chains that contribute unfavorably to solubility.[c]

Res No.	Amino Acid	Sidechain % Burial[a]	Backbone H-bonds[a]	Side Chain H-bonds[a]	Secondary Structure[b]
20	Asn	30.5	3	0	H
46	Thr	19.2	0	0	L
49	Tyr	33.5	0	0	L
77	Gln	39.4	1	0	L
88	Thr	39.5	1	1	L
95	Thr	28.5	1	1	L

[a]calculated using pfis and pdb file 1RGG.

[b]Secondary structure is denoted by L (loop), H (α-helix), S (β-sheet) and was determined using the Swiss-pdb viewer.

[c]a 40% side chain burial cut-off was used. Prolines and position 76 are excluded from the table.

position of a β-turn and conveys a significant level of stability to the protein,[98,102] so it will not be targeted. This leaves three positions which will be targeted for substitution to both serine and aspartic acid. The serine substitutions will be introduced into the T76W background of RNase Sa (pI = 3.5),[46]and the aspartic acid substitutions will be introduced into a basic variant of RNase Sa (3K, pI = 6.4)[46] with the T76W substitution in order to facilitate the measurement of solubility at both positive and negative net charge.

Figure 8 shows the candidate positions (black), negatively charged surface residues (red), positively charged surface residues (blue), and position 76 (yellow). The electrostatic features of the surface of the protein around the three candidate positions differ from each other. Threonine 46 is in a region on the surface of the protein that is relatively lacking of charged residues (Figure 8(a)). It is also a significant distance from the site of the T76W mutation. Asparagine 20 is in a region of the protein that has a number of acid residues on the surface of the protein (Figure 8(c)), and it is likely that this region is more heavily hydrated than the region around threonine 46, since carboxylates are kosmotropic and bind water tighter than water binds itself.[64] The region around glutamine 77 (Figure 8(b)) contains both acidic and basic residues, and it is also directly adjacent to the T76W mutation site (both in sequence and in space). The different environments of these three positions may yield insight into the context dependence of making surface mutations to increase protein solubility.

(a) (b) (c)

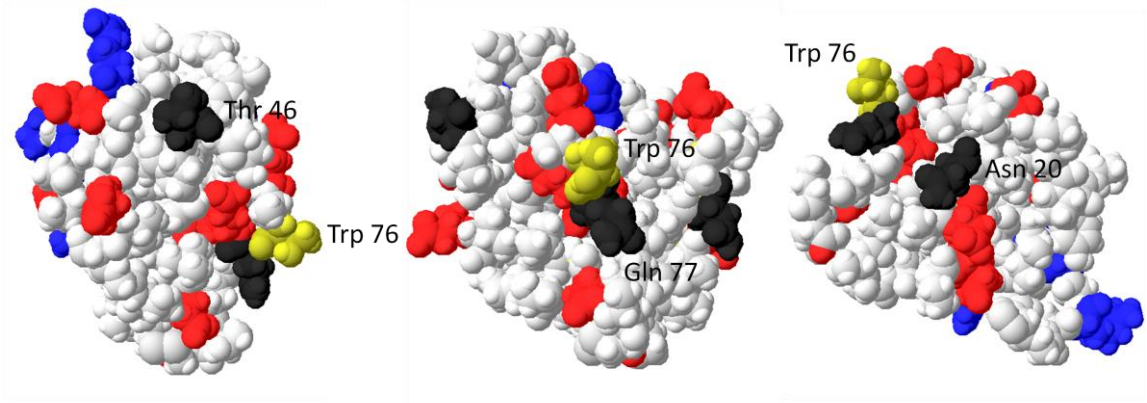


Figure 8. Three views of RNase Sa showing candidate positions for substitution. The candidate positions, threonine 46 (a), glutamine 77 (b) and asparagine 20 (c), are shown in black. Acidic residues are shown in red, basic residues are shown in blue, and the site of the T76W mutation is shown in yellow. These figures were created using the Swiss PDB viewer.[103]

Table 6. Melting temperatures for the serine and aspartic acid variants of T76W RNase Sa in 30 mM MOPS at pH 7.0.

Protein	T_m [a]	ΔT_m [b]
Wild type	47.4	0.0
T76W	49.8	2.4
N20S T76W	49.1	1.7
T46S T76W	49.6	2.2
Q77S T76W	47.2	-0.2
T76W (3K)	50.4	3.0
N20D T76W (3K)	50.6	3.2
T46D T76W (3K)	50.6	3.2
Q77D T76W (3K)	50.6	3.2

[a] Tm=T (in °C), where $\Delta G°=0$. The error is ±0.1 °C.
[b] ΔTm=Tm(mutant)−Tm(wild type) (in °C).

Stability of T76W RNase Sa variants

To determine if any of the amino acid substitutions significantly destabilizes RNase Sa, thermal unfolding experiments were performed, and unfolding was followed by circular dichroism. Table 6 shows the change in melting temperature of the various serine and aspartic acid T76W variants of RNase Sa. Of the eight variants used in this study, seven show an increase in melting temperature relative to WT RNase Sa of 1.7 ˚C to 3.2 ˚C. One variant, Q77S T76W, shows a slight decrease in melting temperature of -0.2 ˚C, but this is not significant enough to cause the protein to be unfolded at room temperature. These results indicate that the variants used in this study will remain folded under experimental conditions.

Solubility measurements of the serine variants of T76W RNase Sa

Table 7 lists the solubility values measured for the T76W serine variants of RNase Sa in ammonium sulfate, PEG-8000, and isopropanol at pH 4.25 (net charge is approximately -1). Figure 9 shows the relative increase in solubility (normalized to T76W) for these variants in ammonium sulfate (black bars), PEG-8000 (white bars), and isopropanol (grey bars). In ammonium sulfate, the solubility of T76W is 2.3 mg/ml. Upon introduction of either the T46S or the Q77 S mutation, the solubility increase to 4.3 mg/ml and 3.8 mg/ml respectively; this is a relative increase in solubility of over 1.8 and 1.6 fold respectively. In contrast, the N20S mutation does not significantly affect the solubility of T76W RNase Sa as the solubility increased by only 3% (within experimental error). The results in PEG-8000 and isopropanol, however, do not show any significant increase in solubility for these mutations. The solubility of T76W in

Table 7. Solubility measurements of RNase Sa T76W serine variants in 50 mM sodium acetate pH 4.25.

Protein	Solubility (mg/ml)		
	Ammonium sulfate[a]	PEG-8000[b]	Isopropanol[c]
T76W	2.3	6.3	5.4
T76W N20S	2.4	6.6	5.7
T76W T46S	4.3	6.9	4.8
T76W Q77S	3.8	4.6	4.0

[a] 1.1 M ammonium sulfate, 25 ˚C.

[b] 5% (w/v) PEG-8000 25 ˚C.

[c] 10% (v/v) Isopropanol, 4 ˚C.

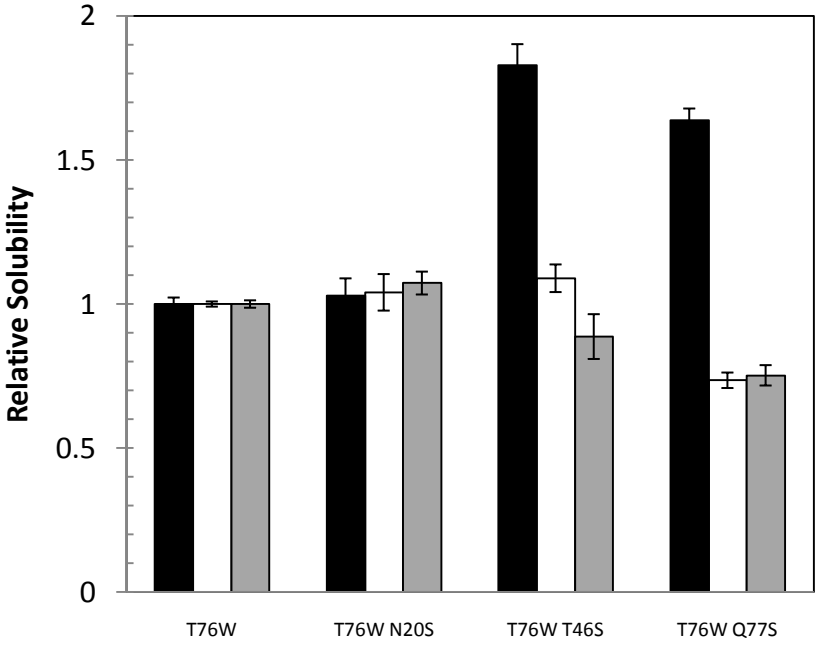


Figure 9. Relative solubility of T76W RNase Sa and serine variants at pH 4.25. The solubility values were normalized to the solubility of T76W. The black bars represent the ammonium sulfate solubility data, the white bars the PEG-8000 data and the grey bars are the isopropanol data.

PEG-8000 is 6.3 mg/ml, and the solubility upon introduction of the N20S or the T46S mutation only increases the solubility to 6.6 mg/ml and 6.9 mg/ml respectively. The solubility of the N20S mutation is within the error of the experiment, and the solubility of the T46S variant is just outside of the experimental error. Surprisingly, the Q77S mutation actually decreased the protein solubility to 4.6 mg/ml. In isopropanol, the results are similar to the results in PEG-8000. The solubility of T76W in isopropanol is 5.4 mg/ml, and the N20S mutation increases the solubility, though only slightly to 5.7 mg/ml (just outside of the experimental error). The solubility of the T46S and Q77S mutations decrease the solubility of T76W slightly to 4.8 and 4.0 mg/ml respectively.

Solubility measurements of the aspartic acid variants of T76W RNase Sa (3K)

To be able to determine if changes in solubility are due to changes in net charge alone, aspartic acid mutations were made within a T76W RNase Sa (3K) background (pI = 6.4)[46] This will allow us to make solubility measurements at net positive and net negative charge at pH values where the introduced aspartic acid residues are unprotonated and charged. Table 8 lists the solubility of the T76W (3K) aspartic acid variants of RNase Sa in ammonium sulfate, PEG-8000, and isopropanol at pH 5 and pH 8 as well as the calculated net charge values. Figure 10 shows the relative increase in solubility (normalized to T76W (3K)) at pH 8) at net negative charge for these variants in ammonium sulfate (black bars), PEG-8000 (white bars), and isopropanol (grey bars). The solubility of T76W (3K) in ammonium sulfate is 3.1 mg/ml. Upon introduction of one of three aspartic acid substitutions, the solubility increases significantly, though to varying degrees. The most significant increases in solubility were seen for the T46D and

Table 8. Solubility measurements and net charge calculations for RNase Sa T76W (3K) apartic acid variants.

Protein	Solubility (mg/ml)						Z_{net}[f]	
	Ammonium sulfate[a]		PEG-8000[b]		Isopropanol[c]			
	pH 5[d]	pH 8[e]	pH 5[d]	pH 8[e]	pH 5[d]	pH 8[e]	pH 5	pH 8
T76W (3K)	2.5	3.1	7.6	7.5	7.5	10	3.0	-0.4
T76W N20D (3K)	3.6	4.9	10	9.0	9.3	12	2.0	-1.4
T76W T46D (3K)	5.2	6.9	13	12	10	15	2.0	-1.4
T76W Q77D (3K)	4.3	6.4	13	11	11	14	2.0	-1.4

[a]1.1 M ammonium sulfate, 25 ˚C.

[b]5% (w/v) PEG-8000 25 ˚C.

[c]10% (v/v) Isopropanol, 4 ˚C.

[d]50 mM citrate buffer.

[e]50 mM tricine buffer.

[f] calculated using pK values from Laurents *et al.*[96] measured in the WT RNase Sa background assuming that the noted aspartic acid residues have an unperturbed pKa of 3.67 as reported by Thurkill *et al*[104].

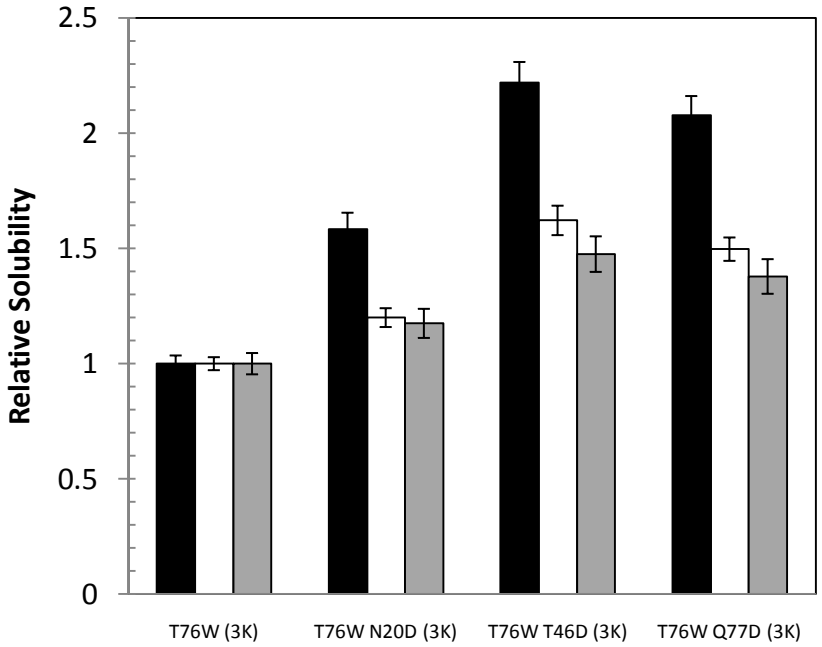


Figure 10. Relative solubility of T76W RNase Sa (3K) aspartic acid variants at negative net charge. The solubility values were normalized to the solubility of T76W (3K). The black bars represent the ammonium sulfate solubility data, the white bars are the PEG-8000 data and the grey bars are the isopropanol data.

Q77D substitutions with solubilities of 6.9 mg/ml and 6.4 mg/ml respectively. This corresponds to a greater than doubling of the solubility with a relative increase in solubility of 2.2 and 2.1 respectively. N20S also increased the solubility, though to a lesser extent than the other variants, to 4.9 mg/ml and had a relative solubility of 1.6. Unlike the serine results, the increases in solubility are also seen in PEG-8000 and isopropanol. In PEG-8000 the solubility of T76W (3K) at pH 8 is 7.5 mg/ml. This increases to 9.0 mg/ml, 12 mg/ml, and 11 mg/ml for N20D, T46D, and Q77D, corresponding to relative solubilities of 1.2, 1.6, and 1.5. In isopropanol the solubility of T76W (3K) is 10 mg/ml and increases to 12 mg/ml, 15 mg/ml, and 14 mg/ml for the N20D, T46D and Q77D respectively. This corresponds to relative increases in solubility of 1.2 for N20D, 1.5 for T46D, and 1.4 for Q77D.

Introduction of either the N20D, T46D, or the Q77D mutation at net negative charge increases the calculated net charge on the protein by approximately one charge unit (from -0.4 to -1.4 at pH 8, see Table 8 right side). Because the solubility of a protein is lowest near the pI and increases as negative or positive net charge increases,[44] we might expect that these mutations are simply shifting the pH versus solubility curve to more acidic pH values. If this were the case, we would expect the solubility to be lower for these mutations at net positive charge. To determine if this is the case for these variants or if solubility also increases at positive net charge, we measured the solubility of these variants at pH 5 where the calculated net charge on T76W (3K) is 3.0 and the net charge for the aspartic acid variants decreases to 2.0. The measured solubility values are listed in Table 8 and the relative solubility values are plotted in

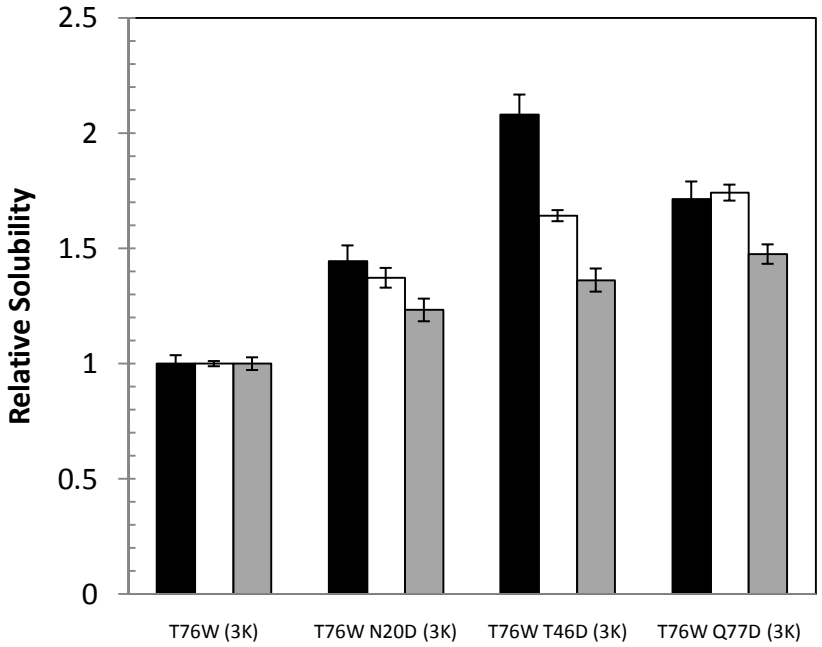


Figure 11. Relative solubility of T76W RNase Sa (3K) aspartic acid variants at positive net charge. The solubility values were normalized to the solubility of T76W (3K). The black bars represent the ammonium sulfate solubility data, the white bars are the PEG-8000 data and the grey bars are the isopropanol data.

Figure 11. The results obtained are very similar to those measured at net negative charge, suggesting that the increase in solubility seen due to the aspartic acid substitutions is not due to increases in net charge but rather is due to the new aspartic acid and is effective at both positive and negative net charge. In summary, the solubility of the three aspartic acid variants increased as at pH 5 in all three of the precipitants used. The solubility of T76W (3K) in ammonium sulfate is 2.5 mg/ml and the solubility of the aspartic acid variants increase by a factor of 1.4 to 2.1. In PEG-8000 the solubility of T76W (3K) is 7.6 mg/ml and the relative solubilities of the aspartic acid variants are between 1.4 and 1.7. In isopropanol, the solubility of T76W (3K) is 7.5 and the relative increases in solubility are between 1.2 and 1.5 fold for the aspartic acid variants. The increases in solubility seen at net positive charge are significant, but in general, are slightly less than those seen at net negative charge, suggesting that net charge may play a partial role in the change in solubilities of these variants.

The effectiveness of serine surface mutations in increasing the solubility of T76W RNase Sa

The solubility results from the serine variants of T76W are inconclusive. The solubility measurements in ammonium sulfate show a significant increase in solubility for the T46S and Q77S variants. In PEG-8000 and in isopropanol, the solubility is not increased significantly by any of the three serine mutations, and the solubility even decreases, though only slightly, for Q77S in these precipitants. These results seem to somewhat follow the behavior of the solubility scales presented in the previous chapter, where T76S has a significantly higher solubility in ammonium sulfate than it does in

either PEG-8000 or isopropanol (see chapter III), though T76S does still show an increase in solubility in these scales. It is likely that serine mutations were unsuccessful in increasing the solubility of T76W in PEG-8000 and isopropanol because these substitutions were too conservative to even partially overcome the significant decrease in solubility of the T76W substitution which lowers the solubility of wild-type RNase Sa by two orders of magnitude. Unfortunately, even though the solubilities of these variants are low, it was not possible to get accurate or reproducible measurements of solubility in the absence of precipitant. Though these results were not conclusive as to the effect of serine mutations on the solubility of T76W RNase Sa, there are examples in the literature where serine mutations did effect protein solubility.[105-108]

The effectiveness of aspartic acid surface mutations in increasing the solubility of T76W RNase Sa

The solubility results from the aspartic acid variants of T76W(3K) RNase Sa clearly show that this strategy was effective for increasing the solubility of this protein in all three precipitants studied. The ammonium sulfate results show the greatest increase in solubility, and two of the substitutions (T46D and Q77D) each more than double the solubility at net negative charge. The PEG-8000 and isopropanol results show more moderate increases in solubility of between 1.2 and 1.6 fold increases in solubility. The results at net positive charge are very similar to the results at net negative charge. This shows that changing the net charge is not the dominant mechanism by which aspartic acid substitutions increase protein solubility. As proposed in the preceding chapter, the

high propensity for hydration of the carboxylate side chains of aspartic acid and glutamic acid may explain their favorable contribution to protein solubility.[64]

At all three substitutions sites, and aspartic acid substitutions increase solubility; however, there does appear to be a context dependence for the substitution site of the aspartic acid residues. At both positive and negative net charge and in all three precipitants, the T46D and Q77D variants show about an equal effect on protein solubility. However, the N20D variant, while still showing an increase in solubility relative to T76W (3K), has a significantly lower solubility than the other two aspartic acid variants in all three precipitants. Interestingly, the N20S mutation is the only one of the serine variants that does not increase solubility in ammonium sulfate. This implies that context does play a role in determining the effect of surface mutations on protein solubility. Asparagine 20 is only 6.5 Å from the nearest charged residue at position 17 (position 17 is an aspartic acid in WT RNase Sa but is changed to a lysine in 3K) and several other charged residues are also in the vicinity (see Figure 8(c)). In contrast, threonine 46 is nearly 12 Å from the closest charged residue, and substitutions at this position are much more effective at increasing the solubility. Glutamine 77 is also on a highly charged region of the protein surface, but it is also immediately adjacent to the T76W mutation site. This suggests that aspartic acid mutations more effectively increase protein solubility if the residue being substituted is on a region of the protein surface lacking charge or next to a hydrophobic residue or patch.

Making hydrophobic to hydrophilic surface mutations to increase protein solubility compared with this approach

Another common approach for increasing protein solubility seen in the literature is the replacement of hydrophobic with hydrophilic residues.[71-81] We have discussed potential shortfalls that can be encountered with this approach in preceding chapters. If this strategy was used for increasing the protein solubility of T76W RNase Sa, tyrosine 49 would have been selected as the most exposed hydrophobic residue on the surface of the protein (see Table 5), and mutation of this residue would have resulted in a decrease in the stability of RNase Sa.[98] Even if we change our cutoff to a higher level of side chain burial, the next least buried residues are isoleucine 58 (51% buried) and tyrosine 30 (60 % buried). Substitution of these residues to hydrophilic residues would likely have resulted in a substantial decrease in protein stability due to the loss of buried hydrophobic surface area. If any increase in protein solubility were seen for these substitutions, it would likely be less significant than the increases in solubility seen in this study, because the introduced hydrophilic residues would be partially buried.

Summary

We have tested a previously proposed strategy for increasing the protein solubility of a low solubility variant of RNase Sa. Using this approach, we identified three polar surface residues as targets for substitution and we individually substituted these three positions with either serine or aspartic acid. We measured the solubility of these variants in three precipitants and found that the serine variants only showed an increased solubility in ammonium sulfate, but the aspartic acid variants increased the

solubility in all three precipitants. The increased solubility of the aspartic acid variants was seen regardless of whether the net charge on the protein was positive or negative. We believe that this approach will be much more effective at increasing protein solubility than simple hydrophobic to hydrophilic surface mutations.

CHAPTER V

TOWARDS A MOLECULAR UNDERSTANDING OF PROTEIN SOLUBILITY: INCREASED NEGATIVE SURFACE CHARGE CORRELATES WITH INCREASED SOLUBILITY

General understanding of the intrinsic properties of proteins that determine protein solubility is poor[4] and could be improved by a quantitative study examining a group of proteins with available structural information and measured solubility values. The purpose of this study is two-fold: to compare the solubility results obtained with two different classes of precipitants, salts (ammonium sulfate) and long-chain polymers (polyethylene glycol 8000) for a number of different proteins and to examine the molecular properties of the proteins used in this study with the hopes of better understanding the solubility results obtained in order to gain insight into the intrinsic factors that influence protein solubility.

Proteins are folded under experimental conditions

For this study, we are interested in examining the solubility of folded proteins. While the low solubility of the unfolded state stands as a challenge for those studying the denatured state ensemble, the solubility of the native state is the focus of this study because of its relevance to crystallographers, protein chemists, and those producing protein pharmaceuticals.[42] The precipitants used in this study are common

crystallographic precipitants, if the protein is folded in solution, the precipitate is expected to be native protein.[39,109] To confirm this, we looked at the effect of ammonium sulfate and PEG-8000 on stability by thermal unfolding experiments. Originally this study was designed to include isopropanol as a precipitant. Thomas and Dill[70] investigated the mechanism by which alcohols destabilize proteins and found that it was complex and dependent upon protein sequence and structure. They concluded that alcohols destabilize proteins mainly by weakening hydrophobic interactions. We found that the concentration of isopropanol required to achieve precipitation for many of the proteins in this study was great enough that a mixture of folded and unfolded protein would be present under experimental conditions (unpublished data), so the class of organic solvents was excluded from this study. Use of isopropanol or other denaturing organic solvents for studying protein solubility should be reserved for proteins that remain folded in the presence of solvent concentrations necessary to induce precipitation. The temperatures at which precipitations by organic solvents are performed can be lowered in order to diminish the denaturing effects.

Figures 12 and 13 show changes in observed melting temperature as a function of ammonium sulfate and PEG-8000, respectively. For ammonium sulfate, the melting temperature increases for all of the proteins studied. This is to be expected because sulfate ions increase the surface tension of bulk water, and the folded state is favored due to the reduced water-protein interface as compared to the unfolded state.[63,69]

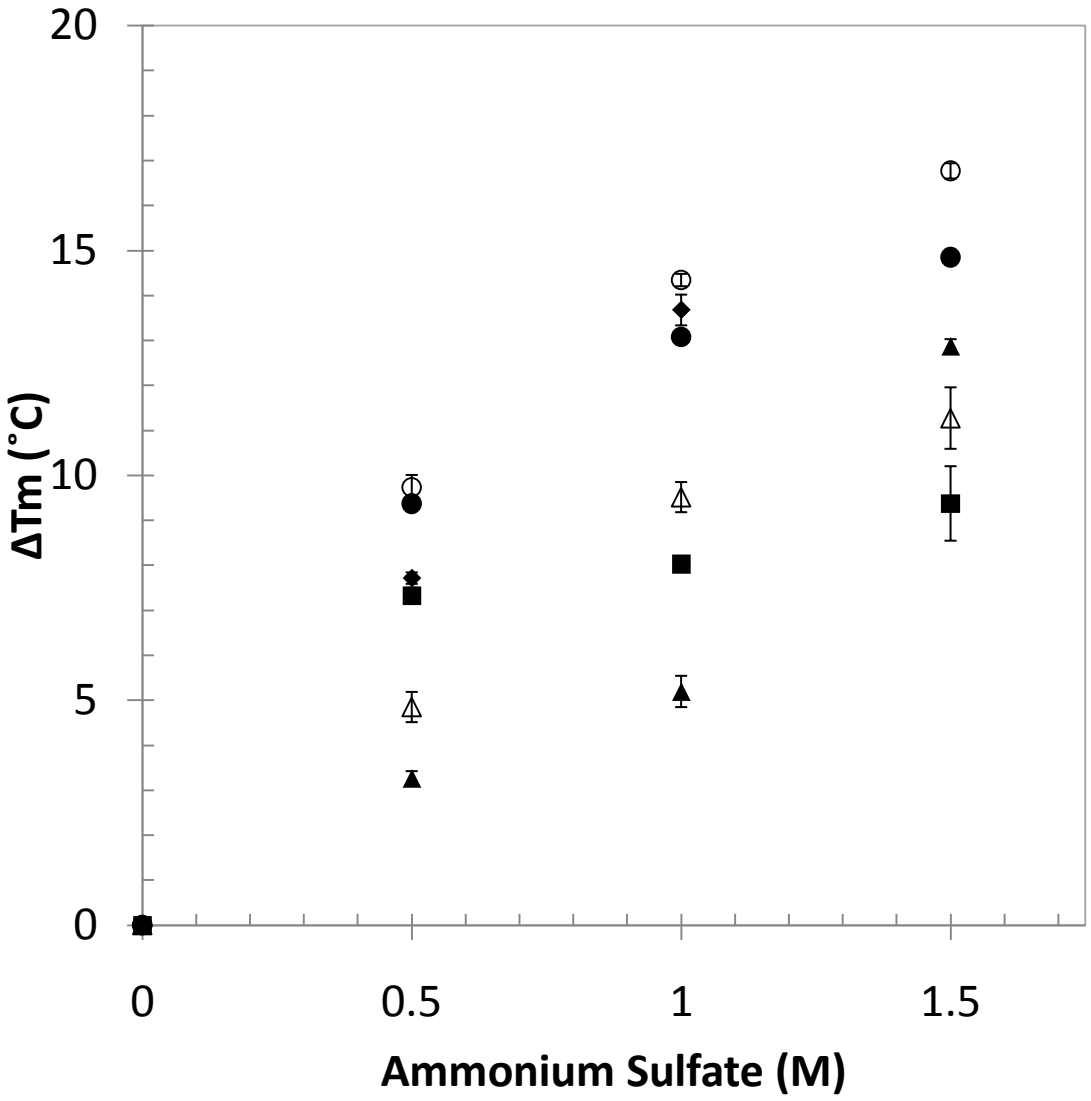


Figure 12. Change in melting temperature as a function of ammonium sulfate concentration for RNase Sa (filled diamonds), α-chymotrypsin (filled circles), lysozyme (open triangles), human serum albumin (filled triangles), ovalbumin (filled squares), and α-lactalbumin (open circles). Thermal denaturation was followed by circular dichroism; see materials and methods section for details.

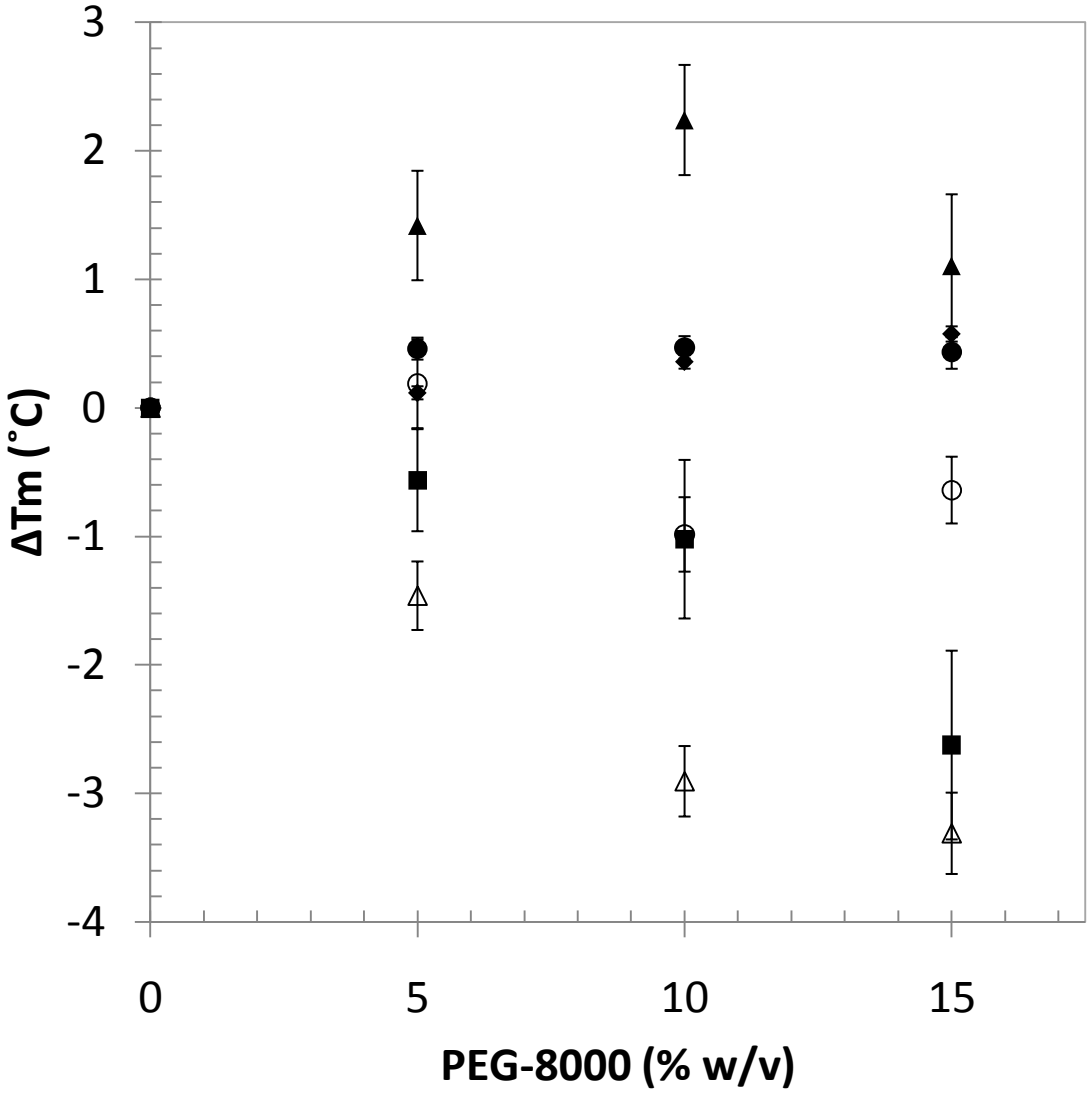


Figure 13. Change in melting temperature as a function of polyethylene glycol concentration for RNase Sa (filled diamonds), α-chymotrypsin (filled circles), lysozyme (open triangles), human serum albumin (filled triangles), ovalbumin (filled squares), and α-lactalbumin (open circles). Thermal denaturation was followed by circular dichroism; see materials and methods section for details.

For example, the degree of stabilization at 1M ammonium sulfate varies from an increase in T_m of 5.2°C for human serum albumin to an increase of 14.3°C observed for α-lactalbumin. One might propose that the amount of surface buried and the degree of unfolding may influence the degree of stabilization for an individual protein; however, in this situation that does not appear to be the case. The protein with the greatest degree of stabilization by ammonium sulfate is α-lactalbumin. It forms a molten globule upon unfolding,[110] and likely buries less surface area than proteins that more fully unfold. This differential stabilization by ammonium sulfate deserves further investigation than will be addressed in this study. For the purposes of this study, it is sufficient to say that all of the proteins used here are stabilized by ammonium sulfate and are folded under experimental conditions. Furthermore, it has been shown that upon precipitation by ammonium sulfate, protein in both the solution and solid phase remains folded.[109]

In contrast to ammonium sulfate, high molecule weight polyethylene glycols are not expected to interact with or have a significant effect on protein stability;[50,51] however, at high concentrations they may destabilize some proteins,[17] and PEG molecules with a molecular weight of 6000 daltons or less have been shown to destabilize some proteins.[68] Figure 13 shows that our data agree with these previous observations; the effect of PEG-8000 on stability is small relative to the effect of ammonium sulfate; though, in the case of lysozyme and ovalbumin we observeed a slight decrease in stability. For example, the effect of PEG-8000 on T_m for all proteins ranges from an increase in T_m of 2.2°C for human serum albumin to a decrease in T_m of 2.9°C for lysozyme at 10% (w/v) PEG-8000. For the two proteins which show a

decrease in stability, the decrease in stability is not enough to cause a significant change in the population of unfolded protein present at room temperature; the T_m's of lysozyme and ovalbumin in the absence of precipitant are 70.5°C and 70.8°C respectively (data not shown), and the maximum concentration of PEG-8000 used is less than 15 % (w/v) for these two proteins. As has been reported elsewhere, fibrinogen exhibits a complicated multi-state unfolding curve[111] that is not amenable to this type of analysis; however, the curves clearly show that fibrinogen is folded under the conditions used in this study (data not shown). In conclusion, the proteins used in this study are folded under experimental conditions in the presence of either ammonium sulfate or polyethylene glycol 8000.

Solubility measurements rapidly reach equilibrium

Since solubility values are defined at equilibrium, it is necessary to insure that samples are allowed adequate time to reach equilibrium. For all proteins, samples were prepared in duplicate and were either centrifuged and quantified immediately or allowed to sit for 24 hours before centrifugation. Measurements made immediately and at 24 hours agree with each other (data not shown), signifying that protein precipitations reach equilibrium quickly after precipitation by either PEG-8000 or ammonium sulfate. This was seen for all protein and precipitant combinations tested in this study and agree with previously reported results for amorphous protein precipitation.[18,37]

Solubility curves in polyethylene glycol

The solubilities of seven proteins were determined as a function of PEG-8000. Figure 14 shows the plot of log solubility as a function of PEG-8000 concentration.

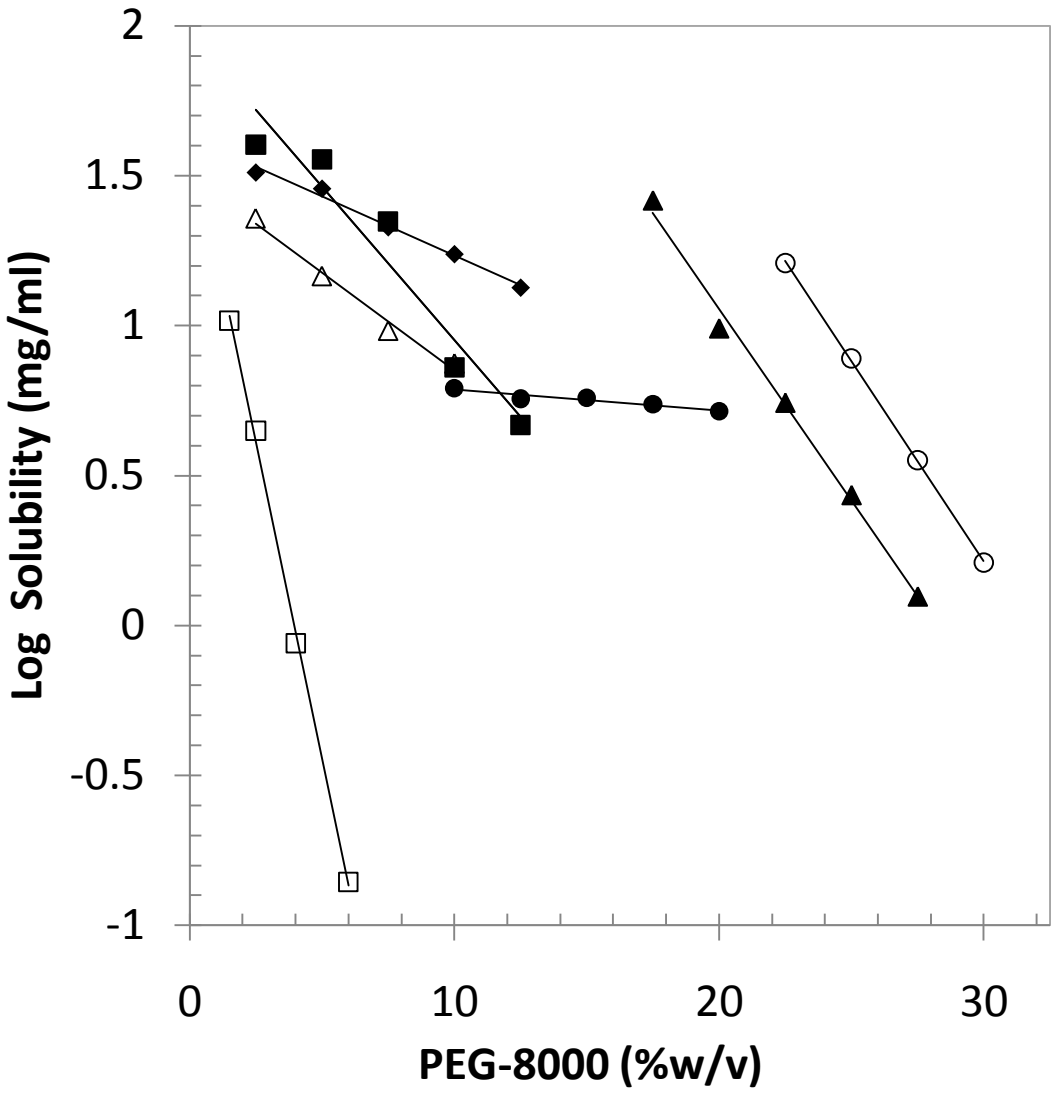


Figure 14. The solubility of several proteins in polyethylene glycol-8000. The solubility of RNase Sa (filled diamonds), α-chymotrypsin (filled circles), lysozyme (open triangles), human serum albumin (filled triangles), ovalbumin (filled squares), α-lactalbumin (open circles), and fibrinogen (open squares) were measured at room temperature (25˚C) and in pH 7.0 5mM citrate buffer. Equation 6 was fit to the data, and the fitted parameters can be found in Table 9.

Table 9. Fit values for PEG-8000 and ammonium sulfate solubility curves.

Protein	PEG-8000[a]			Ammonium sulfate[a]		
	Slope (β)	Intercept (log S_0)	R^2	Slope (β)	Intercept (log S_0*)	R^2
α-chymotrypsin	-0.01	0.9	0.92	-0.92	2.5	1.00
lysozyme	-0.07	1.5	0.99	-1.49	3.5	0.99
human serum albumin	-0.13	3.6	0.99	-2.68	7.5	0.97
RNase Sa	-0.04	1.6	0.99	-1.63	3.4	1.00
α-lactalbumin	-0.13	4.2	1.00	-0.17	1.6	0.98
fibrinogen	-0.42	1.7	1.00	-4.13	3.5	0.98
ovalbumin	-0.10	2.0	0.93	-2.46	6.0	0.97

[a]Equation 6 was fit to the data from Figures 14 and 15 respectively.

Equation 6 (see Chapter I) was fit to the data and linear fit parameters are given in Table 9 (left side). The dependence of solubility on PEG-8000 concentration (β) varies over a 40 fold range from -0.01 for α-chymotrypsin to -0.42 for fibrinogen. As has been reported,[57] PEG precipitation curves are linear over a wide range of PEG concentrations and can be extrapolated to zero PEG concentrations, yielding an estimate of solubility in the absence of precipitant from the y-intercept (log S_0) of the fit. The relative solubility of the proteins is determined by the log S_0 values which vary between 0.9 for fibrinogen and 4.2 for α-lactalbumin. The log S_0 values indicate that human serum albumin and α-lactalbumin are the most soluble of the proteins and this makes sense given that they are present at high concentrations in their respective biological environments (human plasma[112] and bovine milk[113] respectively).

Solubility curves in ammonium sulfate

The solubilities of the proteins were determined as a function of ammonium sulfate concentration. Figure 15 shows the plot of log solubility versus ammonium sulfate concentration. The data were similarly fit by equation 6, and the best fit values are given in Table 9 (right side). The linearity of a salting out curve does not extend to the y-axis due to salting-in at low concentrations of salt, so log S_0* will be used in place of log S_0 to signify that the y-intercept is projected from the salting-out region. The dependence of solubility on ammonium sulfate concentration (β) was found to be similar for all proteins except α-lactalbumin. For those 6 proteins β varies only about 4 fold from -0.92 for α-chymotrypsin to -4.13 for fibrinogen compared to β values for PEG-

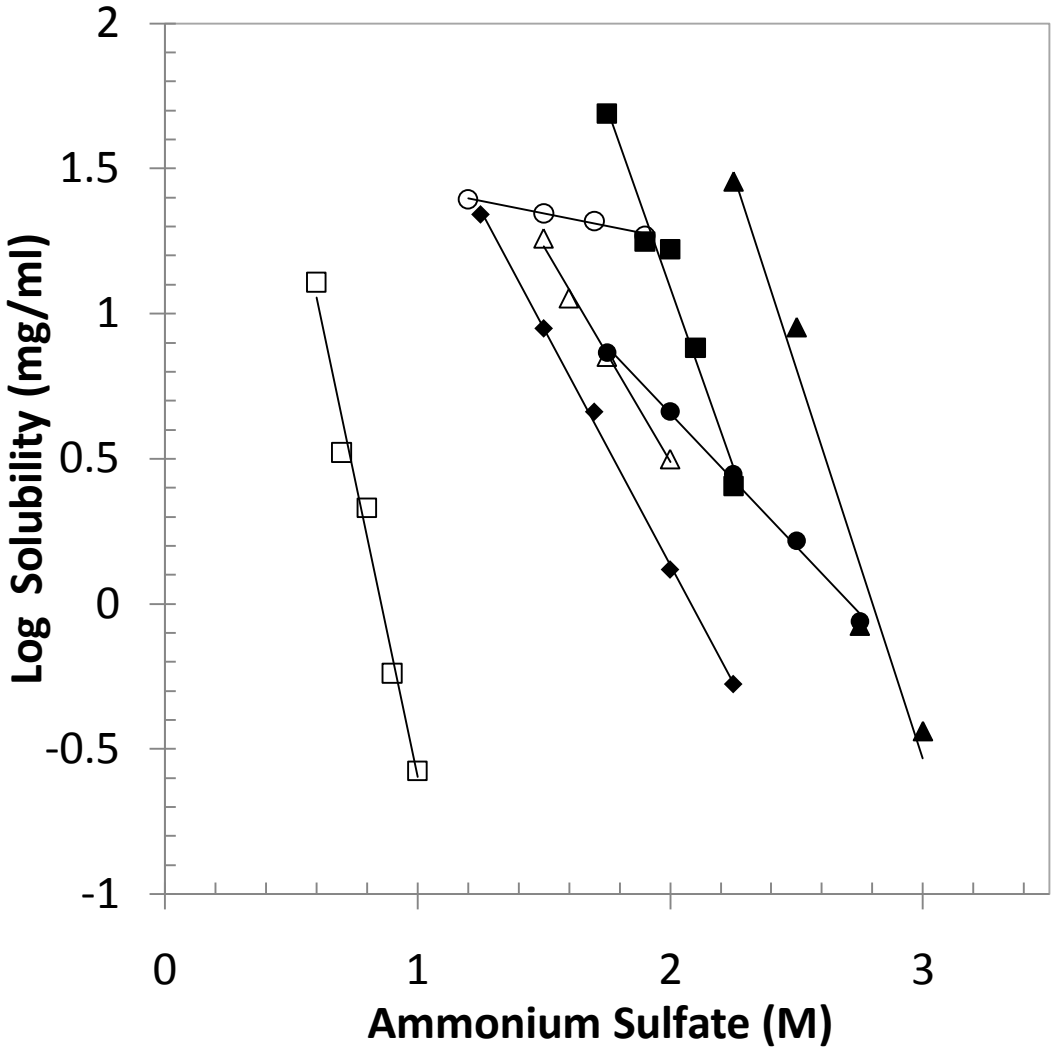


Figure 15. The solubility of several proteins in ammonium sulfate. The solubility of RNase Sa (filled diamonds), α-chymotrypsin (filled circles), lysozyme (open triangles), human serum albumin (filled triangles), ovalbumin (filled squares), α-lactalbumin (open circles), and fibrinogen (open squares) were measured at room temperature (25˚C) and in pH 7.0 5 mM citrate buffer. Equation 6 was fit to the data, and the fitted parameters can be found in Table 9.

concentration is more variable for PEG-8000 than for ammonium sulfate. α-lactalbumin is a clear outlier with a β value of -0.17 that is 13 fold lower than the average β value and 5 fold lower than the next closest β value of -0.92. This suggests that the ability of 8000 which vary over a 40 fold range. Clearly the dependence of solubility on precipitant ammonium sulfate to precipitate α-lactalbumin is reduced in comparison to the other 6 proteins. The case of α-lactalbumin will be addressed further in the Discussion. Based on the log S_0* values, human serum albumin is still predicted to have a high solubility, and α-chymotrypsin is still predicted to have a low solubility, as seen with the results in PEG-8000. Interestingly, α-lactalbumin, which was predicted to have the highest solubility by PEG-8000, is now predicted to have the lowest solubility.

Comparing protein solubility in PEG-8000 and ammonium sulfate

To compare the solubility measured with PEG-8000 to the solubility measured using ammonium sulfate, the log S_0 values from the two fits were evaluated. Figure 16 shows the plot of log S_0* obtained with ammonium sulfate (log S_0* $(NH_4)_2SO_4$) versus log S_0 obtained with PEG-8000 (log S_0 PEG). A remarkably strong correlation between the solubility results of ammonium sulfate and those of PEG-8000 is seen. This suggests that log S_0* $(NH_4)_2SO_4$ is a parameter related to protein solubility in the absence of precipitant. Since log S_0 PEG can be used to estimate solubility in the absence of buffer, the correlation of log S_0 PEG with log S_0* $(NH_4)_2SO_4$ suggests that log S_0 $(NH_4)_2SO_4$ can be used qualitatively to determine differences in solubility.

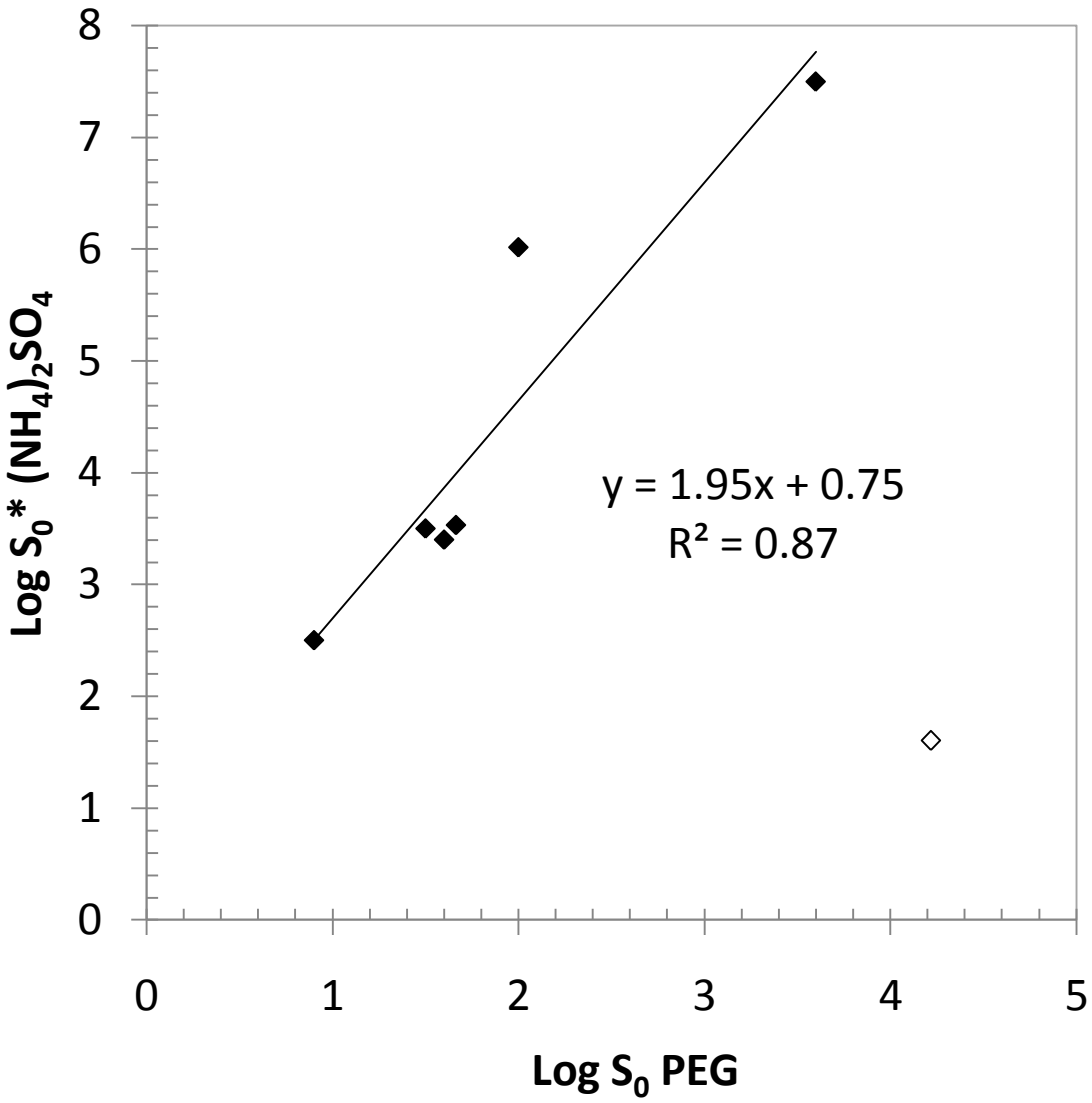


Figure 16. A comparison of the solubility data obtained in ammonium sulfate and PEG-8000. Log S_0 values obtained from PEG-8000 precipitations are plotted against Log S_0* values from ammonium sulfate precipitations for all of the proteins. The data correlate strongly, suggesting a relationship between solubility results in PEG-8000 and ammonium sulfate. α-lactalbumin is shown as an open diamond and is excluded from the fit.

The solubility of α-lactalbumin warrants further discussion. In the case of polyethylene glycol precipitation, α-lactalbumin is predicted to have the highest solubility of the proteins in this study. This is not surprising due both to the fact that α-lactalbumin is present in high concentrations in bovine milk[113] and our ability to make stock concentrations of α-lactalbumin that are in excess of 100 mg/ml. α-lactalbumin has a β value in PEG-8000 that is intermediate of the slopes observed for the other proteins. In the case of ammonium sulfate, α-lactalbumin is predicted to have the lowest solubility among the proteins studied, and the slope observed with α-lactalbumin is a clear outlier. It is the smallest slope observed, 13 fold lower than the average slope, and 5 fold lower than the next closest slope in ammonium sulfate. This suggests that ammonium sulfate is not as effective of a precipitant for α-lactalbumin as it is for the other proteins. This may in part be due to the high surface charge on α-lactalbumin; two thirds of the exposed surface residues carry a charge at pH 7 (data not shown) and nearly a third of the accessible surface area is charged (see Table 10, column 9). We have suggested before that ammonium sulfate may underestimate the contribution of charged surface residues to protein solubility.[37] This likely is related to the mechanism by which ammonium sulfate lowers protein solubility, ie raising surface tension and competing for waters with the protein surface. The high number of charged surface area on α-lactalbumin (roughly equal amounts positive and negative) likely affects the ability of ammonium sulfate to act as a precipitant. The kosmotropic carboxylates on the protein surface compete strongly for water molecules with the sulfate ions, and the chaotropic

Table 10. Protein properties and surface properties used for correlations.

Protein	Mw (kDa)	Amino acids	pI[a]	Charge[a]	Absolute charge[a]	Fraction of ASA[b]				
						Nonpolar	Polar	Charged	Positive	Negative
α-chymotrypsin	25.2	241	8	3.1	3.1	0.52	0.48	0.18	0.12	0.06
lysozyme	14.3	129	10	7.9	7.9	0.48	0.52	0.23	0.18	0.05
human serum albumin	66.5	585	6	-12.2	12.2	0.59	0.41	0.27	0.12	0.14
RNaseSa	10.5	96	3.5	-6.6	6.6	0.56	0.44	0.14	0.06	0.08
α-lacalbumin	14.2	123	5	-6.6	6.6	0.50	0.50	0.30	0.14	0.15
fibrinogen	160	1424	6.8	-3.7	3.7	0.58	0.42	0.19	0.11	0.09
ovalbumin	42.8	385	5.3	-10.5	10.5	0.52	0.48	0.21	0.09	0.12

a-pI and charge were calculated at pH 7 using protein calculator.[114]

b-Fraction nonpolar, polar, charged, positively charged, and negatively charged were calculated as fractions of total accessible surface area.

amino and guanadino groups may lower the water surface tension at the protein water interface, partially opposing the effect of ammonium sulfate. Due to the unique nature of the salting our curve of α-lactalbumin, the ammonium sulfate data will be fit without α-lactalbumin in the subsequent correlations.

In an attempt to determine the intrinsic factors to influence protein solubility, we looked at the several intrinsic protein properties and examined them with respect to protein solubility by comparing them to log S_0 values obtained in this study. We looked at fundamental properties of the protein such as size (molecular weight) and net charge. We also looked at properties of the surface of the protein including polarity and charge by determining the fraction of the surface area of the protein that is either polar, nonpolar, charged, negatively charged, or positively charged. By looking at the correlation of these properties with protein solubility, we were able to determine their relative importance for determining protein solubility.

Correlation of molecular weight and net charge with solubility measurements

To investigate the contribution of intrinsic factors to protein solubility (see Table 10 columns 1-6), log S_0 values for PEG-8000 and ammonium sulfate were plotted versus molecular weight (Figure 17 (a) and(b)), net charge (Figure 17 (c)and (d)), and the absolute value of the net charge (Figure 17 (e) and(f)). Linear fits were made to the data, and R^2 values are given. Due to the mechanism of polyethylene glycol precipitation being related to excluded volume; solubility may increase with protein size or molecular weight; however, no correlation with molecular weight was observed. In general, the solubility of a given protein is at a minimum near the isoelectric point and increases with

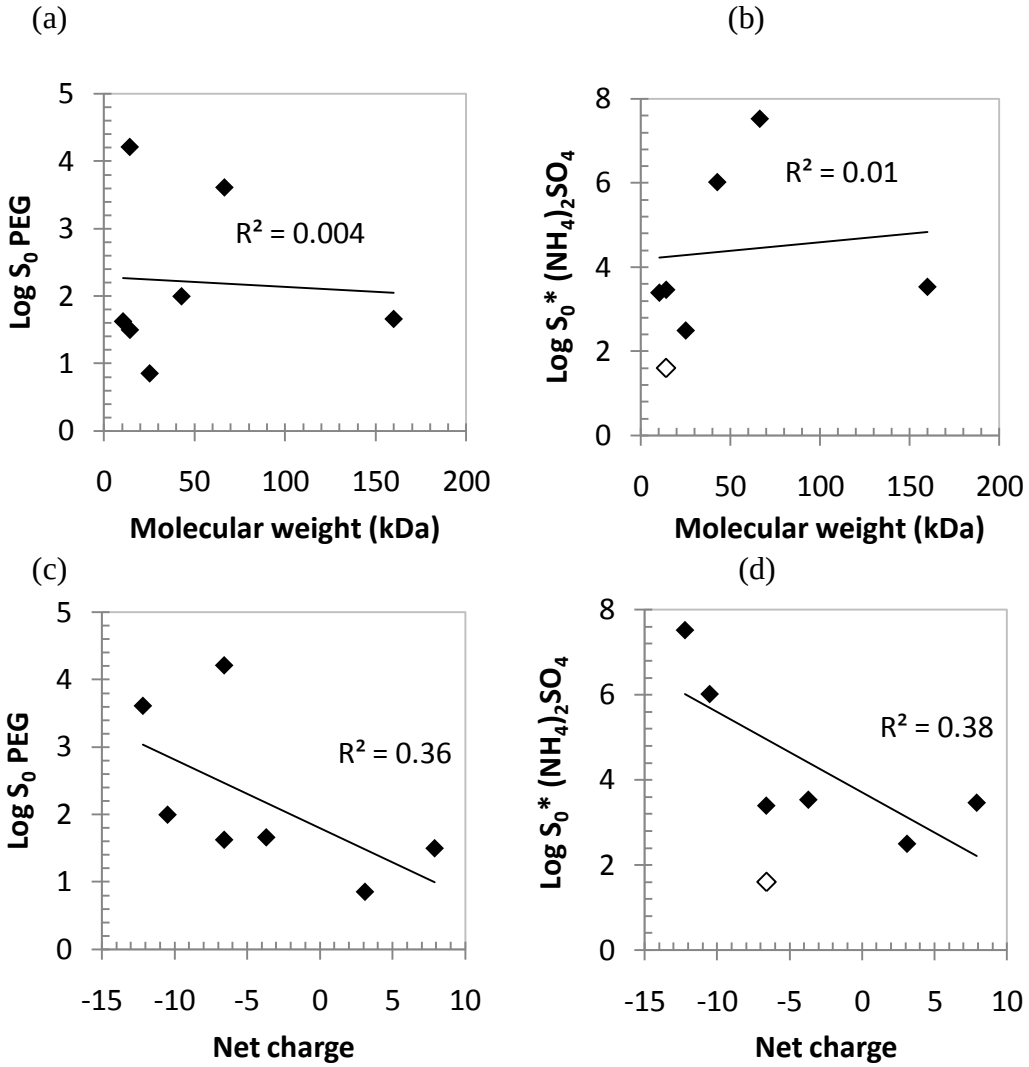


Figure 17. The correlation of molecular weight and net charge with PEG-8000 and ammonium sulfate solubility measurements. Log S_0 values versus molecular weight (a and b), net charge (c and d), and the absolute net charge (e and f) are shown. The lines and R^2 values are from linear least squares fits. α-lactalbumin is shown as open diamonds and is excluded from the ammonium sulfate fits.

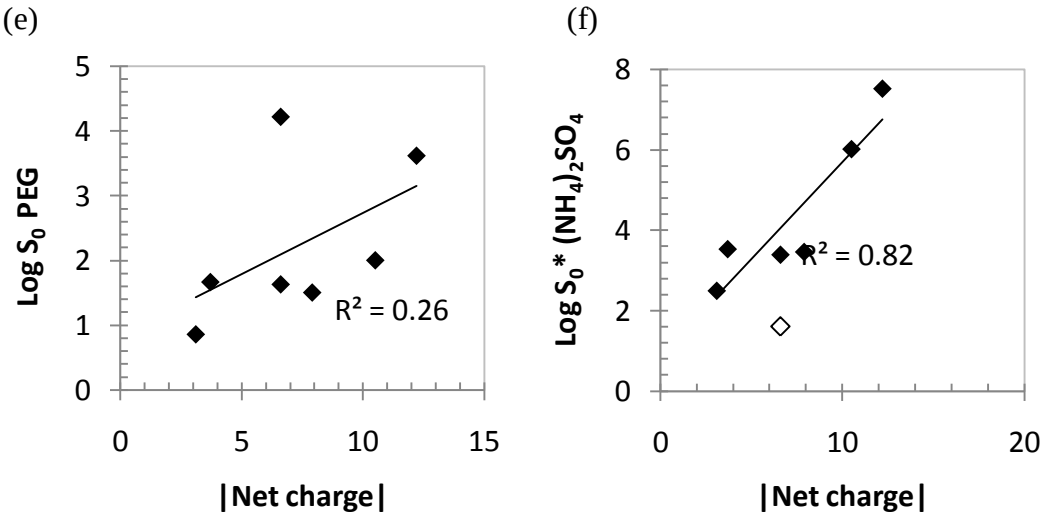


Figure 17. Continued.

the absolute value of the net charge.[16,44] In order to determine if net charge plays a role in determining the solubility of a group of proteins, the net charge and absolute value of net charge were plotted versus log S_0. A weak correlation was observed in all 4 cases. In the case the absolute value of net charge, a weak positive correlation was observed, suggesting that with increasing positive or negative net charge, protein solubility increases. For net charge versus pH, a weak negative correlation was observed. This suggests that, on average for this set of proteins, negatively charged proteins are more soluble than positively charged proteins; though, more data points are required to determine if this is true for a larger set of proteins.

Correlation of the intrinsic properties of the accessible surface area with protein solubility

Because protein solubility is influenced largely by interactions between water and the protein surface, the correlation of solubility with the intrinsic properties of the surface of the protein were investigated. The accessible surface areas (ASA) of all atoms in the proteins were determined and the fractions polar, nonpolar, charged, positively charged, and negatively charged were calculated (see Table 10 columns 7-11). Figure 18 depicts the protein solubility as a function of fraction ASA that is polar or nonpolar for PEG-8000 (Figure 18 (a)and (b)) and ammonium sulfate (Figure 18 (c) and (d)). For PEG-8000, the correlation of the percentage of polar and nonpolar surface residues with protein solubility is very poor, though the correlation is positive for polar residues and negative for nonpolar residues, as might be predicted. This suggests that the surface polarity has a minimal contribution to protein solubility in PEG-8000. For

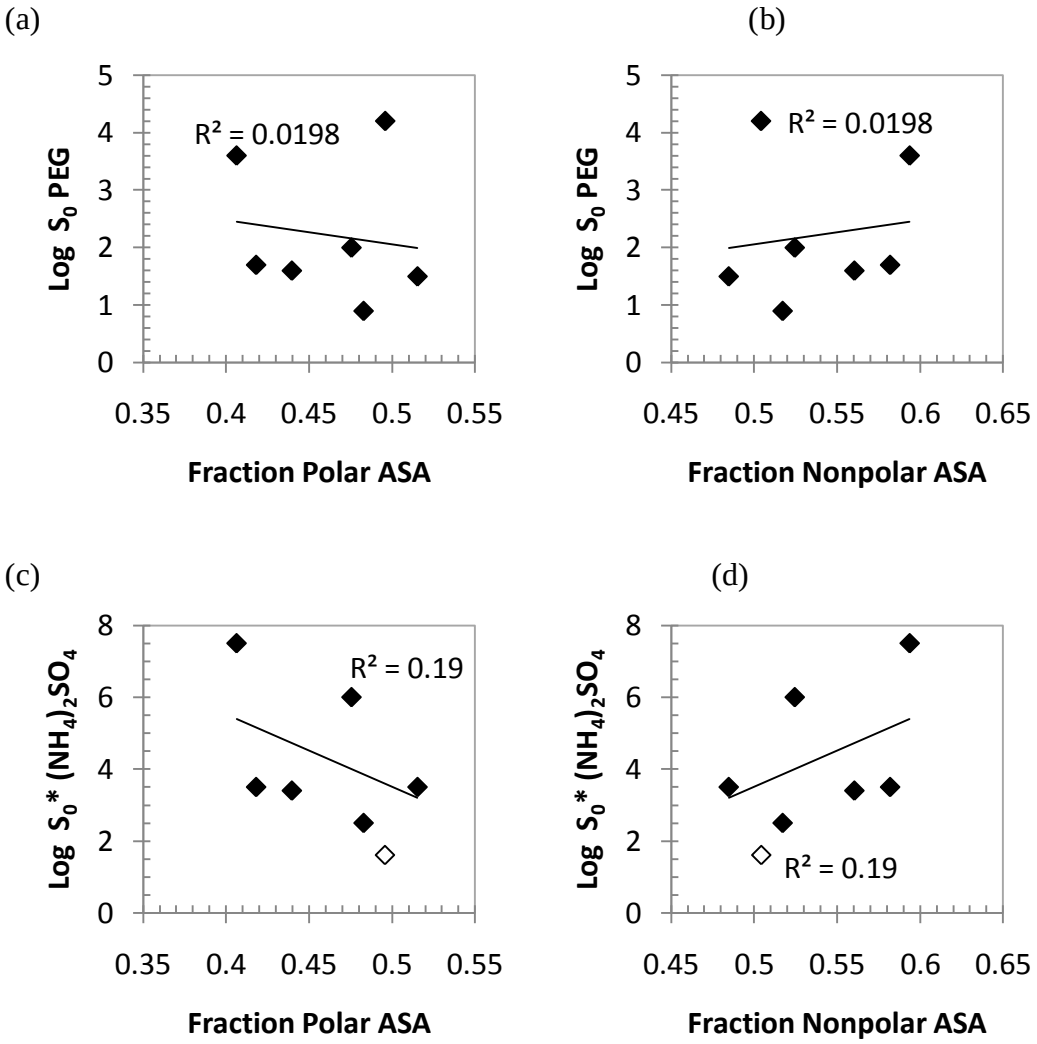


Figure 18. The correlation of fraction polar and nonpolar ASA with PEG-8000 and ammonium sulfate solubility measurements. The ASA for all atoms was calculated using PDB files and either pfis[86] or naccess.[87] Carbon and sulfur atoms are considered nonpolar and nitrogen and oxygen atoms are considered polar. Log S_0 values versus fraction polar ASA (a and c) and fraction nonpolar ASA (c and d) are shown. The lines and R^2 values are from linear least squares fits. α-lactalbumin is shown as open diamonds and is excluded from the ammonium sulfate fits.

ammonium sulfate, a better correlation was observed, but the correlation with polar and nonpolar surface residues is negative and positive respectively, the opposite of what we would have expected and what was observed in PEG-8000.

The contribution of the ASA that is charged, positively charged, and negatively charged was evaluated (see Table 10 columns 9-11). Figure 19 depicts the correlations of PEG-8000 and ammonium sulfate with the fraction charged (Figure 19 (a) and (b)), positively charged (Figure 19 (c) and(d)), and negatively charged (Figure 19 (e) and(f)) ASA. We find a strong correlation between solubility in PEG-8000 and fraction of charged ASA and a more moderate correlation for ammonium sulfate. A correlation between solubility and fraction positively charged ASA was not observed for either PEG-8000 or ammonium sulfate; however, a very strong correlation was observed between solubility and fraction of negatively charged ASA in both PEG-8000 and ammonium sulfate. This strong correlation suggests that negatively charged surface area plays a significant role in determining protein solubility. This is supported by our previous findings that aspartic and glutamic acids contribute more favorably to protein solubility than any of the other 18 amino acids.[37] In order to understand the difference in contribution to protein solubility of negative versus positive charges, one needs to understand the properties of negatively and positively charged groups in proteins. Negatively charged groups in proteins include the kosmotropic carboxylate groups of aspartic acid and glutamic acid residues. Positively charged groups include the chaotropic amino and guanidino groups of lysine and arginine. Kim Collins' work[64,99-101] on ions in solution describes Hofmeister series dependence for hydration of ions in

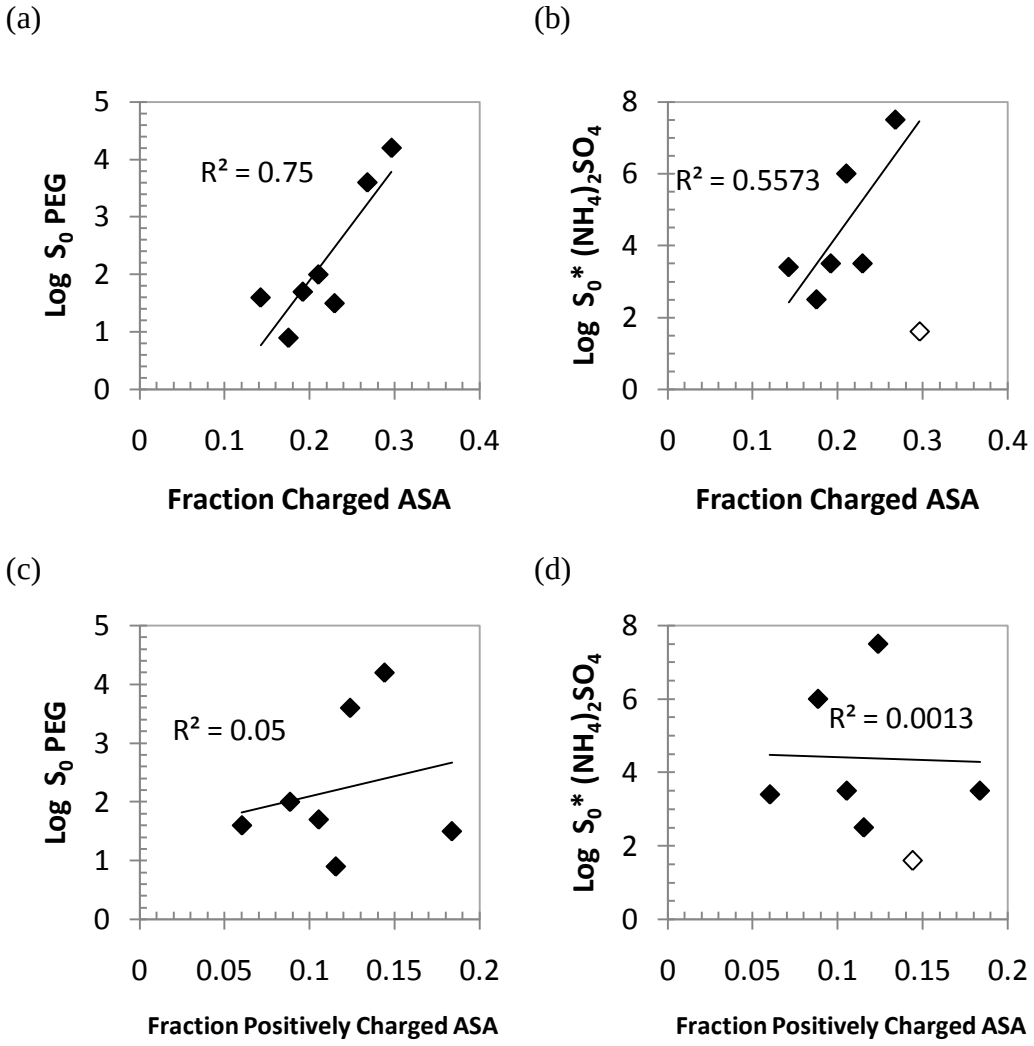
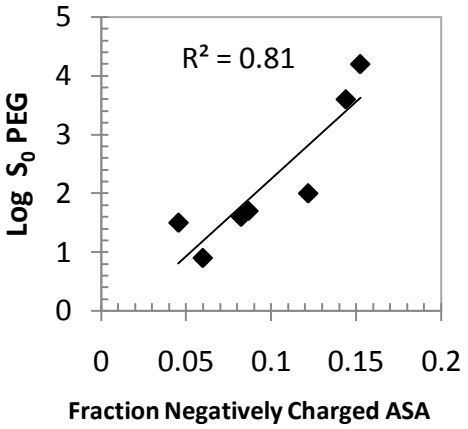


Figure 19. The correlation of the fraction of ASA that is charged (a and b), positively charged (c and d), and negatively charged (e and f) at pH 7.0 with PEG-8000 and ammonium sulfate solubility measurements. The ASA for all atoms was calculated using PDB files and either pfis[86] or naccess.[87] The oxygen atoms glutamic acid and aspartic acid side chains and the c-terminus are considered negatively charged and the nitrogen atoms from arginine and lysine side chains and the n-terminus are considered positively charged at pH 7. The lines and R^2 values are from linear least squares fits. α-lactalbumin is shown as open diamonds and is excluded from the ammonium sulfate fits.

(e)



(f)

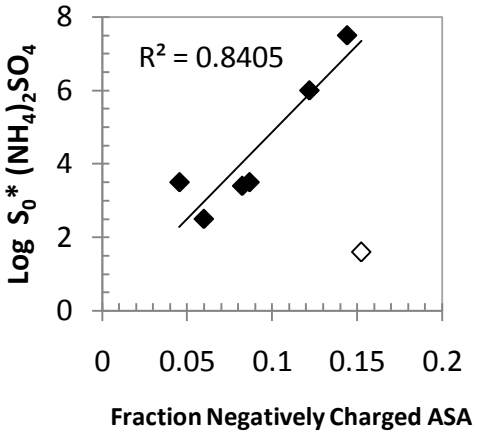


Figure 19. Continued.

solution. Collins showed that kosmotropes are highly hydrated and bind water more tightly than water binds itself, whereas chaotropes bind water weaker than water binds itself and remain largely unhydrated in solution. Therefore, the differential contribution to solubility of negative and positive groups on the protein surface appears to be due to the differential hydration of the carboxylates which bind water tightly and the amino and guanadino groups which bind water weakly.

Conclusions

Negative surface charge was found to have the strongest correlation with increased protein solubility as measured by ammonium sulfate and PEG-8000 precipitation experiments. This is best explained by the strong water binding properties of glutamic and aspartic acid.[64] No correlation between solubility and positive surface charge was seen. A correlation was not observed when we investigated the surface of the protein with respect to polarity. While comparing the two precipitants, we found that ammonium sulfate markedly increases protein stability for all proteins in this study, while polyethylene glycol can have slight stabilizing or destabilizing effects. We found a remarkable correlation between the solubility results obtained with PEG-8000 and ammonium sulfate, even though the mechanisms by which they lower solubility is very different. This suggests that the solubility experiments using these two precipitants are probing similar intrinsic properties of the protein, making the choice between precipitants largely one of convenience. Since polyethylene glycol precipitations can yield a quantitative estimate of solubility in buffer alone and ammonium sulfate can only

determine comparative solubility, polyethylene glycol would be a better choice of precipitant if absolute solubility values are of interest.

CHAPTER VI

SUMMARY

In this study, we employ the use of precipitants to measure protein solubility, because solubility measurements in buffer alone are often difficult or impossible to make for proteins, due to their non-ideal behaviors at high concentrations. Precipitants are useful for the following reasons: (1) they allow you to work at lower protein concentrations where the behavior of protein solutions are closer to ideality, (2) they allow you to induce amorphous precipitation and reach equilibrium rapidly, avoiding complications involving super-saturation, (3) far less protein is required to make solubility measurements than in buffer alone. Solubility values obtained from precipitants are best used comparatively, as determination of accurate solubility values alone is not usually possible. Precipitants fall into three main classes: salts, organic solvents, and polymers. We have used protein solubility measurements made using representatives from each class to gain insight into the intrinsic factors that influence protein solubility.

In Chapter III, 20 variants of RNase Sa were used as a model system to study the amino acid contribution to protein solubility. The 20 variants of RNase Sa differed by the identity of the amino acid at position 76, a surface position entirely exposed to solvent. This allowed us to determine the contribution to protein solubility of all 20 amino acids. We measured the solubility of the 20 variants in PEG-8000 and

isopropanol and compared them with measurements made in ammonium sulfate. WT RNase Sa (threonine at position 76) is near the center of each solubility scale and has an average solubility value as compared to the other 19 variants. Amino acids roughly divide into three classes based on their contribution to protein solubility: amino acids that increase protein solubility (aspartic acid, glutamic acid, serine, lysine, and glycine), those that have solubilities similar to threonine (histidine, asparagine, proline, glutamine, and alanine), and those that decrease protein solubility (isoleucine, valine, leucine, cysteine, tyrosine, phenyalanine, and tryptophan).

Aspartic acid, glutamic acid, and serine had the most favorable contribution to protein solubility, and several polar amino acids (threonine, glutamine, and asparagine) did not have favorable contributions to protein solubility. This led us to propose the following strategy for increasing protein solubility: substitute exposed hydrophobic, asparagine, glutamine, and threonine residues for aspartic acid or glutamic acid if changing the net charge does not matter or to serine if the net charge needs to stay the same. This strategy has two advantages over simply changing hydrophobic to hydrophilic residues: (1) hydrophobic residues are buried more often than they are on the surface of the protein, so there may not be many hydrophobic residues on the surface of the protein to choose from if the structure is know, or if the structure is not know, blind substitution of buried hydrophobic residues may lead to destabilizing the protein, (2) as seen in Chapter III, not all hydrophilic residues contribute favorably to protein solubility. We tested this strategy for increasing protein solubility on a variant of RNase Sa with low solubility (T76W) in Chapter IV. We made surface mutations to serine or aspartic

acid at three exposed positions that had polar amino acids that contributed unfavorably to protein solubility in the wild type protein (asparagine 20, threonine 46, and glutamine 77). We measured the solubilities of these variants using ammonium sulfate, PEG-8000, and isopropanol. Serine variants only showed an increase in solubility in ammonium sulfate; however, the aspartic acid variants showed significant increases in solubility in all three precipitants, and these results were independent of whether the net charge on the protein was positive or negative.

To gain insight into the intrinsic factors that influence protein solubility, we measured the solubility of seven proteins in Chapter V. We compared the following properties of these proteins to the solubility results: net charge, molecular weight, and the amino acid composition of the surface of the protein including the surface area that was polar, non-polar, charged, negatively charge, and positively charge. We found that negative surface charge had the strongest correlation to protein solubility. No correlation to positive surface charge, net charge, molecular weight, or surface polarity was seen.

Protein solubility is an equilibrium property between the solid and solution phases and is determined by the respective chemical potentials of the two phases. Protein solubility is complicated and cannot be determined by any single property of the protein; however, we find that negative surface charges play an important role. This role has been seen several times throughout this study. Aspartic acid and glutamic acid have the most favorable contributions to protein solubility of all 20 natural amino acids. Aspartic acid surface mutations significantly increase the solubility of a low solubility

variant of RNase Sa. In a group of proteins studied, increasing negative surface charge shows the strongest correlation with increasing protein solubility. How can the role of negative charges be understood? A protein commonly has a minimum solubility near the pI of the protein, and the solubility increases as net positive or negative charge increases, likely due to electrostatic repulsion between protein molecules at higher net charge.[44] Charge repulsion may partially explain the contribution to protein solubility seen for the negatively charged residues in proteins; however, if this was the primary mechanism by which negative charge increases protein solubility, we would expect to see the same behavior for positive charges. This behavior is not seen; the contribution to solubility of positive charges is variable and not as favorable as negative charges. We would also expect to see a diminished or even negative effect on protein solubility of adding negative charges at net positive charge; this is also not seen. The difference in solubilities between positively and negatively charged residues in proteins may best be explained by differences in hydration. Kim Collins'[64] has extensively studied the hydration properties of ions in solution. He has found that kosmotropic ions, like the carboxylate groups of aspartic acid and glutamic acid, bind water tighter than water binds itself, and are preferentially hydrated in solution. In contrast, chaotropic ions, like the guanidinium group of arginine and the amino group of lysine, are weakly hydrated, bind water more weakly than water binds itself, and primarily lack waters of hydration in solution.

REFERENCES

1. Bagby, S., Tong, K. I., Ikura, M., Thomas L. James, V. D. & Uli, S. (2001). Optimization of protein solubility and stability for protein nuclear magnetic resonance. *Methods in Enzymol.* **339,** 20-41.
2. Caldwell, G. W., Ritchie, D. M., Masucci, J. A., Hageman, W. & Yan, Z. (2001). The new pre-preclinical paradigm: compound optimization in early and late phase drug discovery. *Curr. Top Med. Chem.* **1**, 353-66.
3. Riès-kautt, M. & Ducruix, A. (1997). Inferences drawn from physicochemical studies of crystallogenesis and precrystalline state. *Methods in Enzymol.* **276**, 23-59.
4. Schein, C. H. (1993). Solubility and secretability. *Curr. Opin. in Biotechnol.* **4**, 456-461.
5. Fowler, S. B., Poon, S., Muff, R., Chiti, F., Dobson, C. M. & Zurdo, J. (2005). Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin. *Proc. Natl. Acad. Sci. U S A* **102**, 10105-10.
6. Ricci, M. S. & Brems, D. N. (2004). Common structural stability properties of 4-helical bundle cytokines: possible physiological and pharmaceutical consequences. *Curr. Pharm. Des.* **10**, 3901-11.
7. Evans, P., Wyatt, K., Wistow, G. J., Bateman, O. A., Wallace, B. A. & Slingsby, C. (2004). The P23T cataract mutation causes loss of solubility of folded gammaD-crystallin. *J. Mol. Biol.* **343**, 435-44.

8. Pande, A., Annunziata, O., Asherie, N., Ogun, O., Benedek, G. B. & Pande, J. (2005). Decrease in protein solubility and cataract formation caused by the Pro23 to Thr mutation in human gamma D-crystallin. *Biochemistry* **44**, 2491-500.

9. Kim, W. & Hecht, M. H. (2006). Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's Abeta42 peptide. *Proc. Natl. Acad. Sci. U S A* **103**, 15824-9.

10. Kaytor, M. D. & Warren, S. T. (1999). Aberrant Protein Deposition and Neurological Disease. *J. Biol. Chem.* **274**, 37507-37510.

11. Bellotti, V. & Merlini, G. (1996). Current concepts on the pathogenesis of systemic amyloidosis. *Nephrol. Dial. Transplant* **11 Suppl 9**, 53-62.

12. Eaton, W. A. & Hofrichter, J. (1990). Sickle cell hemoglobin polymerization. *Adv. Protein Chem.* **40**, 63-279.

13. Peters, T. (1996). *All about albumin: biochemistry, genetics, and medicalapplications,* Academic Press, Inc, San Diego, Ca.

14. Stec, B., Rao, U. & Teeter, M. M. (1995). Refinement of purothionins reveals solute particles important for lattice formation and toxicity. Part 2: structure of beta-purothionin at 1.7 A resolution. *Acta Crystallogr D Biol Crystallogr* **51**, 914-24.

15. Ahn, H. C., Juranic, N., Macura, S. & Markley, J. L. (2006). Three-dimensional structure of the water-insoluble protein crambin in dodecylphosphocholine micelles and its minimal solvent-exposed surface. *J. Am. Chem. Soc.* **128**, 4398-404.

16. Cohn, E. J. & Edsall, J. T. (1943). *Proteins, Amino Acids, and Peptides*, Hafner Publishing Company, New York.

17. Arakawa, T. & Timasheff, S. N. (1985). Theory of protein solubility. *Methods Enzymol.* **114**, 49-77.

18. Feher, G. & Kam, Z. (1985). Nucleation and growth of protein crystals: general principles and assays. *Methods Enzymol.* **114**, 77-112.

19. Trevino, S. R., Scholtz, J. M. & Pace, C. N. (2008). Measuring and increasing protein solubility. *J. Pharm. Sci*.

20. Middaugh, C. V., D. (1992). Protein solubility. In *Stability of Protein Pharmaceuticals, Pt. A* (Ahern, T. M., M., ed.), pp. 109-134. Plenum Press, New York.

21. Wilkinson, D. L. & Harrison, R. G. (1991). Predicting the solubility of recombinant proteins in Escherichia coli. *Biotechnology (N Y)* **9**, 443-8.

22. Ito, T. & Wagner, G. (2004). Using codon optimization, chaperone co-expression, and rational mutagenesis for production and NMR assignments of human eIF2 alpha. *J. Biomol. NMR* **28**, 357-67.

23. Idicula-Thomas, S. & Balaji, P. V. (2005). Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. *Protein Sci.* **14**, 582-92.

24. Schein, C. H. (1990). Solubility as a function of protein structure and solvent components. *Biotechnology (N Y)* **8**, 308-17.

25. Scallon, B., Cai, A., Solowski, N., Rosenberg, A., Song, X. Y., Shealy, D. & Wagner, C. (2002). Binding and functional comparisons of two types of tumor necrosis factor antagonists. *J. Pharmacol. Exp. Ther.* **301**, 418-26.

26. Leader, B., Baca, Q. J. & Golan, D. E. (2008). Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov.* **7**, 21-39.

27. Walsh, M. & Jayne, D. (2007). Rituximab in the treatment of anti-neutrophil cytoplasm antibody associated vasculitis and systemic lupus erythematosus: past, present and future. *Kidney Int.* **72**, 676-82.

28. Gelfand, E. W. (2001). Antibody-directed therapy: past, present, and future. *J Allergy Clin. Immunol.* **108**, S111-6.

29. Lobo, E. D., Hansen, R. J. & Balthasar, J. P. (2004). Antibody pharmacokinetics and pharmacodynamics. *J. Pharm. Sci.* **93**, 2645-68.

30. Liu, J., Nguyen, M. D., Andya, J. D. & Shire, S. J. (2005). Reversible self-association increases the viscosity of a concentrated monoclonal antibody in aqueous solution. *J Pharm. Sci.* **94**, 1928-40.

31. Harn, N., Allan, C., Oliver, C. & Middaugh, C. R. (2007). Highly concentrated monoclonal antibody solutions: direct analysis of physical structure and thermal stability. *J. Pharm. Sci.* **96**, 532-46.

32. Dani, B., Platz, R. & Tzannis, S. T. (2007). High concentration formulation feasibility of human immunoglubulin G for subcutaneous administration. *J Pharm. Sci.* **96**, 1504-17.

33. Campbell, P. & Marcus, R. (2003). Monoclonal antibody therapy for lymphoma. *Blood. Rev.* **17**, 143-52.

34. Shire, S. J., Shahrokh, Z. & Liu, J. (2004). Challenges in the development of high protein concentration formulations. *J. Pharm. Sci.* **93**, 1390-402.

35. Tandford, C. (1961). *Physical Chemistry of Macromolecules*, John Wiley and Sons, Inc., New York.

36. Melander, W. & Horvath, C. (1977). Salt effect on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series. *Arch. Biochem. Biophys.* **183**, 200-15.

37. Trevino, S. R., Scholtz, J. M. & Pace, C. N. (2007). Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *J. Mol. Biol.* **366**, 449-60.

38. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L. Y. (2004). A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc.* **126**, 8933-9.

39. McPherson, A. (2004). Introduction to protein crystallization. *Methods* **34**, 254-65.

40. Brems, D. N. (1988). Solubility of different folding conformers of bovine growth hormone. *Biochemistry* **27**, 4541-4546.

41. Tayyab, S., Qamar, S. & Islam, M. (1993). Protein solubility: An old issue gaining momentum. *Medical Science Research* **21**, 805-809.

42. Middaugh, C. R. & Volkin, D. B. (1992). Protein Solubility. In *Stability of Protein Pharmaceuticals* (Ahern, T. J. & Manning, M. C., eds.), pp. 109-134. Plenum Press, New York.

43. Cohn, E. J. & Edsall, J. T. (1965). *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, Hafner Publishing Co, New York.

44. Tanford, C. (1961). *Physical Chemistry of Macromolecules*, John Wiley and Sons, Inc., New York.

45. Rupley, J. A. (1968). Comparison of protein structure in the crystal and in solution: IV. Protein solubility. *J. Mol. Biol.* **35**, 455-476.

46. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. (2001). The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci.* **10**, 1206-15.

47. Chase, S. F. & Laue, T. M. (2008). The determination of protein valence by capillary electrophoresis. *PACE Setter* **12**, 1-5.

48. Winzor, D. J. (2004). Determination of the net charge (valence) of a protein: a fundamental but elusive parameter. *Anal. Biochem.* **325**, 1-20.

49. Poillon, W. N. (1980). Noncovalent inhibitors of sickle hemoglobin gelation: effects of aliphatic alcohols, amides, and ureas. *Biochemistry* **19**, 3194-3199.

50. Arakawa, T. & Timasheff, S. N. (1985). Mechanism of poly(ethylene glycol) interaction with proteins. *Biochemistry* **24**, 6756-62.

51. Atha, D. H. & Ingham, K. C. (1981). Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *J. Biol. Chem.* **256**, 12108-17.

52. Byrne, N. & Angell, C. A. (2010). The solubility of hen lysozyme in ethylammonium nitrate/H2O mixtures and a novel approach to protein crystallization. *Molecules* **15**, 793-803.

53. Evans, D. R., Romero, J. K. & Westoby, M. (2009). Concentration of proteins and removal of solutes. *Methods Enzymol.* **463**, 97-120.

54. Pohl, T. (1990). Concentration of proteins and removal of solutes. *Methods Enzymol.* **182**, 68-83.

55. Mirarefi, A. Y. & Zukoski, C. F. (2004). Gradient diffusion and protein solubility: use of dynamic light scattering to localize crystallization conditions. *Journal of Crystal Growth* **265**, 274-283.

56. Shih, Y., Prausnitz, J. & Blanch, H. (1992). Some Characteristics of Protein Precipitation by Salts. *Biotechnol. Bioengin.* **40**, 1155-1164.

57. Middaugh, C. R., Tisel, W. A., Haire, R. N. & Rosenberg, A. (1979). Determination of the apparent thermodynamic activities of saturated protein solutions. *J. Biol. Chem.* **254**, 367-70.

58. Shiau, K. & Chen, T. (1997). Initial protein concentration effects on precipitation by salt. *Biotechnol. Bioengin.* **53**, 202-206.

59. Annunziata, O., Payne, A. & Wang, Y. (2008). Solubility of lysozyme in the presence of aqueous chloride salts: common-ion effect and its role on solubility and crystal thermodynamics. *J. Am. Chem. Soc.* **130**, 13347-52.

60. Przybycien, T. M. & Bailey, J. E. (1989). Aggregation kinetics in salt-induced protein precipitation. *AIChE Journal* **35**, 1779-1790.

61. Saluja, A., Crampton, S., Kras, E., Fesinmeyer, R. M., Remmele, R. L., Jr., Narhi, L. O., Brems, D. N. & Gokarn, Y. R. (2009). Anion binding mediated precipitation of a peptibody. *Pharm. Res.* **26**, 152-60.

62. Kumar, V., Sharma, V. K. & Kalonia, D. S. (2009). Effect of polyols on polyethylene glycol (PEG)-induced precipitation of proteins: Impact on solubility, stability and conformation. *Int. J. of Pharm.* **366**, 38-43.

63. Creighton, T. E. (1993). *Proteins: Structures and Molecular Properties*. 2nd ed., W. H. Freeman, New York.

64. Collins, K. D. (1997). Charge density-dependent strength of hydration and biological structure. *Biophys. J.* **72**, 65-76.

65. Winzor, D. J. & Wills, P. R. (2006). Molecular crowding effects of linear polymers in protein solutions. *Biophys. Chem.* **119**, 186-95.

66. Bhat, R. & Timasheff, S. N. (1992). Steric exclusion is the principal source of the preferential hydration of proteins in the presence of polyethylene glycols. *Protein Sci.* **1**, 1133-43.

67. Lee, J. C. & Lee, L. L. (1979). Interaction of calf brain tubulin with poly(ethylene glycols). *Biochemistry* **18**, 5518-26.

68. Farruggia, B., Garcia, G., D'Angelo, C. & Picó, G. (1997). Destabilization of human serum albumin by polyethylene glycols studied by thermodynamical equilibrium and kinetic approaches. *Int. J. of Biol. Macromolecules* **20**, 43-51.

69. Baldwin, R. L. (1996). How Hofmeister ion interactions affect protein stability. *Biophys. J.* **71**, 2056-63.

70. Thomas, P. D. & Dill, K. A. (1993). Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. *Protein Sci.* **2**, 2050-65.

71. Avramopoulou, V., Mamalaki, A. & Tzartos, S. J. (2004). Soluble, oligomeric, and ligand-binding extracellular domain of the human alpha7 acetylcholine receptor expressed in yeast: replacement of the hydrophobic cysteine loop by the hydrophilic loop of the ACh-binding protein enhances protein solubility. *J. Biol. Chem.* **279**, 38287-93.

72. Dale, G. E., Broger, C., Langen, H., D'Arcy, A. & Stuber, D. (1994). Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase. *Protein Eng.* **7**, 933-9.

73. Das, D. & Georgiadis, M. M. (2001). A directed approach to improving the solubility of Moloney murine leukemia virus reverse transcriptase. *Protein Sci.* **10**, 1936-41.

74. Fan, D., Li, Q., Korando, L., Jerome, W. G. & Wang, J. (2004). A monomeric human apolipoprotein E carboxyl-terminal domain. *Biochemistry* **43**, 5055-64.

75. Fowler, S. B., Poon, S., Muff, R., Chiti, F., Dobson, C. M. & Zurdo, J. (2005). Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin. *Proc. Natl. Acad. Sci. U S A* **102**, 10105-10110.

76. Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlando, R., Davies, D. R. & Craigie, R. (1995). Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc. Natl. Acad. Sci. U S A* **92**, 6057-61.

77. Li, Y., Yan, Y., Zugay-Murphy, J., Xu, B., Cole, J. L., Witmer, M., Felock, P., Wolfe, A., Hazuda, D., Sardana, M. K., Chen, Z., Kuo, L. C. & Sardana, V. V. (1999). Purification, solution properties and crystallization of SIV integrase containing a continuous core and C-terminal domain. *Acta Crystallogr D Biol Crystallogr* **55**, 1906-10.

78. Mosavi, L. K. & Peng, Z. Y. (2003). Structure-based substitutions for increased solubility of a designed protein. *Protein Eng.* **16**, 739-45.

79. Roosild, T. P. & Choe, S. (2005). Redesigning an integral membrane K+ channel into a soluble protein. *Protein Eng. Des. Sel.* **18**, 79-84.

80. Slovic, A. M., Kono, H., Lear, J. D., Saven, J. G. & DeGrado, W. F. (2004). Computational design of water-soluble analogues of the potassium channel KcsA. *Proc. Natl. Acad. Sci. U S A* **101**, 1828-33.

81. Zhang, F., Basinski, M. B., Beals, J. M., Briggs, S. L., Churgay, L. M., Clawson, D. K., DiMarchi, R. D., Furman, T. C., Hale, J. E., Hsiung, H. M., Schoner, B.

E., Smith, D. P., Zhang, X. Y., Wery, J. P. & Schevitz, R. W. (1997). Crystal structure of the obese protein leptin-E100. *Nature* **387**, 206-9.

82. Lesser, G. J. & Rose, G. D. (1990). Hydrophobicity of amino acid subgroups in proteins. *Proteins* **8**, 6-13.

83. Pace, C. N. & Sholtz, J. M. (1997). Measuring the conformational stability of a protein. In *Protein Structure: A Practical Approach* (Creighton, T. E., ed.), pp. 229-321. IRL press, USA.

84. Hebert, E. J., Grimsley, G. R., Hartley, R. W., Horn, G., Schell, D., Garcia, S., Both, V., Sevcik, J. & Pace, C. N. (1997). Purification of ribonucleases Sa, Sa2, and Sa3 after expression in *Escherichia coli*. *Protein Expr. Purif.* **11**, 162-168.

85. Gibbs, S. J., Chu, A. S., Lightfoot, E. N. & Root, T. W. (1991). Ovalbumin diffusion at low ionic strength. *J. Phys. Chem.* **95**, 467-471.

86. Hebert, E. J., Giletto, A., Sevcik, J., Urbanikova, L., Wilson, K. S., Dauter, Z. & Pace, C. N. (1998). Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba, and T1. *Biochemistry* **37**, 16192-200.

87. Hubbard, S. J. & Thorton, J. M. (1993). NACCESS. Department of Biochemistry and Molecular Biology, University College London.

88. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

89. Kollman, J. M., Pandi, L., Sawaya, M. R., Riley, M. & Doolittle, R. F. (2009). Crystal Structure of Human Fibrinogen. *Biochemistry* **48**, 3877-3886.

90. Bhattacharya, A. A., Curry, S. & Franks, N. P. (2000). Binding of the general anesthetics propofol and halothane to human serum albumin. *J. Biol. Chem.* **275**, 38731-38738.

91. LeBeau, A. M., Singh, P., Isaacs, J. T. & Denmeade, S. R. (2009). Prostate-specific antigen as a "chymotrypsin-like" serine protease with unique P1 Substrate Specificity. *Biochemistry* **48**, 3490-3496.

92. Chrysina, E. D., Brew, K. & Acharya, K. R. (2000). Crystal structures of apo- and holo-bovine α-lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions. *J. Biol. Chem.* **275**, 37021-37029.

93. Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. (2007). Triclinic lysozyme at 0.65 A resolution. *Acta Crystallographica Section D* **63**, 1254-1268.

94. Stein, P. E., Leslie, A. G. W., Finch, J. T. & Carrell, R. W. (1991). Crystal structure of uncleaved ovalbumin at 1·95 Å resolution. *J. Mol. Biol.* **221**, 941-959.

95. Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). Ribonuclease from Streptomyces aureofaciens at atomic resolution. *Acta Crystallogr D Biol Crystallogr* **52**, 327-44.

96. Laurents, D. V., Huyghues-Despointes, B. M., Bruix, M., Thurlkill, R. L., Schell, D., Newsom, S., Grimsley, G. R., Shaw, K. L., Trevino, S., Rico, M., Briggs, J. M., Antosiewicz, J. M., Scholtz, J. M. & Pace, C. N. (2003). Charge-charge interactions are key determinants of the pK values of ionizable groups in

ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2). *J. Mol. Biol.* **325**, 1077-92.

97. Shepherd, A. J., Gorse, D. & Thornton, J. M. (1999). Prediction of the location and type of beta-turns in proteins using neural networks. *Protein Sci.* **8**, 1045-55.

98. Trevino, S. R., Schaefer, S., Scholtz, J. M. & Pace, C. N. (2007). Increasing protein conformational stability by optimizing beta-turn sequence. *J. Mol. Biol.* **373**, 211-8.

99. Collins, K. D. (1995). Sticky ions in biological systems. *Proc. Natl. Acad. Sci. U S A* **92**, 5553-7.

100. Collins, K. D. (2004). Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods* **34**, 300-11.

101. Collins, K. D., Neilson, G. W. & Enderby, J. E. (2007). Ions in water: characterizing the forces that control chemical processes and biological structure. *Biophys. Chem.* **128**, 95-104.

102. Gibbs, A. C., Bjorndahl, T. C., Hodges, R. S. & Wishart, D. S. (2002). Probing the structural determinants of type II' beta-turn formation in peptides and proteins. *J. Am. Chem. Soc.* **124**, 1203-13.

103. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-23.

104. Thurlkill, R. L., Grimsley, G. R., Scholtz, J. M. & Pace, C. N. (2006). pK values of the ionizable groups of proteins. *Protein Sci.* **15**, 1214-8.

105. Nasreen, A., Vogt, M., Kim, H. J., Eichinger, A. & Skerra, A. (2006). Solubility engineering and crystallization of human apolipoprotein D. *Protein Sci.* **15**, 190-9.

106. Daujotyte, D., Vilkaitis, G., Manelyte, L., Skalicky, J., Szyperski, T. & Klimasauskas, S. (2003). Solubility engineering of the HhaI methyltransferase. *Protein Eng.* **16**, 295-301.

107. Agren, L., Norin, M., Lycke, N. & Lowenadler, B. (1999). Hydrophobicity engineering of cholera toxin A1 subunit in the strong adjuvant fusion protein CTA1-DD. *Protein Eng.* **12**, 173-8.

108. Park, J. H. & Batt, C. A. (2004). Restoration of a defective Lactococcus lactis xylose isomerase. *Appl. Environ. Microbiol.* **70**, 4318-25.

109. Tobler, S. A., Sherman, N. E. & Fernandez, E. J. (2000). Tracking lysozyme unfolding during salt-induced precipitation with hydrogen exchange and mass spectrometry. *Biotechnol. Bioeng.* **71**, 194-207.

110. Greene, L. H., Grobler, J. A., Malinovskii, V. A., Tian, J., Acharya, K. R. & Brew, K. (1999). Stability, activity and flexibility in α-lactalbumin. *Protein Eng.* **12**, 581-587.

111. Donovan, J. W. & Mihalyi, E. (1974). Conformation of fibrinogen: calorimetric evidence for a three-nodular structure. *Proc. Natl. Acad. Sci. U S A* **71**, 4125-8.

112. Sugio, S., Kashima, A., Mochizuki, S., Noda, M. & Kobayashi, K. (1999). Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng.* **12**, 439-446.

113. Wickström, E., Persson Waller, K., Lindmark-Månsson, H. & Sternesjö, Å. (2010). Short communication: Relationships between [alpha]-lactalbumin and quality traits in bulk milk. *J. Dairy Sci.* **93**, 4577-4581.

114. Protein Calculator, http://www.scripps.edu/~cdputnam/protcalc.html.

VITA

Name:	Ryan Mahnken Kramer
Address:	c/o Dr. J. Martin Scholtz 440 Reynolds Medical Building 1114 TAMU, College Station, TX 77843-1114
Email Address:	kramer513@gmail.com
Education:	B.S.,Biochemistry, Washington State University, 2005 Ph.D., Biochemistry, Texas A&M University, 2011