

Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility

Ryan M. Kramer,[†] Varad R. Shende,[§] Nicole Motl,[†] C. Nick Pace,^{†‡} and J. Martin Scholtz^{†‡*}

[†]Department of Biochemistry and Biophysics and [‡]Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, Texas A&M University, College Station, Texas; and [§]Department of Chemistry, University of Bremen, Bremen, Germany

ABSTRACT Protein solubility is a problem for many protein chemists, including structural biologists and developers of protein pharmaceuticals. Knowledge about how intrinsic factors influence solubility is limited due to the difficulty of obtaining quantitative solubility measurements. Solubility measurements in buffer alone are difficult to reproduce, because gels or supersaturated solutions often form, making it impossible to determine solubility values for many proteins. Protein precipitants can be used to obtain comparative solubility measurements and, in some cases, estimations of solubility in buffer alone. Protein precipitants fall into three broad classes: salts, long-chain polymers, and organic solvents. 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 for the two different precipitants agree closely with each other, suggesting that the two precipitants probe similar properties that are relevant to solubility in buffer alone.

INTRODUCTION

Protein solubility is important to structural biologists (1), the pharmaceutical industry (2), and all scientists who work with protein in solution. Structural studies (1,3,4) and pharmaceutical applications (2,5,6) often require very-high-concentration protein samples. The solubility of a protein in aqueous solution varies from almost completely insoluble to hundreds of milligrams per milliliter. For instance, crambin has been reported to be completely insoluble in water (7), and serum albumins have solubilities of >500 mg/mL (8). Low protein solubility has also been implicated in a number of human diseases (9–12). The P23T mutation in human γ D-crystallin shows a markedly decreased solubility and leads to childhood onset of cataracts (10). Therefore, understanding the factors that contribute to protein solubility is an important area of research.

Protein solubility is a thermodynamic parameter defined as 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 (13,14). Solubility can be influenced by a number of extrinsic and intrinsic factors. Extrinsic factors that influence protein solubility include pH, ionic strength, temperature, and the presence of various solvent additives (3). Varying these extrinsic factors can lead to increased solubility (1,4); however, altering the solution conditions is not always appropriate or sufficient to increase protein solubility to the extent required. The intrinsic factors that influence protein solubility are defined primarily by the amino acids on the protein surface, but a detailed understanding of how one can alter the intrinsic properties of a protein to increase its

solubility is lacking (1,4,15). Recently, our laboratory has taken steps toward elucidating the intrinsic factors that influence protein solubility, with the goal of developing a strategy for making mutations that increase protein solubility (16,17).

One of the reasons for the lack of understanding of intrinsic protein solubility is the difficulty of obtaining quantitative solubility measurements. Two methods used to measure protein solubility in aqueous solution are: 1), adding lyophilized protein to solvent; and 2), concentrating a protein solution by ultrafiltration. Both of these methods require that the concentration of protein in solution be increased until saturation is reached; however, this is often difficult to do, especially with very soluble proteins, because gel-like or supersaturated solutions may form, making it difficult to determine the solubility values accurately (15). When lyophilized protein is added to solvent, the variable water and salt content of the lyophilized powder is difficult to control and can have a significant effect on solubility measurements (15).

One way to avoid the difficulties of measuring protein solubility is to make use of an extraneous agent that lowers the solubility of a protein called a precipitant. Protein precipitants can be divided into three classes: salts, organic solvents, and long-chain polymers (18). These precipitants are 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 (16,17,19–25). Common examples from each of these three classes of precipitants, respectively, are ammonium sulfate, isopropanol, and polyethylene glycol (PEG). The relationship between precipitant concentration to protein solubility is described by the following general expression (13,19):

Submitted August 18, 2011, and accepted for publication January 30, 2012.

*Correspondence: jm-scholtz@tamu.edu

Editor: Doug Barrick.

© 2012 by the Biophysical Society
0006-3495/12/04/1907/9 \$2.00

doi: 10.1016/j.bpj.2012.01.060

$$\text{Log } S = \text{constant} - \beta[\text{Precipitant}], \quad (1)$$

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. The constant is the y -intercept of the solubility plot, and for PEG precipitations is equal to the logarithm of the protein activity. For dilute protein solutions, as the activity constant approaches one, Eq. 1 becomes

$$\text{Log } S = \text{Log } S_0 - \beta[\text{Precipitant}], \quad (2)$$

where S_0 is the solubility in the absence of precipitant. Midgaugh et al. (19) showed that for PEG precipitations, the linearity of Eq. 1 extends to zero precipitant for proteins whose solubility can be accurately measured in buffer alone. In this case, the constant portion of Eq. 2, $\text{Log } S_0$, can be used as an estimate of solubility in the absence of precipitant. For salts, Eq. 1 only describes the salting-out region of the solubility plot. At low salt concentrations, salting-in is observed and the solubility is higher than in the absence of salt. Therefore, the constant obtained from salt precipitations represents a projection of the salting-out region onto the y axis. A relationship between this constant and protein solubility has not been established for salt precipitations and will be investigated in this study.

Each class of precipitants decreases protein solubility by a different mechanism. For salts, kosmotropic ions bind water more tightly than water binds itself (26) 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 (13). Organic solvents, such as alcohols, 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 increases and the protein precipitates (18). By the same mechanism, organic solvents decrease the strength of hydrophobic interactions, leading to decreased protein stability (27). Special care must be taken to ensure that the protein remains folded under experimental conditions if organic solvents are used as precipitants. Long-chain polymers occupy more space in solution than does a protein of similar molecular mass and lower the solubility of a protein through an excluded volume mechanism (25,28), effectively crowding the protein out of solution. Due to the different mechanisms of action employed by each class of precipitants, the results obtained with one precipitant may not coincide with results obtained with another precipitant, and may relate differently to solubility in the absence of precipitant. Because of this limitation and the absence of a quantitative relationship between solubility in the absence and presence of precipitant, solubility measurements are best used in a comparative manner. Our

purpose in this study was twofold: 1), to compare the solubility results obtained with two different classes of precipitants (i.e., salts (ammonium sulfate) and long-chain polymers (PEG-8000)) for a number of different proteins; and 2), to examine the molecular properties of the proteins used in this study with the hope of gaining a better understanding of the solubility results obtained and insight into the intrinsic factors that influence protein solubility.

MATERIALS AND METHODS

For details about the materials and methods used in this work, see the [Supporting Material](#).

RESULTS

Proteins are folded under experimental conditions

For this study, we are interested in examining the solubility of folded proteins. The low solubility of the unfolded state stands as a challenge for those studying the denatured-state ensemble; however, we chose to focus on the solubility of the native state because of its relevance to crystallographers, protein chemists, and developers of protein pharmaceuticals (15). 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 (18,41). To confirm this, we used thermal unfolding experiments to examine the effect of ammonium sulfate and PEG-8000 on stability. Originally, this study was designed to include isopropanol as a precipitant. Thomas and Dill (27) investigated the mechanism by which alcohols destabilize proteins and found that it was complex and dependent on 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 (R. Kramer, C. Pace, and J. Scholtz, unpublished data), and therefore we excluded the class of organic solvents from this study. The 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 to diminish the denaturing effects.

Figs. 1 and 2 show the change in observed melting temperature, T_m , as a function of ammonium sulfate and PEG-8000, respectively. For ammonium sulfate, the T_m 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

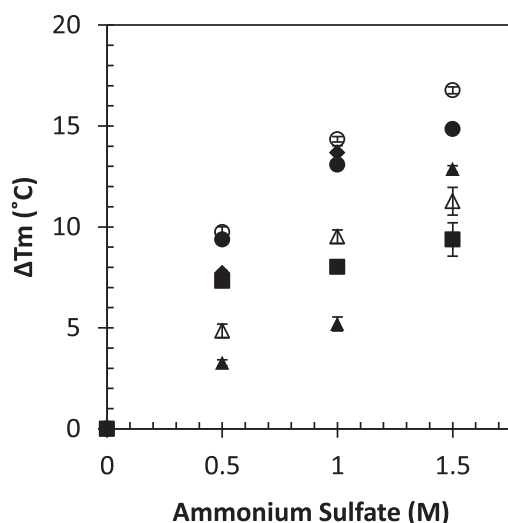


FIGURE 1 Change in T_m as a function of ammonium sulfate concentration for RNase Sa (solid diamonds), α -chymotrypsin (solid circles), lysozyme (open triangles), human serum albumin (solid triangles), ovalbumin (solid squares), and α -lactalbumin (open circles). Thermal denaturation was followed by circular dichroism (see Supporting Material for details).

water-protein interface as compared with the unfolded state (42,43). 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 influence the level 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 ammo-

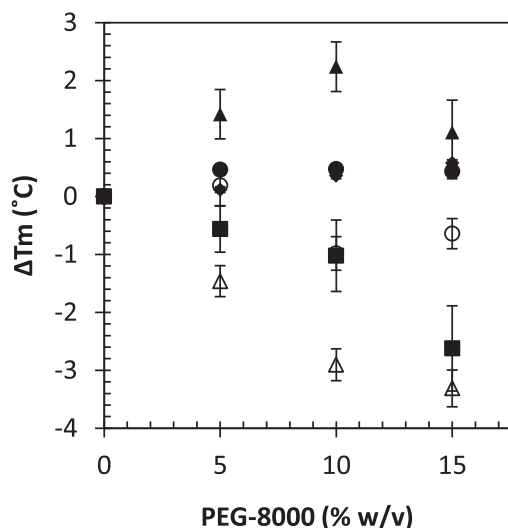


FIGURE 2 Change in T_m as a function of PEG for RNase Sa (solid diamonds), α -chymotrypsin (solid circles), lysozyme (open triangles), human serum albumin (solid triangles), ovalbumin (solid squares), and α -lactalbumin (open circles). Thermal denaturation was followed by circular dichroism (see Supporting Material for details).

nium sulfate is α -lactalbumin. It forms a molten globule upon unfolding (44), and likely buries less surface area than proteins that more fully unfold. This differential stabilization by ammonium sulfate deserves further investigation than can be accomplished here. 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, the protein in both solution and solid phase remains folded (41).

In contrast to ammonium sulfate, high-molecular-mass PEGs are not expected to interact with or have a significant effect on protein stability (25,28); however, at high concentrations they may destabilize some proteins (14), and PEG molecules with a molecular mass of ≤ 6000 Da have been shown to destabilize some proteins (45). Fig. 2 shows that our data agree with those previous observations. The effect of PEG-8000 on stability is small relative to the effect of ammonium sulfate, although in the case of lysozyme and ovalbumin, we observe 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 that show a decrease in stability, the decrease is not enough to cause a significant change in the population of unfolded protein present at room temperature. The T_m -values for 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 $<15\%$ (w/v) for these two proteins. As previously reported (46), fibrinogen exhibits a complicated multistate unfolding curve 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 PEG-8000.

Solubility measurements rapidly reach equilibrium

Because solubility values are defined at equilibrium, it is necessary to ensure that samples are allowed adequate time to equilibrate. For all proteins used in this study, samples were prepared in duplicate and were either centrifuged and quantified immediately or were allowed to sit for 24 h before centrifugation. Measurements obtained immediately and at 24 h yielded the same result (Table 1), 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 agrees with previously reported results for amorphous protein precipitation (16,47). Furthermore, we observed that precipitation could be reversed by the addition of buffer lacking precipitant (data not shown).

TABLE 1 Dependence of solubility measurements on equilibration time

Protein	Precipitant	Equilibration time	Solubility (mg/mL)
α -Chymotrypsin	15% PEG-8000	10 min	5.8 \pm 0.5
		24 h	5.7 \pm 0.2
	2 M ammonium sulfate	10 min	31.9 \pm 0.8
		24 h	31.7 \pm 0.8
Lysozyme	5% PEG-8000	10 min	14.1 \pm 0.2
		24 h	14.1 \pm 0.1
	1.5 M ammonium sulfate	10 min	18.2 \pm 0.7
		24 h	17.7 \pm 0.4
Serum albumin	20% PEG-8000	10 min	9.8 \pm 0.2
		24 h	10.1 \pm 0.2
	2.5 M ammonium sulfate	10 min	8.9 \pm 0.3
		24 h	9.0 \pm 0.6
RNase Sa	10% PEG-8000	10 min	17.3 \pm 0.6
		24 h	17.2 \pm 0.2
	1.5 M ammonium sulfate	10 min	8.9 \pm 0.3
		24 h	8.9 \pm 0.4
Ovalbumin	7.5% PEG-8000	10 min	22.3 \pm 2.1
		24 h	22.2 \pm 1.6
	2 M ammonium sulfate	10 min	16.8 \pm 0.8
		24 h	16.7 \pm 0.6
α -Lactalbumin	25% PEG-8000	10 min	7.8 \pm 1.2
		24 h	8.0 \pm 0.7
	1.5 M ammonium sulfate	10 min	22.2 \pm 0.1
		24 h	22.5 \pm 1.3
Fibrinogen	2.5% PEG-8000	10 min	4.5 \pm 0.2
		24 h	4.3 \pm 0.3
	0.7 M ammonium sulfate	10 min	3.3 \pm 0.2
		24 h	3.4 \pm 0.2

Experiments were performed in pH 7.0 5 mM citrate buffer.

Solubility curves in PEG

The solubility of seven proteins was determined as a function of PEG-8000. Fig. 3 shows the plot of log solubility as a function of PEG-8000 concentration. Equation 3 was fit to the data, and linear fit parameters are given in Table 2 (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 reported previously (19), PEG precipitation curves are linear over a wide range of PEG concentrations and can be extrapolated to zero concentration of PEG, 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, which makes sense given that they are present at high concentrations in their respective biological environments (human plasma (48) and bovine milk (49), respectively).

Solubility curves in ammonium sulfate

The solubility of the proteins was determined as a function of ammonium sulfate concentration. Fig. 4 shows the plot of log solubility versus ammonium sulfate concentration. The data

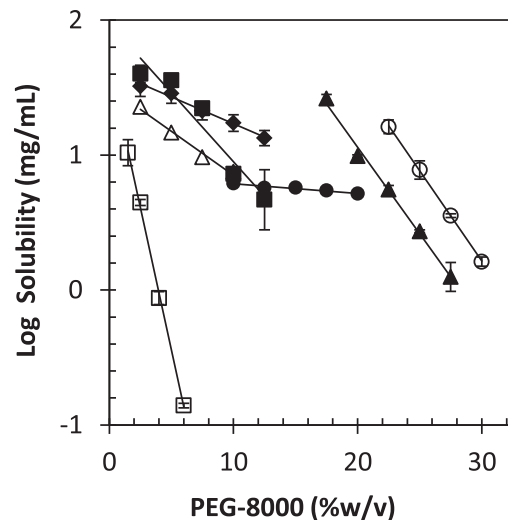


FIGURE 3 Solubility of several proteins in PEG-8000. The solubility of RNase Sa (solid diamonds), α -chymotrypsin (solid circles), lysozyme (open triangles), human serum albumin (solid triangles), ovalbumin (solid squares), α -lactalbumin (open circles), and fibrinogen (open squares) was measured at room temperature (25°C) and in pH 7.0 citrate. Equation 2 was fit to the data, and the fitted parameters can be found in Table 2.

were similarly fit by Eq. 3, and the best-fit values are given in Table 2 (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 six proteins, β varies only ~ 4 -fold from -0.92 for α -chymotrypsin to -4.13 for fibrinogen compared with β -values for PEG-8000, which vary over a 40-fold range. Clearly, the dependence of solubility on precipitant 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 fivefold lower than the next-closest β -value of -0.92 . This suggests that the ability of ammonium sulfate to precipitate α -lactalbumin is reduced in comparison with the other six 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 for PEG-8000. Of interest, α -lactalbumin, which was predicted to have the highest solubility by PEG-8000, is now predicted to have the lowest solubility.

DISCUSSION

Comparing protein solubility in PEG-8000 and ammonium sulfate

To compare the solubility measured using PEG-8000 with that measured using ammonium sulfate, we evaluated the

TABLE 2 Parameters obtained for PEG-8000 and ammonium sulfate solubility curves

Protein [‡]	PEG-8000 ^{1,§}			Ammonium sulfate ^{1,¶}		
	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

¹Equation 2 was fit to the data from Figs. 3 and 4, respectively.

[‡]Protein concentration is expressed in units of mg/mL.

[§]PEG concentration is expressed in units of % w/v.

[¶]Ammonium sulfate concentration is expressed in units of M.

log S_0 -values from the two fits. Fig. 5 shows the plot of log S_0^* obtained with ammonium sulfate (log S_0^* (NH₄)₂SO₄) versus log S_0 obtained with PEG-8000 (log S_0 PEG). A remarkably strong correlation between the solubility results for ammonium sulfate and those for PEG-8000 is seen. This suggests that log S_0^* (NH₄)₂SO₄ is a parameter that is related to protein solubility in the absence of precipitant. Because 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^* (NH₄)₂SO₄ suggests that log S_0 (NH₄)₂SO₄ can be used qualitatively to determine differences in solubility.

The solubility of α -lactalbumin warrants further discussion. In the case of PEG precipitation, α -lactalbumin is predicted to have the highest solubility of the proteins used in this study. This is not surprising given that α -lactalbumin is present in high concentrations in bovine milk (49) and

the fact that we are able 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 fivefold lower than the next-closest slope in ammonium sulfate. This suggests that ammonium sulfate is not as effective as a precipitant for α -lactalbumin as it is for the other proteins. This may be due in part 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 3, column 9). We previously suggested

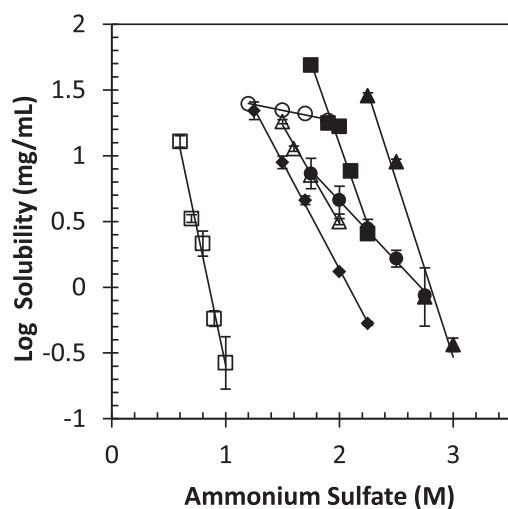


FIGURE 4 Solubility of several proteins in ammonium sulfate. The solubility of RNase Sa (*solid diamonds*), α -chymotrypsin (*solid circles*), lysozyme (*open triangles*), human serum albumin (*solid triangles*), ovalbumin (*solid squares*), α -lactalbumin (*open circles*), and fibrinogen (*open squares*) was measured at room temperature (25°C) and in pH 7.0 citrate. Equation 2 was fit to the data, and the fitted parameters can be found in Table 2.

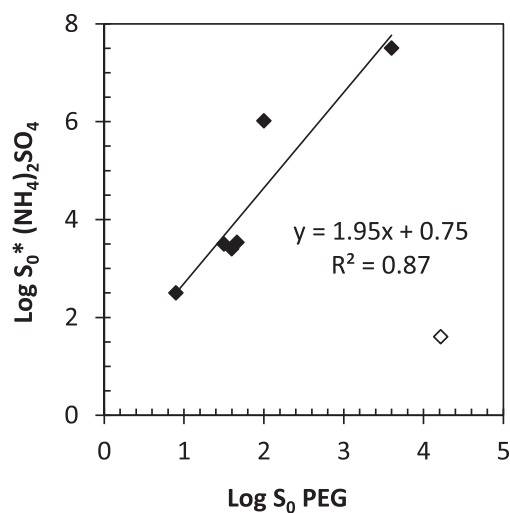


FIGURE 5 Comparison of the solubility data obtained in ammonium sulfate and PEG-8000. Log S_0 -values obtained from PEG-8000 precipitations are plotted against Log S_0^* -values from ammonium sulfate precipitations for all of the proteins. The data correlate strongly, suggesting a relationship between solubility results obtained with PEG-8000 and ammonium sulfate. α -Lactalbumin is shown as an open diamond and is excluded from the fit.

TABLE 3 Protein properties and surface properties used for correlations

Protein	Molecular mass (kDa)	Amino acids	pI*	Charge*	Absolute charge*	Fraction of ASA [†]				
						Nonpolar	Polar	Charged	Positive	Negative
A-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 -Lactalbumin	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

*pI and charge were calculated at pH 7 using Protein Calculator (54).

[†]Fractions of nonpolar, polar, charged, positively charged, and negatively charged were calculated as fractions of total accessible surface area.

that ammonium sulfate may underestimate the contribution of charged surface residues to protein solubility (16). This likely is related to the mechanism by which ammonium sulfate lowers protein solubility (i.e., increasing surface tension and competing for waters with the protein surface). The high level 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 amino and guanidino 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-out curve of α -lactalbumin, the ammonium sulfate data were fit without α -lactalbumin in subsequent correlations.

In an attempt to determine the intrinsic factors that influence protein solubility, we looked at several intrinsic protein properties and examined them with respect to protein solubility by comparing them with $\log S_0$ -values obtained in this study. We looked at fundamental properties of the protein, such as size (molecular mass) 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 was 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 mass and net charge with solubility measurements

To investigate the contribution of intrinsic factors to protein solubility (see Table 3, columns 1–6), we plotted $\log S_0$ -values for PEG-8000 and ammonium sulfate versus molecular mass (Fig. 6, A and B), net charge (Fig. 6, C and D), and the absolute value of the net charge (Fig. 6, E and F). Linear fits were made to the data, and R^2 values are given. Because the mechanism of PEG precipitation is related to the excluded volume, solubility may increase with protein size or molecular mass; however, no correlation

with molecular mass was observed. In general, the solubility of a given protein is at a minimum near the isoelectric point (pI) and increases with the absolute value of the net charge (13,50). To assess whether net charge plays a role in determining the solubility of a group of proteins, we plotted the net charge and absolute value of net charge versus $\log S_0$. A weak correlation was observed in all four cases. In the case of the absolute value of net charge, a weak positive correlation was observed, suggesting that with an increasing positive or negative net charge, protein solubility increases. For

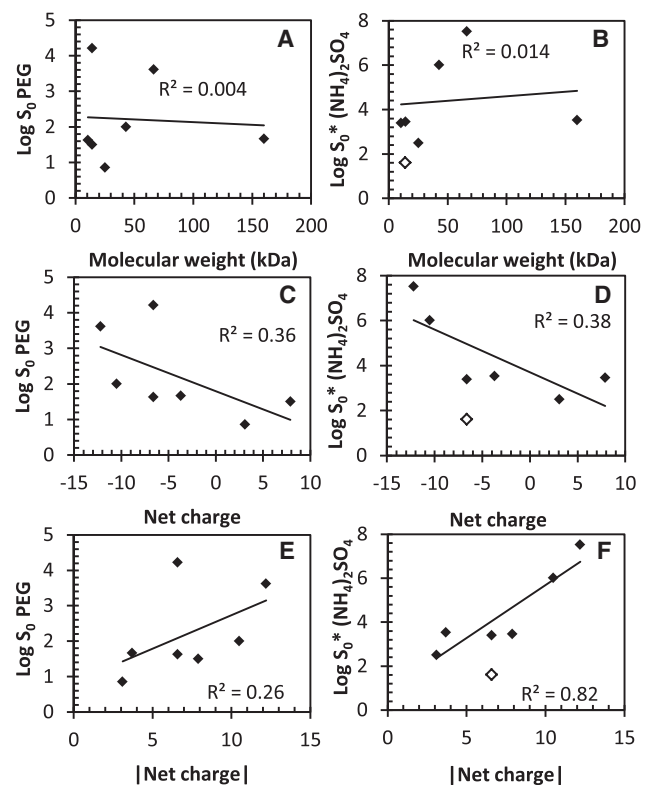


FIGURE 6 Correlation of molecular mass and net charge with PEG-8000 and ammonium sulfate solubility measurements. $\log S_0$ -values versus molecular mass (A and B), net charge (C and D), and absolute net charge (E and F) are shown. The lines and R^2 values are from linear least-squares fits. α -Lactalbumin is shown as open diamonds and is excluded from the ammonium sulfate fits.

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; however, more data points are required to determine whether 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, we investigated the correlation between solubility and the intrinsic properties of the surface of the proteins. The accessible surface areas (ASAs) of all atoms in the proteins were determined and the fractions that were polar, nonpolar, charged, positively charged, and negatively charged were calculated (see Table 3, columns 7–11). Fig. 7 depicts protein solubility as a function of fraction ASA that is polar or nonpolar for PEG-8000 (Fig. 7, A and B) and ammonium sulfate (Fig. 7, C and D). For PEG-8000, the correlation of the percentage of polar and nonpolar surface residues with protein solubility is very poor, although the correlation is positive for polar residues and negative for nonpolar residues, as might be predicted. This suggests that the surface polarity makes a minimal contribution to protein solubility in PEG-8000. For ammonium sulfate, a better correlation was observed, but the correlation with polar and nonpolar surface residues is negative and positive, respectively, which is the opposite of what we would have expected and what was observed in PEG-8000.

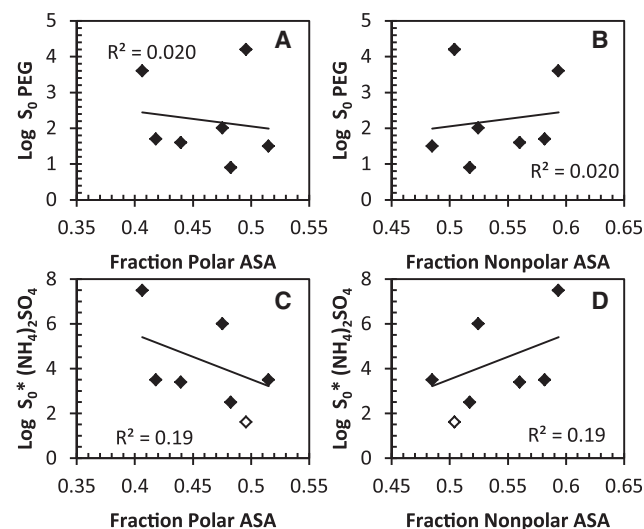


FIGURE 7 Correlation of fractions of polar and nonpolar ASAs with PEG-8000 and ammonium sulfate solubility measurements. The ASA for all atoms was calculated using PDB files and either pfls (31) or NACCESS (32). Carbon and sulfur atoms are considered nonpolar, and nitrogen and oxygen atoms are considered polar. Log S_0 -values versus fraction polar ASA (A and C) and fraction nonpolar ASA (B 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.

The contribution of the ASA that is charged, positively charged, and negatively charged was evaluated (see Table 3, columns 9–11). Fig. 8 depicts the correlations of PEG-8000 and ammonium sulfate with the fraction of charged (Fig. 8, A and B), positively charged (Fig. 8, C and D), and negatively charged (Fig. 8, 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 the fraction of positively charged ASA was not observed for either PEG-8000 or ammonium sulfate; however, a very strong correlation was observed between solubility and the 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 do any of the other 18 amino acids (16). To understand the difference in contribution to protein solubility of negative versus positive

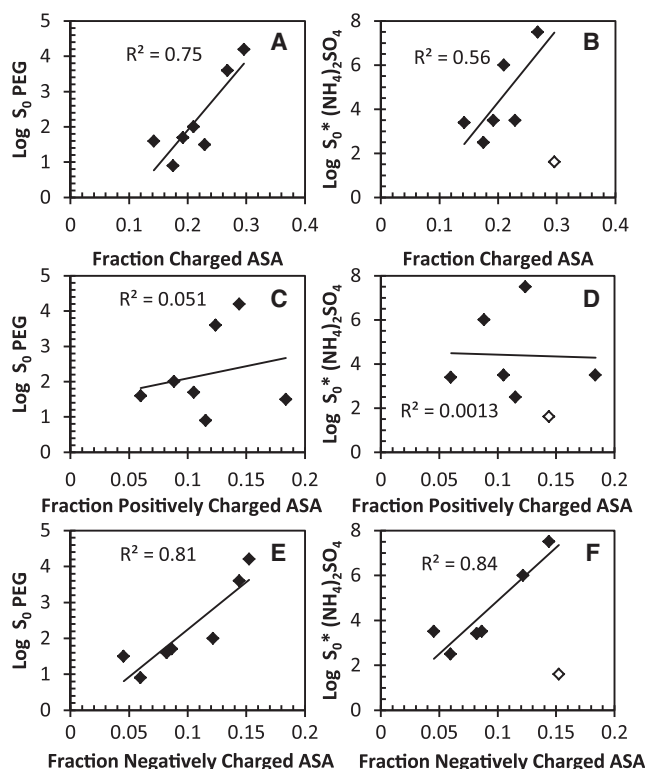


FIGURE 8 Correlation of the fraction of ASA that is charged (A and B), positively charged (C and D), and negatively charged (E and F) at pH 7.0 with PEG-8000 and ammonium sulfate solubility measurements. The ASA for all atoms was calculated using PDB files and either pfls (31) or NACCESS (32). The oxygen 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.

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. In studies on ions in solution, Collins (26,51–53) described a Hofmeister series dependence for hydration of ions in solution. Collins showed that kosmotropes are highly hydrated and bind water more tightly than water binds itself, whereas chaotropes bind water more weakly than water binds itself and remain largely unhydrated in solution. Therefore, the differential contribution to the solubility of negative and positive groups on the protein surface appears to be due to the differential hydration of the carboxylates that bind water tightly and the amino and guanidino groups that bind water weakly.

CONCLUSIONS

In this work, we found that negative surface charge had the strongest correlation with increased protein solubility, as measured by means of ammonium sulfate and PEG-8000 precipitation experiments. This is best explained by the strong water-binding properties of glutamic and aspartic acid (26). No correlation between solubility and positive surface charge was seen. We observed no correlation when we investigated the surface of the protein with respect to polarity. While comparing the two precipitants, we found that ammonium sulfate markedly increased protein stability for all proteins in this study, whereas PEG could have slight stabilizing or destabilizing effects. We found a remarkable correlation between the solubility results obtained with PEG-8000 and ammonium sulfate, even though they employ different mechanisms to decrease solubility. This suggests that solubility experiments using these two precipitants are probing similar intrinsic properties of the protein, making the choice between precipitants largely one of convenience. Because PEG precipitations can yield a quantitative estimate of solubility in buffer alone, and ammonium sulfate can only determine comparative solubility, PEG is a better choice of precipitant when absolute solubility values are of interest.

SUPPORTING MATERIAL

Details about the materials and methods, and references (29,30,33–40), are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(12\)00282-2](http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00282-2).

We thank Gerald Grimsley and Saul Trevino for critical readings of the manuscript.

This work was supported by Amgen, the Robert A. Welch Foundation (BE-1281), and the National Institutes of Health (Molecular Biophysics Training Grant T32GM65088 to R.K.). V.S. contributed to this work during an international rotation at Texas A&M University. N.M. contributed to this work

as part of the Summer Research Experience for Undergraduates program funded by the National Science Foundation.

REFERENCES

1. Bagby, S., K. I. Tong, M. Ikura, V. D. Thomas, L. James, and S. Uli. 2001. Optimization of protein solubility and stability for protein nuclear magnetic resonance. *In Methods in Enzymology*. Academic Press, New York. 20–41.
2. Caldwell, G. W., D. M. Ritchie, ..., Z. Yan. 2001. The new pre-preclinical paradigm: compound optimization in early and late phase drug discovery. *Curr. Top. Med. Chem.* 1:353–366.
3. Riès-kautt, M., and A. Ducruix. 1997. Inferences drawn from physicochemical studies of crystallogenesis and precrystalline state. *In Methods in Enzymology*. C. W. Carter, Jr., editor. Academic Press, New York. 23–59.
4. Schein, C. H. 1993. Solubility and secretability. *Curr. Opin. Biotechnol.* 4:456–461.
5. Fowler, S. B., S. Poon, ..., J. Zurdo. 2005. Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin. *Proc. Natl. Acad. Sci. USA.* 102:10105–10110.
6. Ricci, M. S., and D. N. Brems. 2004. Common structural stability properties of 4-helical bundle cytokines: possible physiological and pharmaceutical consequences. *Curr. Pharm. Des.* 10:3901–3911.
7. Ahn, H. C., N. Juranić, ..., J. L. Markley. 2006. Three-dimensional structure of the water-insoluble protein crambin in dodecylphosphocholine micelles and its minimal solvent-exposed surface. *J. Am. Chem. Soc.* 128:4398–4404.
8. Peters, T. 1996. All About Albumin: Biochemistry, Genetics, and Medical Applications. Academic Press, San Diego, CA.
9. Evans, P., K. Wyatt, ..., C. Slingsby. 2004. The P23T cataract mutation causes loss of solubility of folded γ D-crystallin. *J. Mol. Biol.* 343: 435–444.
10. Pande, A., O. Annunziata, ..., J. Pande. 2005. Decrease in protein solubility and cataract formation caused by the Pro23 to Thr mutation in human γ D-crystallin. *Biochemistry.* 44:2491–2500.
11. Kim, W., and M. H. Hecht. 2006. Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's A β 2 peptide. *Proc. Natl. Acad. Sci. USA.* 103:15824–15829.
12. Kaytor, M. D., and S. T. Warren. 1999. Aberrant protein deposition and neurological disease. *J. Biol. Chem.* 274:37507–37510.
13. Cohn, E. J., and J. T. Edsall. 1943. Proteins, Amino Acids, and Peptides. Hafner Publishing Company, New York.
14. Arakawa, T., and S. N. Timasheff. 1985. Theory of protein solubility. *Methods Enzymol.* 114:49–77.
15. Middaugh, C. R., and D. B. Volkin. 1992. Protein solubility. *In Stability of Protein Pharmaceuticals*. T. J. Ahern and M. C. Manning, editors. Plenum Press, New York. 109–134.
16. Trevino, S. R., J. M. Scholtz, and C. N. Pace. 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.* 366:449–460.
17. Trevino, S. R., J. M. Scholtz, and C. N. Pace. 2008. Measuring and increasing protein solubility. *J. Pharm. Sci.* 97:4155–4166.
18. McPherson, A. 2004. Introduction to protein crystallization. *Methods.* 34:254–265.
19. Middaugh, C. R., W. A. Tisel, ..., A. Rosenberg. 1979. Determination of the apparent thermodynamic activities of saturated protein solutions. *J. Biol. Chem.* 254:367–370.
20. Shiau, K. S., and T. L. Chen. 1997. Initial protein concentration effects on precipitation by salt. *Biotechnol. Bioeng.* 53:202–206.
21. Annunziata, O., A. Payne, and Y. Wang. 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.* 130:13347–13352.
22. Przybycien, T. M., and J. E. Bailey. 1989. Aggregation kinetics in salt-induced protein precipitation. *AIChE J.* 35:1779–1790.
23. Saluja, A., S. Crampton, ..., Y. R. Gokarn. 2009. Anion binding mediated precipitation of a peptibody. *Pharm. Res.* 26:152–160.
24. Kumar, V., V. K. Sharma, and D. S. Kalonia. 2009. Effect of polyols on polyethylene glycol (PEG)-induced precipitation of proteins: Impact on solubility, stability and conformation. *Int. J. Pharm.* 366:38–43.
25. Atha, D. H., and K. C. Ingham. 1981. Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *J. Biol. Chem.* 256:12108–12117.
26. Collins, K. D. 1997. Charge density-dependent strength of hydration and biological structure. *Biophys. J.* 72:65–76.
27. Thomas, P. D., and K. A. Dill. 1993. Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. *Protein Sci.* 2:2050–2065.
28. Arakawa, T., and S. N. Timasheff. 1985. Mechanism of poly(ethylene glycol) interaction with proteins. *Biochemistry.* 24:6756–6762.
29. Pace, C. N., and J. M. Sholtz. 1997. Measuring the conformational stability of a protein. In *Protein Structure: A Practical Approach*. T. E. Creighton, editor. IRL Press, Oxford. 229–321.
30. Hebert, E. J., G. R. Grimsley, ..., C. N. Pace. 1997. Purification of ribonucleases Sa, Sa2, and Sa3 after expression in *Escherichia coli*. *Protein Expr. Purif.* 11:162–168.
31. Hebert, E. J., A. Giletto, ..., C. N. Pace. 1998. Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba, and T1. *Biochemistry.* 37:16192–16200.
32. Hubbard, S. J., and J. M. Thornton. 1993. NACCESS. Department of Biochemistry and Molecular Biology. University College London, London, UK.
33. Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.
34. Kollman, J. M., L. Pandi, ..., R. F. Doolittle. 2009. Crystal structure of human fibrinogen. *Biochemistry.* 48:3877–3886.
35. Bhattacharya, A. A., S. Curry, and N. P. Franks. 2000. Binding of the general anesthetics propofol and halothane to human serum albumin. High resolution crystal structures. *J. Biol. Chem.* 275:38731–38738.
36. LeBeau, A. M., P. Singh, ..., S. R. Denmeade. 2009. Prostate-specific antigen is a “chymotrypsin-like” serine protease with unique P1 substrate specificity. *Biochemistry.* 48:3490–3496.
37. Chrysina, E. D., K. Brew, and K. R. Acharya. 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.
38. Wang, J., M. Dauter, ..., Z. Dauter. 2007. Triclinic lysozyme at 0.65 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* 63:1254–1268.
39. Stein, P. E., A. G. W. Leslie, ..., R. W. Carrell. 1991. Crystal structure of uncleaved ovalbumin at 1.95 Å resolution. *J. Mol. Biol.* 221:941–959.
40. Sevcik, J., Z. Dauter, ..., K. S. Wilson. 1996. Ribonuclease from *Streptomyces aureofaciens* at atomic resolution. *Acta Crystallogr. D Biol. Crystallogr.* 52:327–344.
41. Tobler, S. A., N. E. Sherman, and E. J. Fernandez. 2000–2001. Tracking lysozyme unfolding during salt-induced precipitation with hydrogen exchange and mass spectrometry. *Biotechnol. Bioeng.* 71:194–207.
42. Baldwin, R. L. 1996. How Hofmeister ion interactions affect protein stability. *Biophys. J.* 71:2056–2063.
43. Creighton, T. E. 1993. *Proteins: Structures and Molecular Properties*. W. H. Freeman, New York.
44. Greene, L. H., J. A. Grobler, ..., K. Brew. 1999. Stability, activity and flexibility in α -lactalbumin. *Protein Eng.* 12:581–587.
45. Farruggia, B., G. García, ..., G. Picó. 1997. Destabilization of human serum albumin by polyethylene glycols studied by thermodynamical equilibrium and kinetic approaches. *Int. J. Biol. Macromol.* 20:43–51.
46. Donovan, J. W., and E. Mihalyi. 1974. Conformation of fibrinogen: calorimetric evidence for a three-nodular structure. *Proc. Natl. Acad. Sci. USA.* 71:4125–4128.
47. Feher, G., and Z. Kam. 1985. Nucleation and growth of protein crystals: general principles and assays. *Methods Enzymol.* 114:77–112.
48. Sugio, S., A. Kashima, ..., K. Kobayashi. 1999. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng.* 12:439–446.
49. Wickström, E., K. Persson Waller, ..., A. Sternesjö. 2010. Short communication: relationships between α -lactalbumin and quality traits in bulk milk. *J. Dairy Sci.* 93:4577–4581.
50. Tanford, C. 1961. *Physical Chemistry of Macromolecules*. John Wiley and Sons, New York.
51. Collins, K. D. 1995. Sticky ions in biological systems. *Proc. Natl. Acad. Sci. USA.* 92:5553–5557.
52. Collins, K. D. 2004. Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods.* 34:300–311.
53. Collins, K. D., G. W. Neilson, and J. E. Enderby. 2007. Ions in water: characterizing the forces that control chemical processes and biological structure. *Biophys. Chem.* 128:95–104.
54. Protein Calculator. <http://www.scripps.edu/~cdputnam/protcalc.html>.