

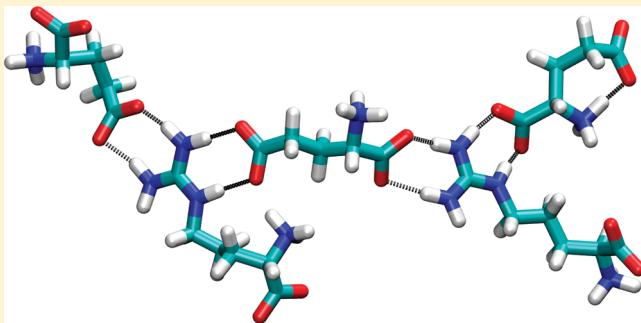
Understanding the Synergistic Effect of Arginine and Glutamic Acid Mixtures on Protein Solubility

Diwakar Shukla and Bernhardt L. Trout*

Department of Chemical Engineering, Massachusetts Institute of Technology, 50 Ames Street, E19-502b, Cambridge, Massachusetts 02139, United States

Supporting Information

ABSTRACT: Understanding protein solubility is a key part of physical chemistry. In particular, solution conditions can have a major effect, and the effect of multiple cosolutes is little understood. It has been shown that the simultaneous addition of L-arginine hydrochloride and L-glutamic acid enhances the maximum achievable solubility of several poorly soluble proteins up to 4–8 times (Golovanov et al., *J. Am. Chem. Soc.*, 2004, 126, 8933–8939) and reduces the intermolecular interactions between proteins. The observed solubility enhancement is negligible for arginine and glutamic acid solutions as compared to the equimolar mixtures. In this study, we have established the molecular mechanism behind this observed synergistic effect of arginine and glutamic acid mixtures using preferential interaction theory and molecular dynamics simulations of *Drosophila* Su(dx) protein (ww34). It was found that the protein solubility enhancement is related to the relative increase in the number of arginine and glutamic acid molecules around the protein in the equimolar mixtures due to additional hydrogen bonding interactions between the excipients on the surface of the protein when both excipients are present. The presence of these additional molecules around the protein leads to enhanced crowding, which suppresses the protein association. These results highlight the role of additive–additive interaction in tuning the protein–protein interactions. Furthermore, this study reports a unique behavior of additive solutions, where the presence of one additive in solution affects the concentration of another on the protein surface.



INTRODUCTION

Protein aggregation and unfolding in solution affects the long-term stability of proteins, the ability to concentrate proteins for structural studies, and storage at high concentration.¹ Additives (also known as cosolvents, cosolute, excipients, etc.) are typically added to aqueous protein solutions to enhance protein stability. L-Arginine is widely used as a formulation additive to inhibit protein aggregation, enhance protein refolding, and as an eluent for the recovery of proteins from an affinity chromatography column.^{2–4} Interestingly though, Golovanov et al.⁵ have shown that the simultaneous addition of L-Arg and L-Glu in equimolar ratios significantly enhances the maximum achievable concentration for three poorly soluble proteins with unrelated sequences (human MAGOH,⁶ fragment 1–153 of murine REF2-1⁷ and WW domains 3 and 4 from *Drosophila* Su(dx) protein (ww34)⁸). An equimolar mixture of L-Arg and L-Glu has been shown to enhance the solubility of these proteins by a factor between 4 and 8 at a concentration of 50 mM of each additive. Single-component solutions of both L-Arg and L-Glu only enhanced the solubility by less than 1.5 times at the same concentration. The authors speculated that the charged side chains of L-Arg and L-Glu interact with oppositely charged residues on the protein surface and cover the adjacently exposed hydrophobic residues, thereby inhibiting the protein aggregation

via reduced hydrophobic interactions. Valente et al.⁹ have experimentally measured the change in the osmotic second virial coefficient (B_{22}) for lysozyme in the presence of the equimolar mixture of L-Arg and L-Glu using self-interaction chromatography. It was found that the equimolar mixture was more effective at reducing the intermolecular attraction between proteins than either single-component solution. The equimolar mixture changed B_{22} values from -1.4×10^{-3} to -0.8×10^{-3} mol mL/g² at 0.2 M concentration of each component. However, this study only shows that the solubility enhancement is due to suppression of intermolecular interactions between proteins, but the molecular mechanism responsible for the suppression of protein–protein interaction has not been established. The exact mechanism is expected to be complicated due to the presence of a multitude of interactions (protein–additive, additive–additive, protein–water, additive–water, etc.) that could potentially affect protein–protein association.

Preferential interaction theory is a thermodynamic framework used to study the effect of additives on protein stability. The preferential interaction coefficient, Γ_{23} is a measure of the excess

Received: May 12, 2011

Revised: September 2, 2011

Published: September 06, 2011

number of additive molecules within the local region in the vicinity of the protein surface as compared to the bulk solution. It can be estimated experimentally via vapor pressure osmometry or dialysis/densimetry.^{10–12} The statistical mechanical definition of the preferential interaction coefficient,

$$\Gamma_{23} = \left\langle n_3^{\text{II}} - n_1^{\text{II}} \left(\frac{n_3^{\text{I}}}{n_1^{\text{I}}} \right) \right\rangle \quad (1)$$

has been used to obtain a theoretical estimate of Γ_{23} .^{13–15} In the above equation, n denotes the number of a specific molecule (subscript 1: water; 2: protein; and 3: additive) in a certain domain (superscript I for the bulk volume outside the vicinity of the protein and superscript II for the volume in the protein vicinity), and angle brackets denote an ensemble average. Timasheff and co-workers have examined the correlation between protein solubility and preferential interactions of proteins with additive solutions.^{16,17} The correlation shows that when the protein is preferentially hydrated (typically for additives with $\Gamma_{23} < 0$), protein solubility was decreased. The opposite behavior was observed when the protein is not preferentially hydrated ($\Gamma_{23} > 0$). When the protein is preferentially hydrated (additive does not interact favorably with the protein surface), the association of proteins is preferred because it reduces the total contact area between protein and additives. Similarly, when the protein is preferentially solvated by additive, the dissociated state is preferred because it will provide more contact surface area for the interaction between protein and additive. It can be seen that the preferential interaction coefficient is related to the protein–protein association in solution, which determines the protein solubility. Higher concentration of additives on the protein surface (positive gamma) has been shown to inhibit protein–protein association and increase protein solubility.^{16–18} Considering that the preferential interaction coefficients and protein solubility are intimately related, the molecular mechanism for the enhancement of local concentration around proteins in an equimolar mixture of L-Arg and L-Glu would also provide molecular insight into the protein solubility enhancement.

Molecular dynamics (MD) simulations of proteins in mixed solvents have been used to calculate preferential interaction coefficients. These simulations could also provide insight into the nature of specific protein–additive interactions, and help explain the molecular mechanism by which additives affect protein–protein interactions.^{4,19} Therefore, we have used preferential interaction theory along with MD simulations to identify the protein–solvent and intrasolvent interactions involved in the molecular mechanism responsible for solubility enhancement. In this study, we have performed MD simulations of the ww34 protein in 50, 100, and 200 mM solutions of L-Arg, L-Glu, and their equimolar mixtures. In addition, two simulations of ww34 in a 400 mM mixture and pure water were also performed. We have reported the theoretical preferential interaction coefficients for all the simulated systems. It was found that the preferential interaction of arginine and glutamic acid with the protein surface in an equimolar mixture of L-Arg and L-Glu was higher than the interaction in single-component solutions. From the mechanistic viewpoint, intrasolvent interactions between arginine and glutamic acid were found to be responsible for this stronger preferential interaction in mixtures. We have used the relationship between the preferential interaction coefficient and protein–protein association suppression to quantitatively relate the

Table 1. Setup of Simulation Systems^a

additive	C (mM)	length (ns)	L-Arg	L-Glu
L-Arg	50	100	12	0
L-Glu	50	100	0	12
Arg+Glu	50	50	12	12
L-Arg	100	50	24	0
L-Glu	100	50	0	24
Arg+Glu	100	50	24	24
L-Arg	200	50	48	0
L-Glu	200	50	0	48
Arg+Glu	200	100	48	48
Arg+Glu	400	100	96	96
NaCl	100	50	24	24

^a 50 mM equimolar mixture of L-Arg and L-Glu implies 50 mM concentration of each component. All systems contain 13333 water molecules and 100 mM NaCl (24 sodium and 24 chloride ions), which is the buffer used in experiments.

local concentration enhancement to the synergistic enhancement of protein solubility in equimolar mixtures of L-Arg and L-Glu.

METHODOLOGY

Simulation Setup. All simulations were performed using the NAMD²⁰ package, with the CHARMM22²¹ force field. The crystal structure of WW domains 3 and 4 from *Drosophila* Su(dx) protein (ww34) was taken from PDB 1TK7.⁸ The TIP3P water model was used.²² The pK_a values for the C-terminal, the N-terminal, and the side chain in an arginine molecule are 1.8, 9.0, and 12.5, respectively. The pK_a value for the side chain group in the glutamic acid is 4.1. The N-terminal and the arginine side chain are protonated, whereas the C-terminal and the glutamic acid side chain are deprotonated at the experimental pH 5.6. Therefore, force field parameters for arginine and glutamic acid were taken from the CHARMM force field with the protonated N terminal, and the side chain, and deprotonated C terminal. The parameters for the N and C terminal were taken from the CTER and NTER parameters available in CHARMM. All simulations were performed in the NpT ensemble with periodic boundary conditions, and full electrostatics were computed using the particle mesh Ewald (PME) method, with a grid spacing of 1 Å or less.²³ The pressure was maintained at 1 atm using the Langevin piston method, with a piston period of 200 fs, a damping time constant of 100 fs, and a piston temperature of 298 K. An integration step of 1 fs was used.²⁴ The initial size of the periodic rectangular box was set to (75 Å)³ in all of the simulations. To set up the simulation systems for various additive concentrations ranging from 50 mM to 400 mM, additive molecules and ions were randomly placed within the simulation box (while assuring no overlap), and subsequently overlapping water molecules were removed. The system was then equilibrated for 1 ns at constant pressure and temperature. All 11 systems described in Table 1 were simulated for a total of 750 ns.

Preferential Interaction Coefficient. MD simulations of protein in mixed solvents are used to measure the preferential binding parameter. The preferential binding parameter, Γ_{23} describes how the additive concentration changes when protein is added to the solution in order to keep the chemical potential of additive constant. The method of calculating preferential interaction parameters, based on a statistical mechanical method applied to an all-atom model with no adjustable parameters,

has been described in the literature.^{13,15} Preferential interaction coefficients (Γ_{23}) are defined by the following equation:^{3,4,14,15}

$$\Gamma_{23} = \rho_3(\infty) \int_0^\infty (g_3 - g_1) dV \quad (2)$$

where subscripts 1, 2, and 3 stand for water, protein, and additive, respectively, g_i is the radial distribution of the component i , and $\rho_3(\infty)$ denotes the density of the additive in the bulk region away from the protein surface. The integral extends from $r = 0$ (protein surface) to ∞ (bulk region). The variation of concentration as a function of the distance from the protein surface is used to calculate Γ_{23} as a function of the distance from the protein until it approaches a constant value. From the simulated trajectory, the number of arginine and water molecules as a function of distance from the protein surface can be obtained. The MD run is saved at periodic time intervals, and these saved frames are used to find Γ_{23} . The expression used to calculate the coefficient is

$$\Gamma_{23}(r, t) = n_3(r, t) - n_1(r, t) \left(\frac{n_3 - n_3(r, t)}{n_1 - n_1(r, t)} \right) \quad (3)$$

where n_3 is the total number of additive molecules, and n_1 is the total number of water molecules. The extent of the local domain around the protein in which $\Gamma_{23}(r, t)$ varies is found to be around 5–6 Å. The value we predict for the preferential interaction coefficient is the average of the instantaneous $\Gamma_{23}(t)$ values. The procedure described above can be used if more than one species is present in the solution to calculate the excess number of molecules of a particular species within the local domain of the protein.

Relationship between Preferential Interaction Coefficient and Protein Solubility. Shulgin and Ruckenstein have derived the relationship between the preferential interaction coefficient and protein solubility using the Kirkwood-Buff theory.²⁵

$$\ln\left(\frac{S_3}{S_1}\right) = \int_1^{x_1} \frac{(V_1 - V_2^\infty - \Gamma_{23}^{(c)} / c_3)}{x_1 V_1} dx_1 - \int_1^{x_1} \frac{(V_2^\infty + \Gamma_{23}^{(c)} / c_3) J_{11}}{V_1} dx_1 \quad (4)$$

where S_3 is the protein solubility in additive solution, S_1 is the solubility in additive-free water plus buffer, V_i is the partial molar volume of component i , V_2^∞ is the partial molar volume of protein at infinite dilution in a mixed solvent, and $J_{11} = \lim_{x_2 \rightarrow 0} ((\partial \ln \gamma_1) / (\partial x_1))_{x_2}$. $\Gamma_{23}^{(c)}$ is the preferential interaction coefficient measured in molar concentrations. It is related to the $\Gamma_{23}^{(m)}$ (expressed in molal concentrations) via the following equation:

$$\Gamma_{23}^{(c)} = (1 - c_3 V_3) \Gamma_{23}^{(m)} - c_3 V_2^\infty \quad (5)$$

If we assume that the partial molar volumes in eq 4 are composition independent, and $\Gamma_{23}^{(c)}$ is proportional to additive concentration, the eq 4 can be simplified to

$$\ln\left(\frac{S_3}{S_1}\right) = - \frac{\Gamma_{23}^{(m)} \ln(a_1)}{c_3 V_1} \quad (6)$$

In addition to the above assumptions, several other simplifications have been used to derive the above equation. The effect of buffer is taken into account indirectly during the calculation of preferential interaction coefficient. The activity coefficient of the protein in the additive solution is assumed to be equal to that at infinite dilution, and the solubility of the additive solution in the

solid phase is assumed to be negligible. These assumptions are valid for poorly soluble solids. The *ww34* protein investigated in this study is a poorly soluble protein. The effect of a mixture of two additives (A and B) on protein solubility is calculated by assuming that their effect is additive.

$$\begin{aligned} \ln\left(\frac{S_{3,A+B}}{S_1}\right) &= \ln\left(\frac{S_{3,A}}{S_{3,B}}\right) + \ln\left(\frac{S_{3,B}}{S_1}\right) \\ &= - \frac{\Gamma_{23,A}^{(m)} \ln(a_1^A)}{c_{3,A} V_1} - \frac{\Gamma_{23,B}^{(m)} \ln(a_1^B)}{c_{3,B} V_1} \end{aligned} \quad (7)$$

The concentration dependence of the activity coefficient of water was calculated from the experimental osmotic coefficient data using the following expression:²⁶

$$\ln(a_1) = - \phi M_1 m_3 \nu \quad (8)$$

The experimental activity coefficient data for L-Arg and L-Glu was taken from the literature.^{27–29}

RESULTS AND DISCUSSION

Preferential Interaction. Preferential interaction coefficients as a function of distance from the protein surface for 100 mM arginine, glutamic acid, and an equimolar mixture of L-Arg and L-Glu are shown in Figure 1. It can be seen that the size of the local domain (region where the concentration varies significantly as compared to the bulk solution) is around 0.6–0.7 nm, which is consistent with the values reported in the literature.^{13,15} The variation of Γ_{23} as a function of simulation time for the simulation of *ww34* in 100 mM arginine is shown in Figure 2. The required simulation time depends on the concentration, size and diffusivity of the additive. It can be seen that preferential interaction coefficient values converge within 50 ns for 100 mM arginine solution. We have performed 100 ns simulations for systems with 50 mM additive concentration. All the simulations provided a converged estimate of the preferential interaction coefficient.

Γ_{23} values as a function of concentration for arginine and glutamic acid in different additive solutions is shown in Figure 3. The preferential interaction data for all ions in solution is included in Table 2. Γ_{23} values for arginine are positive and increase with increasing concentration for both single-component solutions and equimolar mixtures of L-Arg and L-Glu. However, the values for arginine in the mixtures are higher than the corresponding values for single-component arginine solutions. Γ_{23} values for glutamic acid in the single-component solutions are close to zero (between -1 and 1) at all concentrations, indicating that glutamic acid does not interact strongly with the protein surface as compared to arginine. Arginine and glutamic acid should interact with different residues on the protein surface due to their oppositely charged side chains. It has been shown that arginine molecules preferentially interact with charged (both positive and negative) and polar residues (including protein backbone) via hydrogen bonding.^{3,4} Arginine can also interact with aromatic residues on the protein surface via cation-π interactions. Glutamic acid can interact with both positively and negatively charged residues; however, interactions with positively charged residues is expected to be dominant compared to interactions with negatively charged residues. The N-terminal amine group is the only positively charged group in glutamic acid, and it cannot form multiple hydrogen bonds with

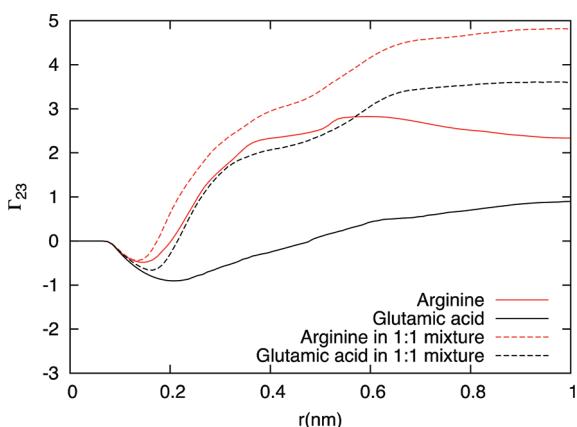


Figure 1. Variation of preferential interaction coefficient as a function of distance from the protein surface for the simulation of ww34 in 100 mM arginine, glutamic acid, and an equimolar mixture of L-Arg and L-Glu.

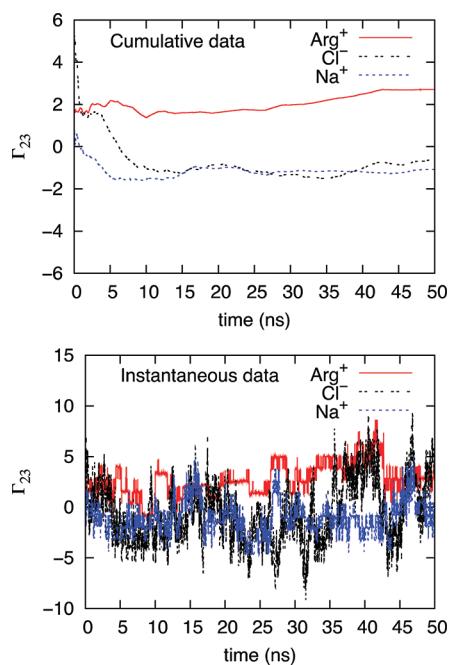


Figure 2. Convergence of the preferential interaction coefficient as a function of simulation time for the simulation of ww34 in 100 mM arginine solution.

the carboxylate groups of negatively charged glutamic and aspartic acid residues on the protein surface. The solvent accessible surface area for different types of residues in ww34 calculated from a 50 ns simulation of the protein in a water box shows that the protein surface is comprised of 30% charged (~15% positive and 15% negative), 35% polar, 10% apolar, 10% aliphatic, and 15% aromatic residues. Therefore, arginine can favorably interact with ~80% of the protein surface, whereas glutamic acid can only interact with less than <50% of the surface. The composition of the protein surface explains the difference between the observed preferential interaction coefficient values for arginine and glutamic acid in the single-component solutions. However, the corresponding values for glutamic acid in the

Table 2. Preferential Interaction Coefficient Values for the Protein ww34 and the Individual Ions in the Aqueous Additive Solution^a

additive	C (mM)	Γ_{23} Arg ⁺	Γ_{23} Glu ⁻	Γ_{23} Na ⁺	Γ_{23} Cl ⁻
L-Arg	50	1.3		1.4	2.1
L-Glu	50		0.9	0.1	0.0
Arg+Glu	50	2.5	2.0	-1.3	-0.3
L-Arg	100	2.7		-1.0	-0.5
L-Glu	100		0.5	0.5	-0.2
Arg+Glu	100	4.5	3.5	-1.0	-0.4
L-Arg	200	7.0		-1.3	-0.3
L-Glu	200		0.8	-0.1	-0.7
Arg+Glu	200	10.5	3.4	-2.1	0.9
Arg+Glu	400	8.8	3.6	-2.8	-1.4

^aThe error bars are on the order of ± 1 .

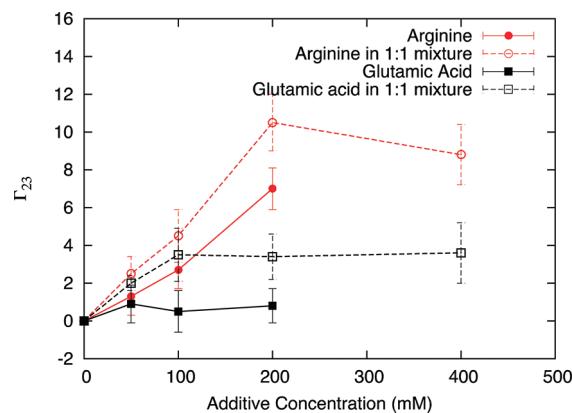


Figure 3. Preferential interaction coefficient values versus concentration for arginine and glutamic acid ions in both single-component additive solutions and equimolar mixtures of arginine and glutamic acid.

equimolar mixtures are higher than the individual component solutions. This observation suggests that there is a synergistic enhancement of the local concentration for both arginine and glutamic acid around the protein surface in the Arg+Glu mixture.

At a low concentration of additive, there are many unoccupied sites on the protein surface for the preferential binding of an additive. (For example, only six arginine molecules are directly hydrogen bonded to the protein surface at 200 mM concentration.) In other words, the presence of one additive on the protein surface does not lead to a reduction in the available binding sites on the surface for the interaction of the other additive. Therefore, when considering only direct protein–additive interactions, the total number of arginine and glutamic acid molecules on the protein surface should be the same as the number in the individual component solutions. The number of directly hydrogen-bonded arginine and glutamic acid molecules in the single-component solution and equimolar mixture are found to be same at all concentrations. However, this observation only accounts for the concentration of one component not being reduced by the presence of the other component. The preferential interaction data described above shows that the local domain concentrations for the mixture are much higher than that for the individual component solutions. It is expected that high local domain concentrations in mixtures should be the result

of intrasolvent interactions between arginine and glutamic acid. Therefore, we have measured the number of arginine and glutamic acid molecules interacting with a particular amino acid on the protein surface to identify sites where additional additive molecules are present.

In order to calculate the number of additive molecules around each amino acid, the following procedure is used: For every additive molecule in solution, the nearest amino acid is determined by computing the distance between the center of mass of an additive molecule and the van der Waals surface of the protein. This data is collected for every simulation frame (1 ps) and then averaged over the entire trajectory. The definition of the local domain is the same as that used for the estimation of the preferential interaction coefficient. The average number of arginine and glutamic acid molecules (within the local domain) coordinated with a particular amino acid is shown in Figure 4

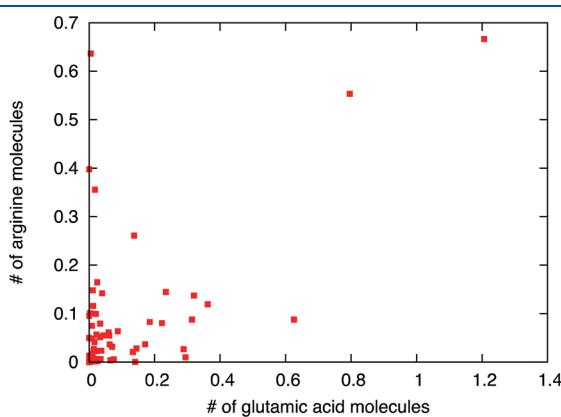


Figure 4. Number of arginine and glutamic acid molecules bound to particular amino acids in ww34 obtained from the simulation of ww34 in a 100 mM equimolar mixture of L-Arg and L-Glu.

(for 100 mM mixture). It can be seen that most of the amino acids are interacting with either arginine or glutamic acid (as indicated by the points close to the x and y axes). However, there are few sites that are occupied by both arginine and glutamic acid molecules. These arginine and glutamic acid binding motifs were identified. It was found that the arginine and glutamic acid molecules around these residues were interacting with each other via multiple hydrogen bonds as shown in Figure 5. In Figure 5a, it can be seen that arginine molecules can form hydrogen bonds with the ASP41 and GLU40 residues on the protein surface, and the guanidinium group on its side chain acts as a favorable site for the interaction with glutamic acid molecules in solution. Similarly, in Figure 5b, two glutamic acid molecules simultaneously interact with ARG71 and ARG79, and the free carboxyl groups interact with the arginine molecule in the solution. In the crystal structure of the arginine glutamate salt,³⁰ it has been observed that arginine and glutamic acid molecules interact via hydrogen bonds between their charged groups. In the pH range 2–9, arginine and glutamic acid have three charged groups each, which lead to a variety of possible interactions between these molecules.

Proposed Hypothesis. On the basis of the observed interactions, we hypothesize that interactions between arginine and glutamic acid are responsible for enhancement of the local concentration in the vicinity of proteins. In particular, when a glutamic acid molecule is interacting with a positively charged residue on the protein surface via hydrogen bonding, the other carboxylate group in the molecule is available for interaction with an arginine molecule in solution. Therefore, due to the presence of two carboxylate groups in glutamic acid, it effectively converts a positively charged site on the protein surface into a negatively charged site. Similarly, if an arginine molecule interacts with an aromatic or charged residue on the protein surface, the other charged groups can interact with other arginine or glutamic acid molecules in solution. In summary, we propose that arginine and glutamic acid molecules interacting with the protein surface act as

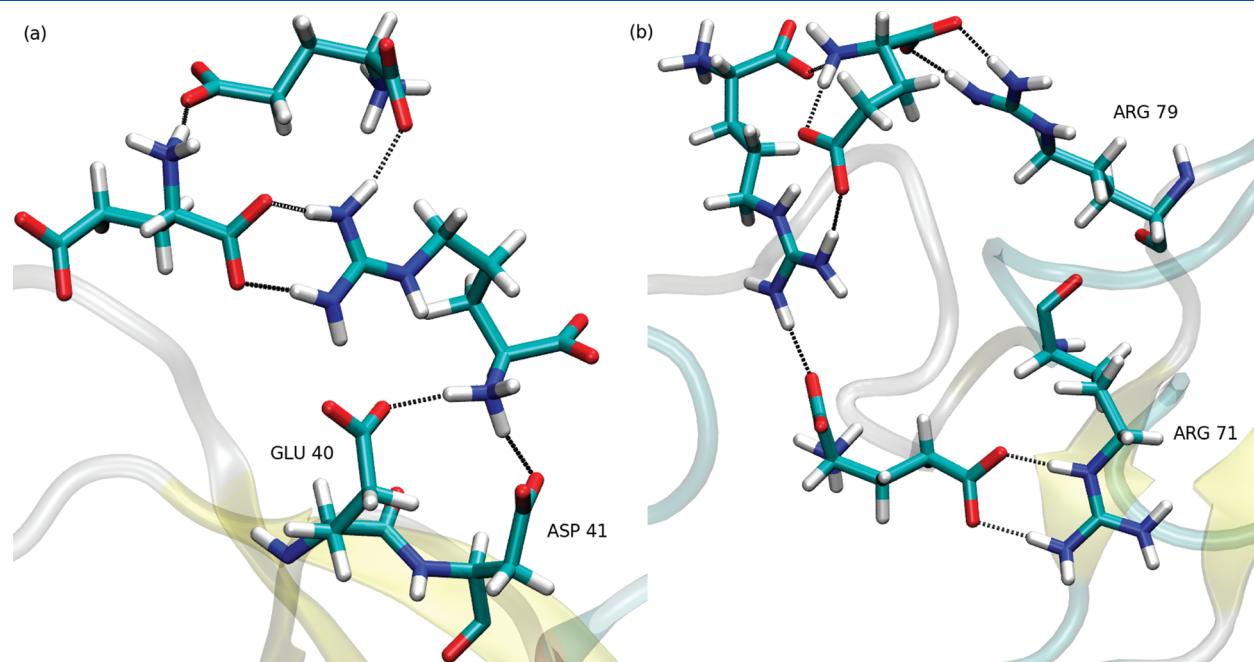


Figure 5. MD snapshots of clusters of arginine and glutamic acid molecules on the protein surface. (a) Arginine molecules acting as a bridge between the protein and glutamic acid molecules in solution. (b) Glutamic acid molecules acting as a bridge between the protein surface and arginine molecules in solution.

Table 3. The Number of Hydrogen Bonds Formed between the Protein and Additives, the Predicted Relative Aggregation Rate (k_A/k_0) Due to Crowding, and Solubilization Factors (S_3/S_1) Calculated Using the Kirkwood–Buff Theory of Solutions

additive	C (mM)	H-bonds protein–Arg	H-bonds protein–Glu	bridge H-bonds	k_A/k_0	S_3/S_1
L-Arg	50	5.0			0.78	2.7
L-Glu	50		2.0		0.90	2.6
Arg+Glu	50	6.6	2.7	2.1	0.51	7.7
L-Arg	100	8.6			0.60	2.4
L-Glu	100		4.4		0.91	2.6
Arg+Glu	100	9.0	6.3	13.2	0.28	54.3
L-Arg	200	15.4			0.27	21.6
L-Glu	200		6.9		0.84	1.9
Arg+Glu	200	17.0	8.8	15.6	0.08	628
Arg+Glu	400	18.4	16.0	28.5	0.08	77.2

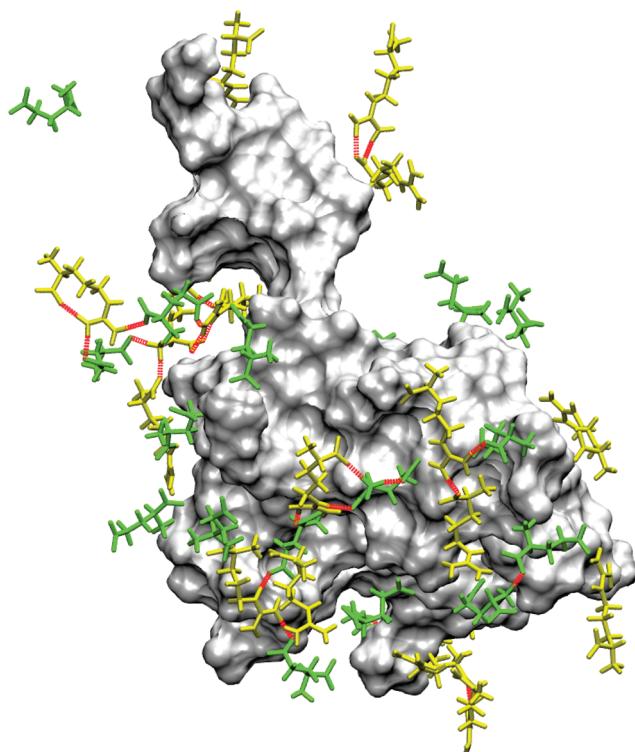


Figure 6. Hydrogen-bonded clusters of L-Arg and L-Glu molecules around the protein surface. Arginine molecules are shown in yellow, glutamic acid molecules are in green, and hydrogen bonds between additives are in red.

additional interaction sites for glutamic acid and arginine molecules in the solution, which could explain the synergistic enhancement of the local concentration in the equimolar mixture of L-Arg and L-Glu. The proposed hypothesis could be verified by calculating the number of arginine (glutamic acid) molecules that are hydrogen bonded to the protein surface with glutamic acid (arginine) molecules acting as a bridge between them.

Hydrogen Bonds. The data for the number of hydrogen bonds between the protein surface and an additive, where the

other additive acts a bridge, is reported in Table 3. It clearly supports the presence of such clusters of additives on the protein surface. These hydrogen-bonded clusters near the protein surface are shown in Figure 6. Now that we have established that such bridged additive molecular clusters exist, we need to establish the number of molecules attached to the protein surface via a bridged molecule and compare it to the number of additional molecules on the protein surface in mixtures as compared to the single-component solution. For this purpose, we have analyzed in detail the preferential interaction and the hydrogen bond data for additive solutions at a 200 mM concentration for both single-component and equimolar solutions. The difference between the preferential interaction data for arginine in the single-component solution and the mixture is 3.5, which implies that, on average, there are approximately 3.5 additional arginine molecules on the protein surface in the mixture as compared to the single-component solution. The corresponding difference for glutamic acid is 1.6. However, the number of direct hydrogen bonds formed between arginine and the protein surface remains the same in both the mixture and the single-component solutions. A similar conclusion can be drawn about the number of directly interacting glutamic acid molecules. The exact number of arginine molecules directly hydrogen bonded to the protein surface was found to be 6.2 in the single-component solution and 5.8 in the equimolar mixture of L-Arg and L-Glu. The number of glutamic acid molecules directly hydrogen bonded to the protein surface was 2.3 in single-component solution and 2.1 in the mixture. The number of bridged hydrogen bond is ~15, which, if we assume that an arginine (or glutamic acid) molecule is hydrogen bonded to a glutamic acid (or arginine) molecule interacting with the protein surface via 2–3 hydrogen bonds, can easily explain the presence of an additional 3.5 arginine and 1.6 glutamic acid molecules on the protein surface in the mixture. We have also counted the exact number of arginine molecules that are hydrogen bonded to the protein surface with a glutamic acid molecule acting as a bridge between them. Similar counting of molecules was also done for glutamic acid. It was found that, on average, 1.7 glutamic acid molecules were hydrogen bonded to the protein surface with an arginine acting as a bridge between them. This value matches exactly with the number of additional glutamic acid molecules on the protein surface in the mixture. Similarly, the number of arginine molecules hydrogen bonded to the protein surface with a glutamic acid molecule acting as a bridge between them was found to be 2.0, which also matches well with the additional number of arginine molecules (3.5) in the L-Arg and L-Glu mixture. The difference arises due to the strict distance and angle cutoff (0.24 nm, 30°) used for the calculation of the number of hydrogen bonds. During the course of simulation, these hydrogen bonds are observed to break and form frequently with an average lifetime of less than 10 ps. Although the arginine molecules may be in the vicinity of the glutamic acid molecules for the entire simulation, they will not be hydrogen bonded for the total simulation time. Therefore, we have calculated the number of contacts between arginine molecules and the glutamic acid molecules hydrogen bonded to the protein surface. The number of contacts was found to be 3.1, and it matches well with the additional number of arginine molecules (3.5) on the protein surface in the 200 mM mixture of L-Arg and L-Glu.

It can be seen that the intrasolvent interactions are responsible for the local domain additive concentration enhancement in the equimolar mixtures of L-Arg and L-Glu. However, the connection

between the local domain additive concentration enhancement and the synergistic protein solubility enhancement still needs to be established.

Protein Solubility Prediction. Shulgin and Ruckenstein have derived a relationship between the preferential interaction coefficient and protein solubility using the Kirkwood–Buff theory.²⁵ A brief derivation was included in the Methodology section for the sake of completeness. Using the relationship between the protein solubility and preferential interaction coefficient, the solubilization factor (S_3/S_1), which is the ratio of solubility in an additive solution and the solubility in the buffer-only solution, can be estimated. Golovanov et al. reported the experimental solubilization factors for ww34 in 50 mM arginine, 50 mM glutamic acid and 50 mM L-Arg and L-Glu mixture. The calculated solubilization factor is ~ 8 , which is higher than the experimental value of ~ 4 . The solubilization factor calculations are extremely sensitive to change in the Γ_{23} values. For example, a decrease in Γ_{23} of 1 would reduce the predicted solubilization factor to 3. Therefore, the experimental value at 50 mM lies within the error bars on the predicted solubilization factor. However, the key point is that the predicted solubilization factors capture the synergistic enhancement of protein solubility in the equimolar mixture.

The predicted solubilization factor for 200 mM mixture is 628, which implies that the solubility of the ww34 would be 628 times more than its solubility in the buffer-only solution. The theoretical framework used for predicting solubilization factors only includes the effect of preferential accumulation of additives on the protein surface leading to significant reduction in protein–protein association at high additive concentration. However, the theory does not include the detrimental effect of additives (especially those with $\Gamma_{23} \gg 0$) on the conformational stability of the protein. Therefore, the predicted values for concentrations greater than 200 mM would decrease if the negative contribution to the conformational stability were included. The solubilization factor for the 400 mM mixture is reduced to 77. This sharp decrease in the solubilization factor is due to the decrease in the concentration of additives on the protein surface as compared to the 200 mM mixture. The preferential interaction coefficient values for both L-Arg and L-Glu in the equimolar mixture are the same at both concentrations, which implies that, although the excess number of molecules on the protein surface are the same in both solutions, the local concentration of additives as compared to the bulk concentration is reduced to almost half in the 400 mM solution as compared to the 200 mM solution. It can be seen from eq 6 that if the Γ_{23} value is kept constant and the concentration is doubled, the solubilization factor would reduce by a factor of e^2 (~ 7.4), which is similar to the reduction in the solubilization factor observed at 400 mM concentration.

The synergistic accumulation of arginine and glutamic acid molecules on the protein surface should also lead to a reduction in protein–protein association due to the crowding around the proteins. Therefore, we have measured the relative association suppression in the mixture of L-Arg and L-Glu using a model system to support our hypothesis that the synergistic enhancement of protein solubility in the equimolar mixture is due to the enhanced crowding of arginine and glutamic acid molecules on the protein surface.

Aggregation Suppression Due to Crowding. The aggregation suppression due to the presence of crowders with a radius equivalent to the radius of gyration for arginine and glutamic acid can be predicted by the procedure outlined in our earlier work.^{4,18}

The effect of crowders is estimated for a model system of two spherical proteins, which is a good approximation for ww34. The preferential interaction data for the crowders is taken from the MD simulation results. The exact procedure used for calculations is included in the Supporting Information. The predicted relative aggregation rate (k_A/k_0) data reported in Table 3 shows that there is a synergistic enhancement of protein association suppression in the L-Arg and L-Glu mixtures. At 50 mM, the aqueous arginine solution is predicted to reduce the rate of association by a factor of 0.78, while the glutamic acid solution is predicted to reduce the rate of association by a factor of 0.9. However, the Arg + Glu mixture is predicted to reduce the rate by a factor of 0.5. Similar but more pronounced trends are observed at higher concentrations with relative aggregation rate reducing to 0.08 for a 200 mM mixture. However, the relative aggregation rate for a 100 mM mixture (100 mM of each component) is comparable to the relative aggregation suppression for a 200 mM arginine solution. Valente et al. have also reported that both the arginine solution (100 mM) and the equimolar mixture of L-Arg and L-Glu (50 mM each) reduce the second virial coefficient of lysozyme by the same amount (3×10^{-4} mol mL/g²). Therefore, it can be seen that the difference between the performance of equimolar mixture and arginine solution decreases at high concentration. Furthermore, the predicted aggregation suppression (0.08) for both 200 mM and 400 mM mixtures are the same, which indicates that there must be a value of concentration at which the synergistic effect levels off. At high concentrations, the favorably interacting sites on the protein surface become occupied. Therefore, the excess concentration of additives on the protein surface, and, correspondingly, their preferential interaction coefficient values, start to decrease due to the exclusion of additives. Similar exclusion at high concentration has been reported for aqueous arginine solutions.^{10,14} Furthermore, the arginine and glutamic acid molecules in the bulk region (away from the protein surface) begin to provide interaction sites for any additional molecules added to the solution, thereby reducing the excess concentration on the surface.

In this study, we have shown that the interaction of arginine and glutamic acid residues with the protein surface and with each other explains the enhancement in protein solubility. However, it is not clear whether this synergistic enhancement of protein stability would be observed for the heat-induced aggregation of proteins as well. Most of the experimental aggregation suppression studies are done at elevated temperature. Arginine has been shown to inhibit heat-induced protein aggregation by suppressing the protein association without any concomitant effect on the conformational stability of the protein. NaGlu has been shown to stabilize proteins by enhancing the conformational stability of proteins due to the preferential exclusion of glutamic acid from the protein surface. Therefore, glutamic acid affects the barrier for unfolding of proteins, and arginine affects the barrier for the association of native or partially unfolded proteins. Therefore, the exact mechanism for the suppression of protein aggregation is expected to be a combination of these two effects. The solubility enhancement is observed due to enhancement in crowding around the charged sites on the protein surface. In heat-induced aggregation, the proteins are in a partially or completely unfolded state with exposed hydrophobic residues. Therefore, the ability of the equimolar mixture of L-Arg and L-Glu to reduce their aggregation would depend on their ability to mask all the exposed hydrophobic patches. If there are contiguous patches of exposed hydrophobic residues, the crowding around

charged sites alone will not be able to mask these patches. Therefore, it is likely that the synergistic may not be observed for heat-induced aggregation of proteins. Furthermore, the effectiveness of the equimolar mixture of L-Arg and L-Glu would also depend on the amino acid composition of the protein surface. We have calculated the surface-exposed amino acid composition of the proteins studied by Golovanov et al. The charged and aromatic amino acid residues that interact strongly with the L-Arg and L-Glu comprise between 44 and 58% (Fragment 1–153 of murine REF2–1: 44%; WW34: 51%; Y14: 55% and Human MagOH: 58%) of the solvent accessible area of these proteins. Similarly, L-Arg and L-Glu mixtures are effective at reducing the self-interaction of lysozyme in solution. For lysozyme, the charged and aromatic amino acids comprise 51% of the total solvent accessible area. This observation supports our conclusion that the mixture of L-Arg and L-Glu would be more effective at enhancing solubility of proteins with charged and aromatic surface-exposed amino acids.

CONCLUSIONS

In this study, we have shown that equimolar mixtures of L-arginine and L-glutamic acid enhance protein solubility due to the synergistic enhancement of the local concentration of additives near the protein surface compared to the single-component solution. The molecular mechanism responsible for this concentration enhancement involves interactions between additives on the protein surface and those in the solution. The crowding due to the presence of an enhanced number of molecules on the protein surface explains the suppression of protein–protein interactions in the equimolar mixture of L-Arg and L-Glu, thus highlighting the role of intrasolvent interactions in suppressing protein–protein interactions. The behavior of L-Arg and L-Glu mixture is similar to that of urea and trimethyl amine oxide (TMAO) mixtures, where TMAO affects the concentration of urea on the protein surface due to an attractive interaction between urea and TMAO.^{31,32} These results raise several questions that require further experimental verification. In particular, the simulations raise the following questions. Does the synergistic effect plateau at a high concentration? Are equimolar mixtures of L-Arg and L-Glu also effective at inhibiting heat-induced protein aggregation? Can the synergistic accumulation of additive on a protein surface due to the additive–additive interactions be exploited for the design of additive mixtures, which could be used for high concentration formulations? Furthermore, in the broad context of stabilizing proteins in formulations, this study reiterates the need to understand all molecular interactions in the formulation in order to determine the overall stability of the formulation.^{33–35}

ASSOCIATED CONTENT

Supporting Information. The details of the methodology to estimate the effect of crowders on protein–protein association and a movie showing additive molecules around proteins. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Phone: (617)-258-5021. Fax: (617)-253-2272. E-mail: trout@mit.edu.

ACKNOWLEDGMENT

The authors acknowledge funding support from the Singapore–MIT Alliance (SMA). The authors also thank Curtiss P. Schneider of the Department of Chemical Engineering at MIT for careful reading of the manuscript and his useful suggestions.

REFERENCES

- (1) Manning, M. C.; Chou, D. K.; Murphy, B. M.; Payne, R. W.; Katayama, D. S. *Pharm. Res.* **2010**, *27*, 544–575.
- (2) Baynes, B. M.; Wang, D. I. C.; Trout, B. L. *Biochemistry* **2005**, *44*, 4919–4925.
- (3) Shukla, D.; Zamolo, L.; Cavallotti, C.; Trout, B. L. *J. Phys. Chem. B* **2011**, *115*, 2645–2654.
- (4) Shukla, D.; Trout, B. L. *J. Phys. Chem. B* **2010**, *114*, 13426–13438.
- (5) Golovanov, A. P.; Hautbergue, G. M.; Wilson, S. A.; Lian, L.-Y. *J. Am. Chem. Soc.* **2004**, *126*, 8933–8939.
- (6) Lau, C.-K.; Diem, M. D.; Dreyfuss, G.; Van Duyne, G. D. *Curr. Biol.* **2003**, *13*, 933–941.
- (7) Tunnicliffe, R. B.; Hautbergue, G. M.; Kalra, P.; Jackson, B. R.; Whitehouse, A.; Wilson, S. A.; Golovanov, A. P. *PLoS Pathog.* **2011**, *7*, e1001244.
- (8) Cornell, M.; Evans, D. A. P.; Mann, R.; Fostier, M.; Flasza, M.; Monthatong, M.; Artavanis-Tsakonas, S.; Baron, M. *Genetics* **1999**, *152*, 567–576.
- (9) Valente, J. J.; Verma, K. S.; Manning, M. C.; William Wilson, W.; Henry, C. S. *Biophys. J.* **2005**, *89*, 4211–4218.
- (10) Schneider, C. P.; Trout, B. L. *J. Phys. Chem. B* **2009**, *113*, 2050–2058.
- (11) Courtenay, E. S.; Capp, M. W.; Anderson, C. F.; Record, M. T. *Biochemistry* **2000**, *39*, 4455–4471.
- (12) Gekko, K.; Timasheff, S. N. *Biochemistry* **1981**, *20*, 4667–4676.
- (13) Shukla, D.; Shinde, C.; Trout, B. L. *J. Phys. Chem. B* **2009**, *113*, 12546–12554.
- (14) Shukla, D.; Trout, B. L. *J. Phys. Chem. B* **2011**, *115*, 1243–1253.
- (15) Baynes, B. M.; Trout, B. L. *J. Phys. Chem. B* **2003**, *107*, 14058–14067.
- (16) Arakawa, T.; Timasheff, S. N. In *Diffraction Methods for Biological Macromolecules Part A*; Academic Press, 1985; Vol. 114, pp 49–77.
- (17) Arakawa, T.; Bhat, R.; Timasheff, S. N. *Biochemistry* **1990**, *29*, 1914–1923.
- (18) Baynes, B. M.; Trout, B. L. *Biophys. J.* **2004**, *87*, 1631–1639.
- (19) Vagenende, V.; Yap, M. G. S.; Trout, B. L. *Biochemistry* **2009**, *48*, 11084–11096.
- (20) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (21) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (23) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (24) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613.
- (25) Shulgina, I. L.; Ruckenstein, E. *Biophys. Chem.* **2005**, *118*, 128–134.
- (26) Prausnitz, J. M.; Lichtenthaler, R. N.; Gomes de Azevedo, E. *Molecular Thermodynamics of Fluid-Phase Equilibria*; Prentice Hall: Upper Saddle River, NJ, 1999.
- (27) Tsurko, E. N.; Neueder, R.; Kunz, W. *Acta Chim. Slov.* **2009**, *56*, 58–64.
- (28) Tsurko, E. N.; Neueder, R.; Kunz, W. *J. Solution Chem.* **2007**, *36*, 651–672.
- (29) Bonner, O. D. *J. Chem. Eng. Data* **1982**, *27*, 422–423.
- (30) Suresh, C. G.; Ramaswamy, J.; Vijayan, M. *Acta Crystallogr., Sect. B: Struct. Sci.* **1986**, *42*, 473–478.

- (31) Zou, Q.; Bennion, B. J.; Daggett, V.; Murphy, K. P. *J. Am. Chem. Soc.* **2002**, *124*, 1192–1202.
- (32) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6433–6438.
- (33) Shukla, D.; Schneider, C. P.; Trout, B. L. *Adv. Drug Deliv. Rev.* **2011**, DOI: 10.1016/j.addr.2011.06.014.
- (34) Schneider, C. P.; Shukla, D.; Trout, B. L. *J. Phys. Chem. B* **2011**, *115*, 7447–7458.
- (35) Shukla, D.; Schneider, C. P.; Trout, B. L. *J. Phys. Chem. Lett.* **2011**, *2*, 1782–1788.