

# Contribution of Charged Groups to the Enthalpic Stabilization of the Folded States of Globular Proteins<sup>†</sup>

Voichita M. Dadarlat and Carol Beth Post\*

Department of Medicinal Chemistry and Molecular Pharmacology, Markey Center for Structural Biology, Purdue Cancer Center, Purdue University, West Lafayette, Indiana 47907

Received: September 1, 2007; In Final Form: November 22, 2007

In this paper we use the results from all-atom molecular dynamics (MD) simulations of proteins and peptides to assess the individual contribution of charged atomic groups to the enthalpic stability of the native state of globular proteins and investigate how the distribution of charged atomic groups in terms of solvent accessibility relates to protein enthalpic stability. The contributions of charged groups is calculated using a comparison of nonbonded interaction energy terms from equilibrium simulations of charged amino acid dipeptides in water (the “unfolded state”) and charged amino acids in globular proteins (the “folded state”). Contrary to expectation, the analysis shows that many buried, charged atomic groups contribute favorably to protein enthalpic stability. The strongest enthalpic contributions favoring the folded state come from the carboxylate (COO<sup>-</sup>) groups of either Glu or Asp. The contributions from Arg guanidinium groups are generally somewhat stabilizing, while NH<sub>3</sub><sup>+</sup> groups from Lys contribute little toward stabilizing the folded state. The average enthalpic gain due to the transfer of a methyl group in an apolar amino acid from solution to the protein interior is described for comparison. Notably, charged groups that are less exposed to solvent contribute more favorably to protein native-state enthalpic stability than charged groups that are solvent exposed. While solvent reorganization/release has favorable contributions to folding for all charged atomic groups, the variation in folded state stability among proteins comes mainly from the change in the nonbonded interaction energy of charged groups between the unfolded and folded states. A key outcome is that the calculated enthalpic stabilization is found to be inversely proportional to the excess charge density on the surface, in support of an hypothesis proposed previously.

## 1. Introduction

Variation in solvent accessible areas, packing density, or van der Waals and electrostatic interactions has been traditionally used to account for changes in protein stability upon amino acid mutations or under different environmental conditions. A quantitative understanding of the factors contributing to protein native-state stability remains an important goal, with benefits for advancing protein design efforts and understanding human disease. Studies to define the origins of structural stability are particularly relevant because of the implication of stability as a causative effect in disease-causing mutations. A decrease in protein stability by 1 to 3 kcal/mol can be a major factor contributing to monogenic human disease.<sup>1–3</sup>

At physiological temperatures, the native structure of most globular proteins is only marginally stable. The folded state is enthalpically but not entropically favored relative to the unfolded random-coil state; however, the enthalpy and entropy of unfolding on a per-residue basis vary in magnitude among globular proteins.<sup>4</sup> The variation in the residue-specific entropy of unfolding is understood to arise primarily from differences in amino acid composition of buried residues, and can be rationalized in terms of changes in the solvation of protein surface area and hydrophobicity.<sup>5</sup> In contrast, the physical origin of the specific enthalpy of unfolding is not well understood.<sup>6,7</sup>

Recently, a computational analysis of protein compressibility showed a correlation involving the distribution of charged atoms

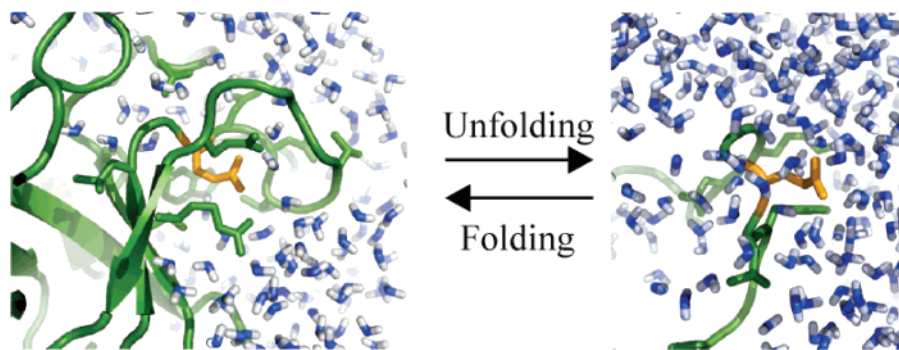
between the surface and interior of the protein;<sup>8</sup> together, the results inferred an unexpected relationship between the enthalpy of unfolding and this charge distribution whereby the per-residue change in enthalpy upon protein unfolding,  $\Delta H^U$ , decreases (the folded state is less stable) as the fraction of charged atoms located on the protein surface increases. We postulated that the variation among proteins in  $\Delta H^U$  is due to differences in the excess of surface charged atoms but in a counter-intuitive way: a more uniform distribution of charge between the surface and interior stabilizes protein structure. Here, we test this postulate by estimating the enthalpic stability from all-atom molecular dynamics (MD) simulations.

The change in solution free energy upon protein folding has three primary sources: (a) the favorable decrease in free energy due to the formation of intraprotein interactions, (b) the change in free energy due to the disruption of amino acid–water interactions (or desolvation) of all amino acids that are buried in the protein core or only partially exposed to water in the protein folded state, and (c) the change in the free energy of the solvent due to solvent reorganization and structural collapse. The focus of this investigation is to estimate the enthalpic contribution of charged amino acids, Asp, Glu, Arg, and Lys, to protein folded state stability as a function of solvent accessibility and the spatial distribution of charged groups.

While there is a general agreement that burial of hydrophobic amino acids in the protein interior is a major driving force for protein folding and stability, the contributions from charged atomic groups and hydrogen bonding to protein folded state

<sup>†</sup> Part of the “Attila Szabo Festschrift”.

\* Corresponding author.



**Figure 1.** The change in the local environment of charged groups upon protein unfolding: a carboxylate group from GLU (in gold) in a protein folded state environment (ribbon diagram, in green, on the left side of the scheme), and in an unfolded state environment (on the right). Water molecules in the folded and unfolded states are shown in blue.

stability has been a much debated issue. Total or partial desolvation of charged and polar groups upon protein folding is widely regarded as a limiting step in protein folding. Some researchers<sup>9–14</sup> have favored a stabilizing contribution from salt bridges and side-chain–side-chain hydrogen bonds,<sup>15</sup> while others have contended that hydrogen bonding and salt bridges make an unfavorable contribution.<sup>16–18</sup> Transfer of a salt bridge between charged atomic groups from water to a nonpolar environment was found to be destabilizing by 10 to 16 kcal/mol.<sup>19</sup> Nonetheless, the increase in the number of salt bridges in the native structure of the thermophilic proteins has been regarded as an indicator that salt bridges could stabilize the protein folded state<sup>6,20</sup> and that intraprotein electrostatic interactions are responsible for the increased stability of thermophilic proteins.<sup>21</sup>

Quantitative measurements of salt bridge contributions to the stability of proteins also lend contradictory results. Salt bridges stabilize the folded state of T4 lysozyme by 5 to 0.5 kcal/mol depending on the experimental system and the location of the salt bridge.<sup>22,23</sup> Certain salt bridges substantially stabilize the folded state of T4 lysozyme, but ionic interactions at the solvent exposed surface have little or no contribution to stabilization. In other cases, surface salt bridges were found to add several kilocalories per mole to protein stability.<sup>24</sup> On the other hand, buried salt bridges were reported to destabilize a coiled-coil protein and Arc repressor protein by 2 to 4 kcal/mol.<sup>25,26</sup> Thus experimental studies of salt bridges submit opposing conclusions of stabilization and destabilization of the protein folded state.

In an earlier study of protein compressibility,<sup>8</sup> we linked the enthalpy of unfolding per residue to the excess charge on the protein surface and found that globular proteins that have lower excess charge (more charges buried) are enthalpically more stable in the folded state. The excess charge,  $\delta_{xs}$ , is the difference between protein surface charge density and the charge density in the protein interior. The inverse relationship between the excess charge and the change in solution enthalpy upon protein unfolding was observed for experimental data from a set of 16 globular proteins. We hypothesized that protein folded state stabilization is partially due to a gain in enthalpy from favorable electrostatic interactions between buried charges as a result of optimal charge distributions evolved to stabilize buried charge in globular proteins. A goal of the present work is to test this hypothesis and investigate the origin of the empirical correlation between the distribution of buried charged groups in native proteins and protein folded state stability.

We use the results from all-atom MD simulations of proteins and peptides to assess the individual contribution of charged atomic groups to the enthalpic stability of the native state of globular proteins and investigate how the distribution of charged

atomic groups in the protein tertiary structure influences their contribution to protein enthalpic stability. The net contribution of each charged atomic group is estimated as a sum of two components: (a) the change in the nonbonded interaction energy of charged amino acids between the unfolded state modeled by solvated dipeptides and the folded state of globular proteins; and (b) the change in the enthalpy of the hydration water due to the collapse of the hydration shells formed by charged atomic groups and subsequent release of certain water molecules to the bulk solvent in the process of protein folding. The concept is illustrated in Figure 1, where a carboxylate group from GLU (in gold), is shown in a protein folded state environment (on the left side of the scheme), and an unfolded state environment (on the right). The analysis shows that most native, charged atomic groups have significant contributions to protein enthalpic stability. While solvent reorganization/release has favorable contributions to folding for all charged atomic groups, the variation in folded state stability among proteins comes mainly from the change in the nonbonded interaction energy of charged amino acids between the unfolded and folded states.

The carboxylate ( $\text{COO}^-$ ) groups from Glu and Asp are the leaders in enhancing the folded state stability. The guanidinium groups from Arg are somewhat stabilizing, while an  $\text{NH}_3^+$  group from Lys only marginally stabilizes the folded state. The average enthalpic gain due to the transfer of a methyl group in an apolar amino acid from solution to the protein interior is described for comparison. The trend in stabilization by the type of charged group is consistent with the distribution of the relative numbers of charged groups as a function of the degree of solvent exposure. Charged groups that are less exposed to solvent contribute more to protein native state enthalpic stability than charged groups that are exposed, and the calculated cumulative effect of charge is inversely proportional to the excess charge density,  $\delta_{xs}$ .

## 2. Theory and Computational Methods

**2.1. Model Systems.** The contributions of individual atomic groups to the enthalpic stability of the folded state were assessed from a comparison of their enthalpy in the folded and unfolded states. The atomic groups considered here are the side-chain carboxylate groups of Asp and Glu ( $\text{COO}^-$ ), the guanidinium group of Arg (Gu); the side-chain amino group of Lys ( $\text{NH}_3^+$ ); and the charged  $\text{Ca}^{2+}$  ions in  $\alpha$ -lactalbumin and trypsin. The apolar group is the methyl group,  $-\text{CH}_3$ , in ALA, VAL, LEU, and ILE. The unfolded state of the protein is modeled here as charged or apolar dipeptides,  $\text{CH}_3-\text{CO}-\text{XXX}-\text{NH}-\text{CH}_3$ , where XXX is Glu, Asp, Arg, Lys, Ala, Val, Leu, Ile, and the corresponding dipeptides are glud, aspd, argd, lysd, alad, vald,

leud, and iled, respectively. This definition of the unfolded state assumes that the unfolded state is a random coil, an ensemble of conformations in which the amino acid chain is highly hydrated and individual residues do not interact with each other. The backbone dihedral angles of the dipeptides are not fixed; they exhibit both  $\alpha$  and  $\beta$ /PPII (stretched) conformations with relative populations specific to each dipeptide.

The folded state corresponds to the native states of four globular proteins: trypsin, ribonuclease A, hen egg-white lysozyme, and  $\alpha$ -lactalbumin. The total number of charged atomic groups in the four proteins is 104. Of these, 28% are  $\text{COO}^-$  (Asp), 16% are  $\text{COO}^-$  (Glu), 17% are from Arg, and the remaining 38% are  $\text{NH}_3^+$  groups from Lys. In vertebrates, 23% of amino acids are charged amino acids,<sup>27</sup> and the relative composition in the charged category is as follows: 26% are Asp, 25% are Glu, 18% are Arg, and 31% are Lys. Therefore, our set of proteins is a reasonably representative set.

**2.2. Change in Enthalpy upon Protein Unfolding.** The net contribution of each atomic group to the change in enthalpy upon protein unfolding,  $\Delta H^U$ , is a sum of two quantities: (a) the change in the enthalpy of the charged group  $g$ , due to a change in the local environment for transfer from the folded state to the unfolded state,  $\Delta H_g^U$ ; and (b) the negative change in the enthalpy of the hydration water molecules that are released to the bulk upon protein folding,  $\Delta H_h^U$ :

$$\Delta H^U = \Delta H_g^U + \Delta H_h^U \quad (1)$$

If  $\Delta H^U > 0$ , the atomic group contributes favorably to protein folding. The sum of  $\Delta H^U$  over all atomic groups in a protein corresponds to the unfolding enthalpy.

**2.2.1. Local Environment of Charged Groups,  $\Delta H_g^U$ .**  $\Delta H_g^U$  of an atomic group is estimated from the difference between the time-average nonbonded energy (a sum of the electrostatic and van der Waals terms) in the unfolded state modeled by the dipeptide solution,  $\langle E_{\text{nb}}^{g,\text{dipep}} \rangle$ , and the corresponding energy in the folded state,  $\langle E_{\text{nb}}^{g,\text{prot}} \rangle$ :

$$\Delta H_g^U = 1/2(\langle E_{\text{nb}}^{g,\text{dipep}} \rangle - \langle E_{\text{nb}}^{g,\text{prot}} \rangle) \quad (2)$$

where  $\langle \dots \rangle$  represents the time average. In this approximation, the difference between the covalent bonding energies in the folded and unfolded states is assumed to be negligible.

The nonbonded energy of an atomic group,  $E_{\text{nb}}^g$ , in either the native protein or dipeptide solution, is a time average of the pairwise sum over the van der Waals and electrostatic interactions of all atoms in the atomic group,  $n_g$ , with the protein (or dipeptide) and the solvent molecules within a specified cutoff distance:

$$\langle E_{\text{nb}}^g \rangle = \left\langle \sum_{i=1}^{n_g} \sum_{j \neq i}^m (E_{\text{vdW}}(r_{ij}) + E_{\text{elec}}(r_{ij})) \right\rangle \quad (3)$$

where  $m$  is the number of all atoms within the cutoff distance.

**2.2.2. Solvent Reorganization,  $\Delta H_h^U$ .** The second term in eq 1,  $\Delta H_h^U$ , is the negative change in solvation water enthalpy due to changes in the solvation shells around charged groups upon protein folding. The released water molecules become bulk water molecules in the peptide solution. Water molecules in the vicinity of proteins are “reorganized” with respect to the bulk water, and the corresponding radial distribution functions show structure up to 8 Å away from charged atomic groups.<sup>28</sup> Three hydration shells are formed: the first hydration shell extends from 0 to 3.2 Å, a second shell is evident between 3.2

and 6 Å, and a less pronounced third hydration shell is formed between 6 and 8 Å.

To calculate the change in hydration water enthalpy and account for its release from the dipeptide surface in the unfolded state, we estimate the average numbers of water molecules in the first, second and third hydration shells for each charged group in the dipeptide solution ( $N_w^{1\text{st}}$ ,  $N_w^{2\text{nd}}$ ,  $N_w^{3\text{rd}}$ ), and the difference in the nonbonded energy of a water molecule in each hydration shell relative to bulk water:  $\Delta H_w^{1\text{st}}$ ,  $\Delta H_w^{2\text{nd}}$ , and  $\Delta H_w^{3\text{rd}}$ . The average change in enthalpy for each water molecule in a hydration shell,  $\Delta H_w^{\text{shell}}$ , is calculated as follows:

$$\Delta H_w^{\text{shell}} = 1/2(\langle E_{\text{nb}}^{\text{shell}} \rangle / N_w^{\text{shell}} - \langle E_{\text{nb}}^{\text{w}} \rangle) \quad (4)$$

where  $\langle E_{\text{nb}}^{\text{shell}} \rangle / N_w^{\text{shell}}$  and  $\langle E_{\text{nb}}^{\text{w}} \rangle$  are the average nonbonded energy of a hydration-shell water molecule and a bulk water molecule, respectively, in the dipeptide solution.  $\langle E_{\text{nb}} \rangle$  is defined according to eq 3.  $N_w^{1\text{st}}$ ,  $N_w^{2\text{nd}}$ , and  $N_w^{3\text{rd}}$  are calculated from postprocessing the trajectories.  $N_w^{1\text{st}}$  is the average number of water molecules in a spherical shell extending to 3.2 Å from a charged group in the peptide solution. To estimate  $N_w^{2\text{nd}}$ , we calculated the number of water molecules in a spherical shell between 3.2 and 6 Å around the charged group, excluding water molecules that belong to the first hydration shell of other polar or apolar groups. In the same manner,  $N_w^{3\text{rd}}$  is calculated to include water molecules between 6 and 8 Å from the charged group, excluding water molecules that are in the first hydration shell of apolar or polar groups. The maximum change in solvation water enthalpy for the release of all hydration water molecules from the surface of the charged group when the protein folds,  $\Delta H_h^{\text{max}}$  is estimated as

$$\Delta H_h^{\text{max}} = N_w^{1\text{st}} \times \Delta H_w^{1\text{st}} + N_w^{2\text{nd}} \times \Delta H_w^{2\text{nd}} + N_w^{3\text{rd}} \times \Delta H_w^{3\text{rd}} \quad (5)$$

For charged groups that are not completely buried upon protein folding, the actual  $\Delta H_h^U$  is less than  $\Delta H_h^{\text{max}}$ . The solvent reorganization contribution can be estimated by factoring in the fractional degree of charge burial,  $1 - \text{SE}$ , where SE is the fractional exposure of the charged group to solvent (a definition of SE is given below). Therefore,

$$\Delta H_h^U = (1 - \text{SE}) \times \Delta H_h^{\text{max}} \quad (6)$$

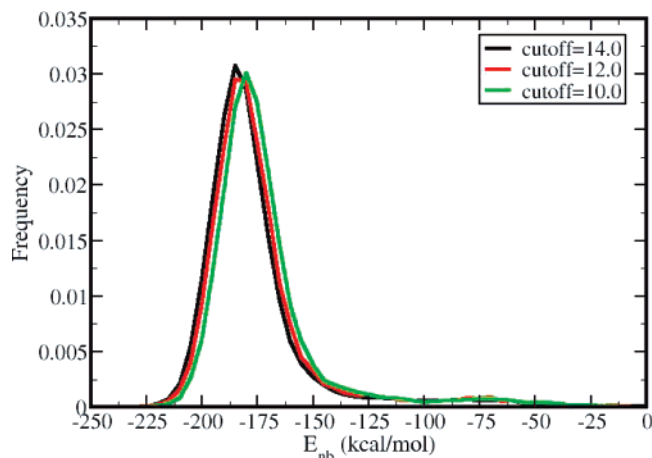
**2.3. Fractional Solvent Exposure for Atomic Groups.** The nonbonded energy of each atomic group depends on its specific microenvironment, the sum of all the interactions with the neighboring atomic groups. We distinguish “environments” according to the location of the charged groups: from fully buried in the protein to gradually exposed to the solvent on the protein surface. The peptide environment is taken here to be similar to the environment encountered by the charged groups in the unfolded protein.

The fractional degree of exposure to solvent, SE, is the ratio of the time average solvent accessible surface area exposed by an atomic group in the folded protein,  $\langle \text{SA}^{g,\text{prot}} \rangle$ , and the time average solvent accessible surface area exposed by the same charged group in the corresponding dipeptide,  $\langle \text{SA}^{g,\text{dipep}} \rangle$ :

$$\text{SE} = \langle \text{SA}^{g,\text{prot}} \rangle / \langle \text{SA}^{g,\text{dipep}} \rangle \quad (7)$$

The time average SA for each atomic group is calculated from the time series of MD simulations using CHARMM<sup>29</sup> with a probe radius of 1.4 Å (using the algorithm of Lee and Richards<sup>30</sup>).





**Figure 2.** The influence of cutoff distance on the distribution of the nonbonded energies for the carboxylate group in glud: 10 Å (black curve), 12 Å (red curve), and 14 Å (green curve). The time-averaged nonbonded energy is  $-180$  kcal/mol for 10 Å,  $-182.5$  kcal/mol for 12 Å, and  $-184$  kcal/mol for 14 Å. The results presented in this study are calculated using a nonbonded cutoff distance of 14 Å.

**2.4. Molecular Dynamics Simulations.** Equilibrium MD simulations for four proteins and eight dipeptides were performed using CHARMM.<sup>29</sup> A solvated  $\text{Ca}^{2+}$  ion that is found buried in both trypsin and  $\alpha$ -lactalbumin, was also simulated in the “unfolded” state. The main simulation boxes contained 3500 to 5000 explicit water molecules (1100 water molecules were included in the dipeptide and ion solutions) as well as charge neutralizing counterions. Additional  $\text{Na}^+$  and  $\text{Cl}^-$  molecules were added up to a final salt concentration of 0.166 M. Periodic boundary conditions and the particle mesh Ewald method<sup>31</sup> for calculating electrostatic interactions were used with a 2 fs time step for the integration of the equations of motion, in an NPT ensemble. Charged atomic groups have protonation states corresponding to pH 7, with the exception of Glu35 in HEW lysozyme. Other details of the simulations were as reported earlier.<sup>32,33</sup> Two-nanosecond MD equilibrium simulation trajectories were post analyzed, and snapshots 10 fs apart were used to calculate the nonbonded energies of charged groups.

Distributions of nonbonded energies are shown in Figure 2 for the example of the  $\text{COO}^-$  group in glud and using a 10, 12, or 14 Å cutoff distance. The distributions are well converged. The time-averaged nonbonded energy is  $-180.0$  kcal/mol for 10 Å,  $-182.5$  kcal/mol for 12 Å, and  $-184.0$  kcal/mol for 14 Å. Our analysis utilizes the time-average nonbonded energy values, which, for the well-converged  $E_{\text{nb}}$  distributions observed in this study, are effectively equivalent to the most likely nonbonded energy; average nonbonded energies are within 1 kcal/mol from the most likely energies. Nonbonded energy cutoffs were set to 14 Å, and all images within 18 Å were included in the calculation.

### 3. Results

**3.1. Distribution of Charged Groups as a Function of Solvent Exposure.** The positions of the four types of charged atomic groups in the protein structure in terms of solvent exposure (eq 7) are summarized in Table 1. The level of solvent exposure, SE, for charged atomic groups is divided into five categories: less or equal to 20% exposed ( $\text{SE}_{0-20}$ ), between 20 and 40% ( $\text{SE}_{20-40}$ ), between 40 and 60% ( $\text{SE}_{40-60}$ ), between 60 and 80% ( $\text{SE}_{60-80}$ ), and between 80 and 100% ( $\text{SE}_{80-100}$ ) solvent exposed. The fraction of charged atomic groups,  $f_{\text{SE}}^g$ , is

**TABLE 1: Numbers of Charged Atomic Groups as a Function of Their Exposure to Solvent, SE (in %) in the Protein Folded State<sup>a</sup>**

group	protein	$N_{0-100}^g$	$N_{0-20}^g$	$N_{20-40}^g$	$N_{40-60}^g$	$N_{60-80}^g$	$N_{80-100}^g$
$\text{COO}^-$ (ASP)	TRPS	6	4	0	1	0	1
	RBNA	5	2	1	0	2	0
	HLYSO	7	1	3	2	1	0
	ALACTA	11	3	3	1	3	1
	$\sum N^g$	29	10	7	4	6	2
	$f_{\text{SE}}^g$		34	24	14	21	7
$\text{COO}^-$ (GLU)	TRPS	4	2	0	1	0	1
	RBNA	5	0	2	1	1	1
	HLYSO	1	0	0	0	1	0
	ALACTA	7	0	2	2	0	3
	$\sum N^g$	17	2	4	4	2	5
	$f_{\text{SE}}^g$		12	24	24	12	28
Gu (ARG)	TRPS	2	0	1	0	1	0
	RBNA	4	1	1	1	1	0
	HLYSO	11	1	1	3	3	3
	ALACTA	1	0	0	0	0	1
	$\sum N^g$	18	2	3	4	5	4
	$f_{\text{SE}}^g$		11	17	22	28	22
$\text{NH}_3^+$ (LYS)	TRPS	14	1	1	4	6	2
	RBNA	10	1	1	4	1	3
	HLYSO	5	0	0	1	4	0
	ALACTA	11	1	1	2	6	1
	$\sum N^g$	40	3	3	11	17	6
	$f_{\text{SE}}^g$		7.5	7.5	27.5	42.5	15
All Crg	$\sum \sum N^g$	104	17	17	23	30	17
All Crg	$f_{\text{SE}}^{\text{tot}}$		17	16	23	30	16

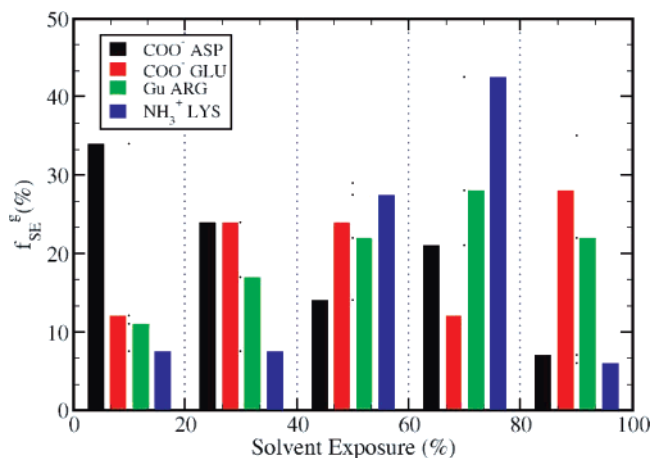
<sup>a</sup>  $N^g$  is the number of charged groups of type “g” with the indicated solvent exposure.

the ratio between the number of charged atomic groups of type  $g$  for each SE range, and the total number of atomic groups of type  $g$ , multiplied by 100.

The distribution of the charged groups by type as a function of solvent exposure is shown in Figure 3. For the proteins studied here, the carboxylate group of Asp has the highest percentage of buried groups (58%) within  $\text{SE}_{0-40}$  (see also Tables 3S through 6S, Supporting Information). In contrast, Lys amino groups are largely solvent exposed and not buried; 42.5% fall within the range of  $\text{SE}_{60-80}$ , while only 15% have less than 40% SE. These data indicate that, on average, the basic amino acids Lys and Arg tend to position themselves closer to the protein–water interface than the acidic amino acids Asp and Glu.

**3.2. Unfolded State Protein Enthalpy.** The time-averaged values  $\langle E_{\text{nb}}^{g,\text{dipep}(g)} \rangle$  and  $\langle \text{SA} \rangle$  for the charged groups in aspd, glud, argd, and lysd, the methyl groups in alad, vald, leud, and ilel, and the  $\text{Ca}^{2+}$  ion were evaluated from distributions from 2 ns MD simulations. (Table 2S, Supporting Information). The methyl group values range from  $-13.2$  kcal/mol for leud to 2.2 kcal/mol for alad. Thus, methyl groups of vald, leud, and ilel ( $\gamma$ ) have negative energies, while the side chain in alad and ilel ( $\delta$ ) has weakly unfavorable nonbonded interactions with its environment. The charged groups have strong, favorable interaction energies ranging from  $-72$  kcal/mol for the guanidinium group in argd to  $-201$  kcal/mol for the carboxylate group in aspd. The carboxylate group in glud has a nonbonded interaction energy of  $-184$  kcal/mol, and that of the  $\text{NH}_3^+$  group in lysd is  $-86$  kcal/mol.

**3.3. Contributions of Apolar Groups to the Enthalpy of Unfolding from the Change in Local Environment,  $\Delta H_g^{\text{U}}$ .** To provide a basis of comparison for the enthalpy of charged groups, the enthalpy change for apolar methyl groups upon protein unfolding was evaluated from the difference of non-

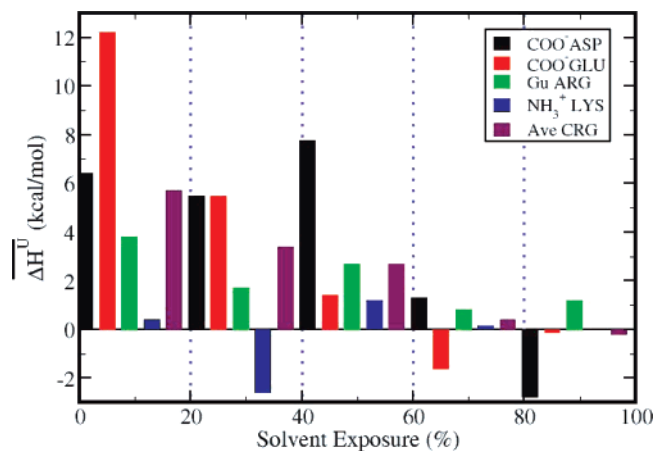


**Figure 3.** Distribution of charged groups in terms of solvent exposure. The fraction of each type of charged groups,  $f_{SE}^g$  (in %) as a function of SE values in the ranges indicated along the abscissa. COO<sup>-</sup> (Asp) – black bars; COO<sup>-</sup> (Glu) – red bars; Gu (Arg) – green bars; NH<sub>3</sub><sup>+</sup> (Lys) – blue bars. The total number of groups for each type is over the four proteins used in this study.

bonded interaction energy of methyl groups in apolar dipeptide molecules (alad, vald, leud, and iled) and that of the corresponding apolar amino acids, Ala, Val, Leu, and Ile in folded proteins (eq 2). The average  $\Delta H_g^U$  for a methyl group is 1.3 kcal/mol and varies with the amino acid in which the methyl group is embedded as well as its position in the protein structure (Table 1S, Supporting Information). The enthalpic gain for a side chain methyl group ranges from 0.8 to 1.5 kcal/mol for Ala, from 0 to 0.6 kcal/mol for Val, from 1 to 1.9 kcal/mol for Ile, and from 0.6 to 1.1 kcal/mol for Leu. The variation observed here is consistent with the site-specific differences in contributions of apolar groups to protein stability from experiment,<sup>10</sup> whereby the local environment, as characterized by the number of methyl and methylene groups in the neighborhood of a specific methyl group, was found to influence the free energy change upon folding.

**3.4. Contributions to the Enthalpy of Unfolding Due to Solvent Reorganization,  $\Delta H_h^U$ .** The change in the number of bound water molecules as a result of the increase in protein SA upon unfolding is accounted for with  $\Delta H_h^U$  calculated from the nonbonded energy differences of a water molecule in pure solvent and one in each hydration shell, plus an estimate of the number of hydration water molecules (see Theory and Computational Methods). Values for each hydration shell of  $N_w^{1st}$ ,  $N_w^{2nd}$ ,  $N_w^{3rd}$ ;  $\langle E_{nb}^{1st} \rangle$ ,  $\langle E_{nb}^{2nd} \rangle$ ,  $\langle E_{nb}^{3rd} \rangle$ ;  $\Delta H_w^{1st}$ ,  $\Delta H_w^{2nd}$ ,  $\Delta H_w^{3rd}$ ; and the maximum change in solvent enthalpy due to reorganization for a fully buried charged group,  $\Delta H_h^{max}$ , are listed in Table 7S (Supporting Information).  $\Delta H_w^{shell}$  depends on the nonbonded energy of a water molecule in pure solvent, which is  $\langle E_{nb}^w \rangle = -9.2$  kcal/mol.  $\Delta H_h^{max}$  ranges from 1.1 to 2.4 kcal/mol.

While water molecules in the first hydration shell of charged groups have lower  $\langle E_{nb}^{1st} \rangle$  than bulk water molecules (by 1 to 3 kcal/mol, Table 7S) due to strong charge–dipole interactions between the charged atomic groups and water molecular dipoles, water molecules in the second and third hydration shells have unsatisfied hydrogen-bonding capabilities so  $\langle E_{nb}^{2nd} \rangle$  and  $\langle E_{nb}^{3rd} \rangle$  are greater than  $\langle E_{nb}^w \rangle$ . The combined result of the integrated effects for the loss of hydration shells is a net favorable enthalpy gain upon protein folding. The contribution of solvent reorganization/release favors protein folding but differentiates among charged atomic groups; the maximum gain in enthalpy upon complete burial of a charged group is from ASP, which



**Figure 4.** Average contribution to protein unfolding enthalpy for charged group type,  $\Delta H^U$ , as a function of solvent exposure, SE (in %): COO<sup>-</sup> (Asp) – black bars; COO<sup>-</sup> (Glu) – red bars; Gu (Arg) – green bars; NH<sub>3</sub><sup>+</sup> (Lys) – blue bars.

contributes 2.4 kcal/mol toward the folded state stability. Complete burial of the same group from GLU, or the charged groups of Arg and Lys, leads to gains in solvent enthalpy of 1.7, 1.9 and 1.1 kcal/mol, respectively. Estimates from vald for the maximum change in the enthalpy due to water molecules near apolar groups (Table 7S, last row) indicate that the change is similar in magnitude (approximately 2.4 kcal/mol) to that for charged groups, even though the mechanism that generates the change is different. Water in the first and third hydration shells of apolar groups has a less favorable nonbonded interaction energy than average water molecules, while water molecules in the second hydration shells have a more favorable interaction energy.

**3.5. Charged Group Enthalpy of Unfolding,  $\Delta H^U = \Delta H_g^U + \Delta H_h^U$ .** The enthalpic contributions to protein unfolding (eq 1) summed over the charged groups from either Asp, Glu, Arg, or Lys residues in each of the four proteins are shown in Table 2 as a function of SE. The individual residue contributions are provided in Tables 3S to 6S (Supporting Information) along with residue values for  $\langle E_{nb}^g \rangle$ ,  $\langle SA \rangle$ , SE,  $\Delta H_g^U$ , and  $\Delta H_h^U$ . Values for  $\Delta H^U$  summed over residue type in each of the four proteins (Table 2) are largely positive in sign, so that the enthalpic contribution from charged groups generally stabilizes the folded state. Integrated over all degrees of solvent exposure (i.e., SE<sub>0–100</sub>), the residue sums for  $\Delta H^U$  are positive except for only two cases: Glu residues in HLYSO (−0.8 kcal/mol) and Lys residues in ALACTA (−1.8 kcal/mol).

The global average of  $\Delta H^U$  over all charged groups,  $\Delta \tilde{H}^U$ , is 2.1 kcal/mol (Table 2, last row), a combination of  $\Delta \tilde{H}_g^U = 1.3$  kcal/mol for protein environmental effects and  $\Delta \tilde{H}_h^U = 0.9$  kcal/mol for solvent reorganization (Tables 8S and 9S, Supporting Information). These charged group global averages are similar to those for methyl enthalpy changes, which are  $\Delta \tilde{H}_g^U = 1.3$  kcal/mol and  $\Delta \tilde{H}_h^U = 2.4$  kcal/mol. Nonetheless, the variation among charged groups is considerably larger than that of methyl groups (Tables 3S to 6S). While methyl groups in apolar amino acids have favorable contributions to the enthalpy of the protein folded state, irrespective of solvent exposure (Table 1S), the charged atomic groups have both favorable and unfavorable contributions to protein folding.

The listings in Tables 3S to 6S also indicate from the positive  $\Delta H_h^U$  values associated with individual charged groups in the four proteins that the formation of hydration shells at the surface of the unfolded protein has the effect of increasing enthalpy.

**TABLE 2: Enthalpy of Unfolding, Change Due to Charged Groups upon Protein Unfolding,  $\Delta H^U = \Delta H_g^U + \Delta H_h^U$  (kcal/mol), Summed over Charged Groups of a Given Type as a Function of Exposure to Solvent, SE (in %)**

group	protein	$\Delta H^U$					
		0–100	0–20	20–40	40–60	60–80	80–100
$g = \text{COO}^-$ (ASP)	TRPS	+30.7	+16.0	0	+14.5	0	+2
	RBNA	+33.1	+8.4	+14.0	+9	+9.8	0.0
	HLYSO	+33.8	+9.9	8.7	+14.3	+9	0.0
	ALACTA	+38.4	+29.9	+15.8	+1.3	-2.9	-5.7
	$\Sigma\Delta H^U$	+136.0	+64.2	+38.5	+31.0	+7.8	-5.5
	$\Sigma N^g$	29	10	7	4	6	2
	$\overline{\Delta H^U}$	+4.7	+6.4	+5.5	+7.8	+1.3	-2.8
$g = \text{COO}^-$ (GLU)	TRPS	+27.9	+24.4	0.0	+4.4	0.0	+9
	RBNA	+18.5	0.0	+19.1	-2.1	-2.3	+3.8
	HLYSO	-8	0.0	0.0	0.0	-8	0.0
	ALACTA	+1.6	0.0	+3.0	+3.9	0.0	-5.3
	$\Sigma\Delta H^U$	+49.0	+24.4	+22.1	+6.2	-3.1	-6
	$\Sigma N^g$	17	2	4	4	2	5
	$\overline{\Delta H^U}$	+2.9	+12.2	+5.5	+1.4	-1.6	-1
$g = \text{Gu}$ (ARG)	TRPS	+2.1	0.0	.1	0.0	+2.0	0.0
	RBNA	+9.0	+2.5	+3.4	+2.8	+3	0.0
	HLYSO	+12.6	+5.0	+1.6	+7.8	+1.7	-1.6
	ALACTA	+5.2	0	0	0	0	+5.2
	$\Sigma\Delta H^U$	+28.9	+7.5	+5.1	+10.6	+4.0	+3.6
	$\Sigma N^g$	18	2	3	4	5	4
	$\overline{\Delta H^U}$	+1.6	+3.8	+1.7	+2.7	+8	+9
$g = \text{NH}_3^+$ (LYS)	TRPS	+8.7	+2.4	-1.3	+6.3	-2	+1.7
	RBNA	+6	-1.3	-3.3	+5.7	-1.0	+6
	HLYSO	+1.7	0.0	0.0	+1.6	.1	0.0
	ALACTA	-1.8	+1	-3.2	-.1	+3.5	-2.1
	$\Sigma\Delta H^U$	+9.2	+1.2	-7.8	+13.5	+2.4	+2
	$\Sigma N^g$	40	3	3	11	17	6
	$\overline{\Delta H^U}$	+2	+4	-2.6	+1.2	.03	
All	$\Sigma\Sigma\Delta H^U$	223.1	97.3	57.9	61.3	11.1	-2.3
All	$\Sigma\Sigma N^g$	104	17	17	23	30	17
All Ave	$\overline{\Delta H^U}$	2.1	5.7	3.4	2.7	.4	-2

<sup>a</sup> Group averages,  $\overline{\Delta H^U}$ , are averages over each charged group type in the indicated SE range, from the four proteins. Global averages,  $\overline{\Delta H^U}$ , are averages over all charged groups in four proteins in the indicated SE range.

Thus, in addition to the well-known increase in entropy upon the release of hydration water molecules during folding, the loss of hydration water molecules also favors folding for enthalpic reasons, which has been noted previously.<sup>34</sup> Nonetheless,  $\Delta H_h^U$  is generally smaller in magnitude (<2.4 kcal/mol) than  $\Delta H_g^U$ , which ranges from -5 to 13 kcal/mol.

**3.5.1. Carboxylate Group in Asp.** Asp residues exhibit the strongest enthalpic interactions stabilizing the folded state.  $\Delta H^U$  for all Asp residues (SE from 0 to 100) is greater than 30 kcal/mol in the case of all four proteins (Table 2). On the basis of these representative globular proteins, carboxylate groups from Asp contribute on average  $\Delta H^U = 4.7$  kcal/mol to protein folded-state stability. The fully buried  $\text{COO}^-$  group of Asp176 contributes  $\Delta H^U = 9.9$  kcal toward the stabilization of the TRPS folded state. This gain in enthalpic stability is more than twice the average enthalpic gain due to the transfer of a  $\text{CH}_3$  group from water in the hydrophobic core of a protein (see above). By showing that opposite effects on stability can occur, we note that a different buried  $\text{COO}^-$  group in TRPS, the  $\text{COO}^-$  of Asp171 (Table 3S), has  $\Delta H^U = -2.6$  kcal/mol and contributes unfavorably to the folded state of trypsin.

**3.5.2. Carboxylate Group in Glu.** Glu carboxylate groups have  $\Delta H^U = 2.9$  kcal/mol and can strongly stabilize tertiary structure, although to a somewhat lesser extent than those of Asp. The interaction of the  $\text{COO}^-$  group in glud with the solvent is not as weak (-184 kcal/mol) as its counterpart in the aspd (-201 kcal/mol). This difference is likely due to the more

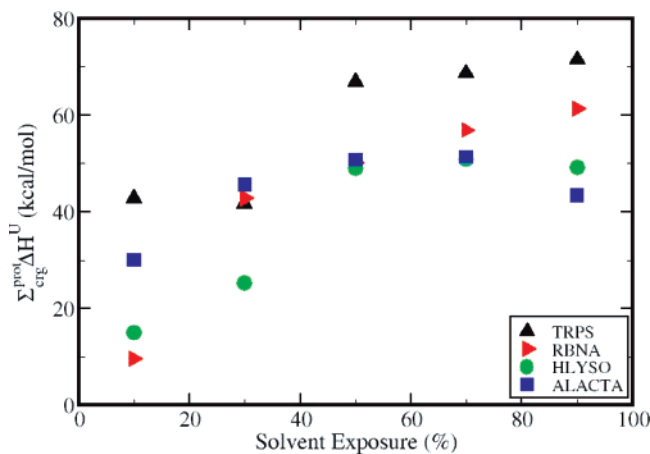
hydrophobic nature of the glutamic acid (it has an extra  $\text{CH}_2$  group). As part of a protein, however, the  $\text{COO}^-$  group of the glutamic acid can contribute as much favorable enthalpy as that of an aspartic acid. In a buried glutamic acid, Glu52,  $\text{COO}^-$  contributes 13.7 kcal/mol to the enthalpy of protein unfolding, and thus favors folding almost 4 times as much as an average methyl group, while the surface-exposed charged group from Glu167 contributes only 0.9 kcal/mol.

**3.5.3. Guanidinium Group in Arg.** The average effect of the guanidinium group of Arg residues to protein stability is slightly favorable by 1.6 kcal/mol (Table 2). Protein folding induces both positive and negative changes in enthalpy with values from different Arg residues ranging from +5.2 kcal/mol for the charged group from Arg11 in  $\alpha$ -lactalbumin (Table 6S) to -2.2 kcal/mol for the charged group from Arg128 (in HEW lysozyme).

**3.5.4. The  $\text{NH}_3^+$  Group in Lys.** Unlike other charged groups, the average contribution of the  $\text{NH}_3^+$  group of Lys is close to negligible:  $\Delta H^U = 0.2$  kcal/mol (Table 2). The folded-state environments are energetically similar to that of solvated Lys, with  $\overline{\Delta H_g^U} = -0.2$  kcal/mol (Table 8S).

**3.6. Charged Atomic Groups and Solvent Exposure.** The group average contributions from different charged group types with given solvent exposure (Table 2),  $\Delta H^U$ , are plotted as a function of solvent exposure in Figure 4. The major stabilizing contributions are from the buried  $\text{COO}^-$  groups of Glu (red) and Asp (black) with SE less than 60%. In particular, the  $\text{COO}^-$



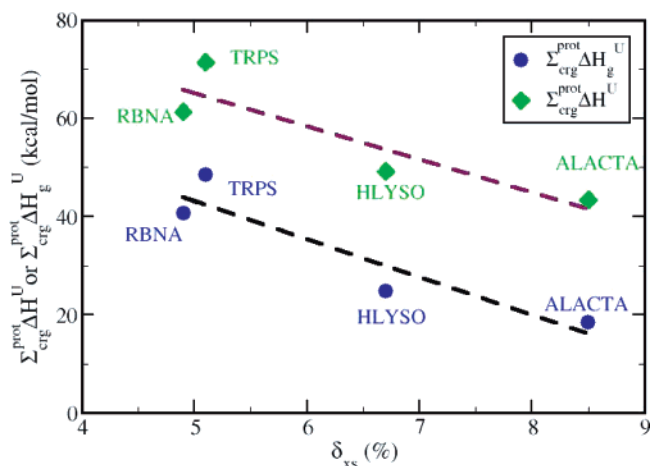


**Figure 5.** Cumulative contribution of buried charged groups to the folded state enthalpy as a function of solvent exposure by protein,  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H^{\text{U}}$  for four proteins: TRPS (black, triangle up symbols), RBNA (red, right triangles), HLYSO (green circles), and ALACTA (blue squares).

groups from Glu, which are less than 20% exposed to solvent, contribute an average of 12.2 kcal/mol. In contrast, groups with more than 60% of their surface exposed to solvent have a much smaller contribution to protein stability. The trend in  $\Delta H^{\text{U}}$  for types of charged groups is roughly consistent with their frequency of occurrence as a function of SE. Comparison of Figures 3 and 4 reveals higher frequency for greater stabilization in the folded state, as expected. Additionally, 58% of Asp, 36% of Glu, and 28% of Arg charged groups are more than 40% buried, while 85% of Lys charged groups are largely exposed to water.

The cumulative contribution of all charged groups of a protein to  $\Delta H^{\text{U}}$  ( $\Sigma_{\text{crg}}^{\text{prot}} \Delta H^{\text{U}}$ ) as a function of solvent exposure is shown in Figure 5. For the four proteins studied here, the cumulative contribution increases up to SE 40–60%, then plateaus or decreases slightly with the addition of the charged groups near the surface. Overall, charged groups contribute 43.4 kcal/mol to the stability of ALACTA, 49.2 kcal/mol for HLYSO, 61.3 kcal/mol for RBNA, and 71.4 kcal/mol for TRPS.

**3.7. Correlation between  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H^{\text{U}}$  and the Excess Charge Density,  $\delta_{\text{xs}}$ .** In an earlier paper,<sup>8</sup> we linked protein folded state stability to the excess charge on the protein surface and found that globular proteins with lower excess charge, or a more uniform charge distribution, have a more positive unfolding enthalpy per residue and thus are enthalpically more stable. The negative correlation between excess charge,  $\delta_{\text{xs}}$ , and experimental  $\Delta H^{\text{U}}$  values was shown for a set of 16 globular proteins (see Figure 5 in ref 8). To establish a causal relationship in this correlation, we examine here whether enthalpies calculated from only charged groups exhibit this trend. Figure 6 shows  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H^{\text{U}}$ , and the contributions from environmental change only,  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H_g^{\text{U}}$ , as a function of  $\delta_{\text{xs}}$  for all charged groups in the four proteins as determined in this study. The inverse correlation observed here for  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H^{\text{U}}$  from charged residues alone supports the postulate that differences among proteins in the specific enthalpy of unfolding are due to the spatial distribution of charge. Because most of the stabilizing effect related to charged groups comes from groups that are largely buried, the relationship observed with experimental data in protein folded state stability, showing increased protein stability with decreasing excess charge, is at least in part due to optimized nonbonded interactions between charged groups with the environment in the native protein folded state.



**Figure 6.** Total protein-based enthalpy contributions,  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H^{\text{U}}$ , from charged groups as a function of excess charge density for each protein (green diamonds and maroon dashed linear regression best fit), and the contributions from environmental change only,  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H_g^{\text{U}}$  (blue circles and black dashed linear regression best fit).

It is interesting to note that the contribution from solvent reorganization, although substantial (approximately 22 kcal/mol for each protein), does not influence the variation in stability among proteins, as shown by the two almost parallel best fits in Figure 6. The correlation coefficient for the linear fit of  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H^{\text{U}}$  (upper line, Figure 6) is  $-0.89$ , and that of  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H_g^{\text{U}}$  (lower line, Figure 6) is  $-0.93$ . Therefore, when comparing the relative stabilities of globular proteins, the enthalpy gain due to solvent release around charged groups appears to provide an approximately constant enhancement to the specific  $\Sigma_{\text{crg}}^{\text{prot}} \Delta H^{\text{U}}$ .

#### 4. Discussion and Conclusions

The energetic analysis of protein charged groups reported here was motivated by a surprising empirical correlation observed between the folded-state spatial distribution of charged groups and the residue-specific enthalpic stability of globular proteins.<sup>8</sup> We sought to determine if the enthalpic contribution of charged groups to  $\Delta H^{\text{U}}$  increasingly favors the folded state as the fraction of buried charged residues increases. If so, the tertiary structures of proteins with more uniform charge distributions should provide a basis for learning what is needed to reduce the substantial energetic penalty that arises from the removal of charged groups from an aqueous environment. Such information can be exploited in future protein design. A key outcome of this study is that the estimated enthalpy due to charged groups for different proteins was found to be inversely related to the excess surface charge  $\delta_{\text{xs}}$  (Figure 6), and thus the results support the hypothesis that buried charge lowers native protein enthalpy and a causal effect for the correlation.<sup>8</sup>

Since the publication by Murphy, Privalov, and Gill<sup>4</sup> correlating heat capacity with the entropy and enthalpy of protein unfolding, the basis for differences among globular proteins in the entropy per residue of unfolding is known to be the amount of buried apolar surface area.<sup>35</sup> Nevertheless, proteins with large amounts of buried apolar surface do not gain stability, regardless of the larger entropy favoring folding, and the origin of the compensating enthalpy is unclear.<sup>7,35,36</sup> Studies of compressibility<sup>8,37,38</sup> identified relationships with the unfolding enthalpy and entropy similar to those of heat capacity, and suggested a rationale based on charge distribution for variations in enthalpy of native, folded proteins.<sup>8</sup> The calculated energies reported here confirm this rationale by demonstrating that enthalpic stability

is gained by strong, favorable electrostatic interactions generated when charged groups are located in the interior as well as on the protein surface. Increasing numbers of buried charged groups necessarily oppose burying in the apolar surface, and thus changes in specific enthalpy by this mechanism would be compensated by opposing ones in the entropy per residue. Further, attributing differences in the specific thermodynamic properties of proteins to charged residues is reasonable because the variation in the fraction of charged residues is relatively large compared to that for uncharged polar and apolar residues.<sup>8</sup>

We find that a number of buried charged groups make significant, favorable contributions to structural stability. The main argument against structural stabilization by charges buried in a protein interior is the large penalty for desolvation and the experimental measures of free energy of transfer between water and organic solvents.<sup>19</sup> On the other hand, it is well recognized that protein interiors are not a homogeneous, hydrophobic medium; numerous uncharged polar groups and some charged polar groups compose electrostatic networks and specific local environments shielded from solvent.<sup>35,39</sup> The question is whether the local environments can compensate sufficiently for the loss of water around a charged group. This study finds that the set of buried charged groups from four globular proteins exhibits a large range of  $\Delta H_g^U$  values that span both favorable and unfavorable contributions to the total enthalpy of unfolding.  $\Delta H_g^U$  for individual charged groups range from 14 kcal/mol to  $-3$  kcal/mol (Tables 3S–6S, Supporting Information), with the average over all groups examined here equal to 2.1 kcal/mol (Table 2), near the value of 1.3 kcal/mol averaged over CH<sub>3</sub> apolar groups (Table 1S, Supporting Information). The different types of charged groups vary in their tendency for being buried; carboxylate groups are most likely to be buried while amino groups are the least likely. Moreover, the interactions of buried Glu and Asp carboxylate groups are found to have an overall lower energy than buried Arg and Lys basic groups (Table 2). Indeed, the most favorable contributions for COO<sup>-</sup> groups occur for fully buried groups, SE<sub>0–20</sub>. Thus, many environments are indeed sufficient to overcome the desolvation penalty, and the total contribution for all charged groups favors the folded state.

What is the explanation for the variations in  $\Delta H_g^U$  for a given type of charged group with similar solvent exposure? While a rigorous explanation is likely complex, we suggest that a simple assessment of the local electrostatic environment may explain how to accommodate a charged group upon protein folding. The number of charged groups and the net charge,  $Q_i^{\text{tot}}$ , was evaluated for a spherical region centered on the group of interest and of radius  $R = 5 \text{ \AA}$ :  $Q_i^{\text{tot}} = \sum q_j$  ( $R < 5 \text{ \AA}$ ). The conjecture is that a local  $Q_i^{\text{tot}} \approx 0$ , generated from two or more groups in addition to the central group, creates a “good neighborhood”, and the atomic group is likely to provide a positive contribution to the enthalpy of folding. This premise is supported by several examples in trypsin. Glu52 is buried, yet has one of the largest favorable contributions to the protein folded state enthalpy: +12 kcal/mol (Table 3S). A 5 Å-radius sphere surrounding this group includes five charged groups (Arg49, Asp53, Glu59, Glu62, and Arg99) and a Ca<sup>2+</sup> ion with a net charge of  $Q_i^{\text{tot}} = 0$ , making a good neighborhood for embedding a charged group. The carboxylate group of Glu62 from the same cluster of charges contributes +9 kcal/mol and its local neighborhood contains six charged groups (Arg49, Glu52, Asp53, Glu59, Arg99, Lys125) and a total charge  $Q_i^{\text{tot}} = -1e$ . On the other hand, Asp171 contributes  $-5$  kcal/mol, and this destabilization derives from six charged groups (Lys125, Lys136, Lys170, Asp176, and Lys200) within 5 Å, but with a

net positive charge of  $+2e$ . A more extensive analysis is underway to establish a quantitative relationship between the structural features of a local environment and the contribution to the folded state stability.

Experimental analysis of the effect of buried charged amino acids on protein structural stability has generated conflicting conclusions. Experimental mutagenesis of buried charged groups have an intrinsic limitation, in that substitution of a given residue, or residue pair for a salt bridge, can disrupt the structure and impact local interactions of many atomic groups, charged and uncharged. As a result, factors other than simple removal of charge come into play. In contrast, the computational approach from MD simulations taken here allowed a direct analysis of the energetics.

The occurrence of both positive and negative  $\Delta H_g^U$  values explains certain conflicting results from mutagenesis studies; depending on which salt bridges are mutated, measurements would indicate either stabilizing or destabilizing effects. In addition, engineering charged residues into a structure for stability likely requires long-range analysis of the full structure rather than simple pairwise mutation of proximal residues with appropriate geometry for substitution of a salt bridge. Our study suggests that engineering attempts to design buried charged groups into a protein<sup>24,40,41</sup> can benefit by consideration of the carefully crafted network of interactions that have evolved in native proteins with uniform charge distribution, perhaps by development of parameters like  $Q_i^{\text{tot}}$ .

**Acknowledgment.** This work was partially supported by grants from the U.S. National Institutes of Health (AI39639 to C. B. P.) and the Sloan Foundation and DOE (to V. M. D.).

**Supporting Information Available:** Tables listing additional enthalpic data derived in this study. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- Wang, Z.; Moul, J. *Hum. Mutat.* **2001**, *17*, 263–270.
- Partridge, A. W.; Therien, A. G.; Deber, C. M. *Proteins* **2004**, *54*, 648–656.
- Yue, P.; Li, Z.; Moul, J. *J. Mol. Biol.* **2005**, *353*, 459–473.
- Murphy, K. P.; Privalov, P. L.; Gill, S. J. *Science* **1990**, *247*, 559–561.
- Yang, A. S.; Sharp, K. A.; Honig, B. V. *J. Mol. Biol.* **1992**, *227*, 889–890.
- Xiao, L.; Honig, B. H. *J. Mol. Biol.* **1999**, *289*, 1435–1444.
- Lazaridis, T.; Archontis, G.; Karplus, M. *Adv. Protein Chem.* **1995**, *47*, 231–306.
- Dadarlat, V. M.; Post, C. B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14778–14783.
- Perutz, M. F. *Science* **1978**, *201*, 1187–1191.
- Serrano, L.; Kellis, J.; Cann, P.; Matouschek, A.; Fersht, A. S. *J. Mol. Biol.* **1992**, *224*, 783–804.
- Marqusee, S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8898–8902.
- Habermann, S. M.; Murphy, K. P. *Protein Sci.* **1996**, *5*, 1229–1239.
- Myers, J. K.; Pace, C. N. *Biophys. J.* **1996**, *71*, 2033–2039.
- Grimsley, G. R.; Shaw, K. L.; Fee, L. R.; Alston, R. W.; Huyghues-Despointes, B. M.; Thurlkill, R. L.; Scholtz, J. M.; Pace, C. N. *Protein Sci.* **1999**, *8*, 1843–1849.
- Ragone, R. *Protein Sci.* **2001**, *10*, 2075–2082.
- Dill, K. A. *Biochemistry* **1990**, *29*, 7133–7155.
- Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211–226.
- Honig, B. H.; Yang, A. S. *Adv. Protein Chem.* **1995**, *46*, 27–58.
- Honig, B. H.; Hubbel, W. L. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 5412–5416.
- Kumar, S.; Tsai, C. J.; Nussinov, R. *Biochemistry* **2002**, *41*, 5359–5374.
- Szilágyi, A.; Závodszy, P. *Structure* **1995**, *8*, 493–504.
- Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 2403–8.



- (23) Sun, D. P.; Sauer, U.; Nicholson, H.; Matthews, B. W. *Biochemistry* **1991**, *30*, 7143–7153.
- (24) Perl, D.; Mueller, U.; Heinemann, U.; Schmid, F. X. *Nat. Struct. Mol. Biol.* **2000**, *7*, 380–383.
- (25) Schneider, J. P.; Lear, J. D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 5742–5743.
- (26) Waldburger, C. D.; Schildbach, J. F.; Sauer, R. T. *Nat. Struct. Biol.* **1995**, *2*, 122–128.
- (27) King, J. L.; Jukes, T. H. *Science* **1969**, *164*, 788–798.
- (28) Dadarlat, V. M.; Post, C. B. *Biophys. J.* **2006**, *91*, 4544–4554.
- (29) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (30) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379–400.
- (31) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (32) Dadarlat, V. M.; Post, C. B. *J. Phys. Chem. B* **2001**, *105*, 715–724.
- (33) Dadarlat, V. M. *Biophys. J.* **2005**, *89*, 1433–1445.
- (34) Cheng, Y.; Rossky, P. *Nature* **1998**, *392*, 696–699.
- (35) Honig, B. H. *J. Mol. Biol.* **1999**, *293*, 283–293.
- (36) Makhatazde, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307–425.
- (37) Phelps, D. K.; Post, C. B. *J. Mol. Biol.* **1995**, *254*, 544–551.
- (38) Phelps, D. K.; Rossky, P. J.; Post, C. B. *J. Mol. Biol.* **1998**, *276*, 331–337.
- (39) Warshel, A. *J. Biol. Chem.* **1998**, *273*, 27035–27038.
- (40) Godoy-Ruiz, R.; Perez-Jimenez, R.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M. *J. Mol. Biol.* **2004**, *336*, 313–318.
- (41) Godoy-Ruiz, R.; Perez-Jimenez, R.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M. *Biophys. J.* **2005**, *89*, 3320–3331.

# Supplementary Material for “Contribution of Charged Groups to the Enthalpic Stabilization of the Folded States of Globular Proteins”

Voichita M. Dadarlat and Carol Beth Post\*

Department of Medicinal Chemistry and Molecular Pharmacology,  
Markey Center for Structural Biology, Purdue Cancer Center  
Purdue University, West Lafayette, IN 47907

November 18, 2007

---

\*This work was partially supported by grants from the U.S. National Institutes of Health (AI39639 to C. B. P.) and the Sloan Foundation and DOE (to V. M. D.)

## Table Captions for Supplementary Material

### Table 1S.

Enthalpy of unfolding and nonbonded energy for  $\text{CH}_3$  groups in apolar residues and dipeptides.

### Table 2S.

The “unfolded state” nonbonded energy (in kcal/mol), and time average solvent accessible surface areas,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ), of apolar and charged groups for solvated dipeptides and  $\text{Ca}^{2+}$ .

### Table 3S.

Enthalpic contribution of charged groups to protein folded state stability in TRPS. Non-bonded energy of group  $g$ ,  $\langle E_{nb}^g \rangle$ , (in kcal/mol), time average SASA,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ), Exposure to solvent,  $SE$ , (in %), and enthalpic gain upon folding,  $\Delta H_g^U$ , solvent reorganization,  $\Delta H_h^U$ , and total enthalpy change,  $\Delta H^U = \Delta H_g^U + \Delta H_h^U$  (in kcal/mol).

### Table 4S.

Enthalpic contribution of charged groups to protein folded state stability in RBNA. Non-bonded energy of group  $g$ ,  $\langle E_{nb}^g \rangle$ , (in kcal/mol); time average SASA,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ); Exposure to solvent,  $SE$ , (in %); and enthalpic gain upon folding,  $\Delta H_g^U$ , the change in enthalpy upon solvent reorganization,  $\Delta H_h^U$ , and the total enthalpy change,  $\Delta H^U = \Delta H_g^U + \Delta H_h^U$  (in kcal/mol).

### Table 5S.

Enthalpic contribution of charged groups to protein folded state stability in HLYSO. Non-bonded energy of group  $g$ ,  $\langle E_{nb}^g \rangle$ , (in kcal/mol); time average SASA,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ); exposure to solvent,  $SE$ , (in %); and enthalpic gain upon unfolding from environmental change,  $\Delta H_g^U$ , solvent reorganization,  $\Delta H_h^U$ , and the total change,  $\Delta H^U = \Delta H_g^U + \Delta H_h^U$  (in kcal/mol).

### Table 6S.

Enthalpic contribution of charged groups in ALACTA. Nonbonded energy of group  $g$ ,  $\langle E_{nb}^g \rangle$ , (in kcal/mol); time average SASA,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ); exposure to solvent,  $SE$ , (in %); enthalpic gain upon folding from charged groups,  $\Delta H_g^U$ , solvent reorganization,  $\Delta H_h^U$ , and the net change in enthalpy,  $\Delta H^U$  (in kcal/mol).

### Table 7S.

Contribution of solvent “reorganization” to the change in system enthalpy upon protein unfolding: the average number of water molecules in the first, second and third hydration



shells,  $N_w^{1st}$ ,  $N_w^{2nd}$ ,  $N_w^{3rd}$ ; the average nonbonded energy for each,  $\langle E_{nb}^{1st} \rangle$ ,  $\langle E_{nb}^{2nd} \rangle$ ,  $\langle E_{nb}^{3rd} \rangle$ ; the change in enthalpy of a water molecule,  $\Delta H_w^{1st}$ ,  $\Delta H_w^{2nd}$ ,  $\Delta H_w^{3rd}$ ; and the maximum change in solvent enthalpy upon complete charged group burial,  $\Delta H_h^{max}$ . Energies are in kcal/mol.  $\Delta H_w^{shell}$  is calculated as the difference between  $\langle E_{nb}^{1st} \rangle$  and the average nonbonded energy of a water molecule in the dipeptide solution, -9.2 kcal/mol.

**Table 8S.**

Enthalpy of unfolding - change in local environment of charged group upon protein unfolding,  $\Delta H_g^U$  (kcal/mol), summed over charged groups of a given type as a function of exposure to solvent, SE, (in %). Group averages,  $\overline{\Delta H_g^U}$ , are averages over each charged group type in the indicated SE range, from 4 proteins. Global averages,  $\widetilde{\Delta H_g^U}$ , are averages over all charged groups in 4 proteins in the indicated SE range.

**Table 9S.**

Enthalpy of unfolding - contribution from solvent reorganization and structural collapse,  $\Delta H_h^U$  (kcal/mol), summed over charged groups of a given type as a function of exposure to solvent, SE, (in %). Group averages,  $\overline{\Delta H_h^U}$ , are averages over each charged group type in the indicated SE range, from 4 proteins. Global averages,  $\widetilde{\Delta H_h^U}$ , are averages over all charged groups in 4 proteins in the indicated SE range.

Table 1: Enthalpy of folding and nonbonded energy for terminal CH<sub>3</sub> groups in apolar amino acids and their analog dipeptides

Group	$\langle E_{nb}^{g,dipep(g)} \rangle$	$\langle E_{nb}^{g,prot} \rangle$	$\Delta H_g^U$
CH <sub>3</sub> (alad)	+2.2	-	-
C $\delta_1$ ( $\delta_2$ )H <sub>3</sub> (vald)	-5.5	-	-
C $\gamma_1$ ( $\gamma_2$ )H <sub>3</sub> (leud)	-13.1	-	-
C $\gamma_2$ H <sub>3</sub> (iled)	+1.2	-	-
CH <sub>3</sub> (ALA)		.2 - .8	.8 - 1.5
CH <sub>3</sub> (VAL)		-5.5 - 6.6	.0 - .6
CH <sub>3</sub> (ILE)		.8 - -2.6	+1.0 - 1.9
CH <sub>3</sub> (LEU)		-14.2 - -15.2	.6 - 1.1
Average			$\propto$ +1.3

Table 2: The “unfolded state” nonbonded energy (in kcal/mol), and time average solvent accessible surface areas,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ), of apolar and charged groups for solvated dipeptides and  $\text{Ca}^{2+}$ .

Group	$\langle E_{nb}^{g,dipep(g)} \rangle$	$\langle SA^{g,dipep(g)} \rangle$
$\text{CH}_3(\text{alad})$	+2.25	77
$\text{C}_{\gamma_1}\text{H}_3(\text{vald})$	-5.6	68
$\text{C}_{\gamma_2}\text{H}_3(\text{vald})$	-5.2	63
$\text{C}_{\delta_1}\text{H}_3(\text{leud})$	-13.2	71
$\text{C}_{\delta_2}\text{H}_3(\text{leud})$	-13.2	70
$\text{C}_{\delta}\text{H}_3(\text{iled})$	+1.23	77
$\text{C}_{\gamma_2}\text{H}_3(\text{iled})$	-1.82	61
$\text{COO}^-(\text{aspd})$	-201	97
$\text{COO}^-(\text{glud})$	-184	98
$\text{Gu}(\text{argd})$	-72	142
$\text{NH}_3^+(\text{lysd})$	-86	59
$\text{Ca}^{2+}$	-606	122



Table 3: Enthalpic contribution of charged groups to protein folded state stability in TRPS. Nonbonded energy of group  $g$ ,  $\langle E_{nb}^g \rangle$ , (in kcal/mol), time average SASA,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ), Exposure to solvent,  $SE$ , (in %), and enthalpic gain upon folding,  $\Delta H_g^U$ , solvent reorganization,  $\Delta H_h^U$ , and total enthalpy change,  $\Delta H^U = \Delta H_g^U + \Delta H_h^U$  (in kcal/mol).

Group	$\langle E_{nb}^g \rangle$	$\langle SA \rangle$	SE	$\Delta H_g^U$	$\Delta H_h^U$	$\Delta H^U$
COO <sup>-</sup> (ASP84)	-210.5	0	0	+4.8	2.4	7.2
COO <sup>-</sup> (ASP176)	-216.0	0	0	+7.5	2.4	9.9
COO <sup>-</sup> (ASP171)	-191	2	2	-5	2.4	-2.6
COO <sup>-</sup> (ASP145)	-227.5	52	52	+13.3	1.2	14.5
COO <sup>-</sup> (ASP133)	-200.5	79	81	-.3	.5	.2
COO <sup>-</sup> (ASP53)	-201.5	17	18	+.3	1.2	1.5
COO <sup>-</sup> (GLU62)	-202.0	0	0	+9	1.7	10.7
COO <sup>-</sup> (GLU52)	-208.0	0	0	+12	1.7	13.7
COO <sup>-</sup> (GLU59)	-191	44	45	+3.5	.9	4.4
COO <sup>-</sup> (GLU167)	-185.8	99	100	+.9	0.0	.9
Gu(ARG49)	-69.5	37	26	-1.3	1.4	.1
Gu(ARG99)	-74.8	94	67	+1.4	.6	2.0
NH <sub>3</sub> <sup>+</sup> (LYS43)	-82.5	18	31	-1.8	.5	-1.3
NH <sub>3</sub> <sup>+</sup> (LYS69)	-84.2	37	63	-.9	.4	-.5
NH <sub>3</sub> <sup>+</sup> (LYS89)	-91.1	27	46	+2.6	.6	3.2
NH <sub>3</sub> <sup>+</sup> (LYS91)	-79.6	40	68	-3.2	.4	-2.8
NH <sub>3</sub> <sup>+</sup> (LYS125)	-87.3	54	92	+.7	.1	.8
NH <sub>3</sub> <sup>+</sup> (LYS136)	-84.0	29	49	-1.0	.6	-.4
NH <sub>3</sub> <sup>+</sup> (LYS139)	-87.4	50	84	+.7	.2	.9
NH <sub>3</sub> <sup>+</sup> (LYS149)	-89	10	16	+1.5	.9	2.4
NH <sub>3</sub> <sup>+</sup> (LYS170)	-85.6	39	66	-.2	.4	.2
NH <sub>3</sub> <sup>+</sup> (LYS186)	-86.3	32	54	+.2	.5	.7
NH <sub>3</sub> <sup>+</sup> (LYS200)	-86.2	36	62	+.1	.4	.5
NH <sub>3</sub> <sup>+</sup> (LYS202)	-85.7	38	66	-.2	.4	.2
NH <sub>3</sub> <sup>+</sup> (LYS208)	-90.3	28	48	+2.2	.6	2.8
NH <sub>3</sub> <sup>+</sup> (LYS217)	-89.4	45	76	+1.7	.3	2.0

Table 4: Enthalpic contribution of charged groups to protein folded state stability in RBNA. Nonbonded energy of group  $g$ ,  $\langle E_{nb}^g \rangle$ , (in kcal/mol); time average SASA,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ); Exposure to solvent,  $SE$ , (in %); and enthalpic gain upon folding,  $\Delta H_g^U$ , the change in enthalpy upon solvent reorganization,  $\Delta H_h^U$ , and the total enthalpy change,  $\Delta H^U = \Delta H_g^U + \Delta H_h^U$  (in kcal/mol)

Group	$\langle E_{nb}^g \rangle$	$\langle SA \rangle$	SE	$\Delta H_g^U$	$\Delta H_h^U$	$\Delta H^U$
COO <sup>-</sup> (ASP14)	-197.7	14	14	-1.7	2.1	.4
COO <sup>-</sup> (ASP38)	-220.8	59	61	+9.9	.9	10.8
COO <sup>-</sup> (ASP53)	-199.0	61	63	-1.0	.9	-.1
COO <sup>-</sup> (ASP83)	-225.4	22	23	+12.2	1.8	14.0
COO <sup>-</sup> (ASP121)	-212.5	8	8	+5.8	2.2	8.0
COO <sup>-</sup> (GLU2)	-209.5	28	29	+12.8	1.2	14.0
COO <sup>-</sup> (GLU9)	-191.0	82	84	+3.5	.3	3.8
COO <sup>-</sup> (GLU49)	-177.9	41	42	-3.1	1.0	-2.1
COO <sup>-</sup> (GLU86)	-192	38	38	+4	1.1	5.1
COO <sup>-</sup> (GLU111)	-178.8	79	80	-2.6	.3	-2.3
Gu(ARG10)	-73.8	26	18	+.9	1.6	2.5
Gu(ARG33)	-75.8	33	23	+1.9	1.5	3.4
Gu(ARG39)	-71.7	101	72	-.2	.5	.3
Gu(ARG85)	-75.4	61	43	+1.7	1.1	2.8
NH <sub>3</sub> <sup>+</sup> (LYS1)	-82.6	58	100	-1.7	0.0	-1.7
NH <sub>3</sub> <sup>+</sup> (LYS7)	-81.6	9	15	-2.2	.9	-1.3
NH <sub>3</sub> <sup>+</sup> (LYS31)	-83.5	44	75	-1.3	.3	-1.0
NH <sub>3</sub> <sup>+</sup> (LYS37)	-91.3	29	50	+2.7	.6	3.3
NH <sub>3</sub> <sup>+</sup> (LYS41)	-77.9	15	25	-4.1	.8	-3.3
NH <sub>3</sub> <sup>+</sup> (LYS61)	-91.8	48	82	+2.9	.2	3.1
NH <sub>3</sub> <sup>+</sup> (LYS66)	-85.4	30	51	-.3	.5	.2
NH <sub>3</sub> <sup>+</sup> (LYS91)	-82.0	33	55	-2.0	.5	-1.5
NH <sub>3</sub> <sup>+</sup> (LYS98)	-84.3	56	95	-.9	.1	-.8
NH <sub>3</sub> <sup>+</sup> (LYS104)	-92.4	35	59	+3.7	.5	4.2

Table 5: Enthalpic contribution of charged groups to protein folded state stability in HLYSO. Nonbonded energy of group  $g$ ,  $\langle E_{nb}^g \rangle$ , (in kcal/mol); time average SASA,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ); exposure to solvent,  $SE$ , (in %); and enthalpic gain upon unfolding from environmental change,  $\Delta H_g^U$ , solvent reorganization,  $\Delta H_h^U$ , and the total change,  $\Delta H^U = \Delta H_g^U + \Delta H_h^U$  (in kcal/mol).

Group	$\langle E_{nb}^g \rangle$	$\langle SA \rangle$	SE	$\Delta H_g^U$	$\Delta H_h^U$	$\Delta H^U$
COO <sup>-</sup> (ASP48)	-218.0	25	25	+8.5	1.8	10.3
COO <sup>-</sup> (ASP66)	-216.0	1	0	+7.5	2.4	9.9
COO <sup>-</sup> (ASP18)	-201.5	36	37	+.3	1.5	1.8
COO <sup>-</sup> (ASP52)	-191.0	20	21	-5.0	1.9	-3.1
COO <sup>-</sup> (ASP87)	-213.3	48	49	+6.2	1.2	7.4
COO <sup>-</sup> (ASP101)	-212.5	55	56	+5.8	1.1	6.9
COO <sup>-</sup> (ASP119)	-202.0	61	63	+.5	.9	1.4
COO <sup>-</sup> (GLU7)	-181.2	61	62	-1.4	.6	-.8
COOH(GLU35)	-52.3	16	-	-		
Gu(ARG5)	-68.2	45	32	-1.9	1.6	-.3
Gu(ARG14)	-74.9	136	97	+1.5	.1	1.6
Gu(ARG21)	-72.3	101	72	+.2	.7	.9
Gu(ARG45)	-73.4	78	55	+.7	1.1	1.8
Gu(ARG61)	-78.2	27	19	+3.1	1.9	5.0
Gu(ARG68)	-69.6	128	91	-1.2	.2	-1.0
Gu(ARG73)	-68.4	91	65	-1.8	.8	-1.0
Gu(ARG112)	-74.3	73	52	+1.2	1.2	2.4
Gu(ARG114)	-74.2	97	69	+1.1	.7	1.8
Gu(ARG125)	-76.8	67	48	+2.4	1.2	3.6
Gu(ARG128)	-67.4	140	100	-2.2	0	-2.2
NH <sub>3</sub> <sup>+</sup> (LYS13)	-87.8	38	65	+.9	.4	1.3
NH <sub>3</sub> <sup>+</sup> (LYS33)	-82.3	41	70	-1.9	.3	-1.6
NH <sub>3</sub> <sup>+</sup> (LYS96)	-82.2	37	62	-1.9	.4	-1.5
NH <sub>3</sub> <sup>+</sup> (LYS97)	-89.0	38	65	+1.5	.4	1.9
NH <sub>3</sub> <sup>+</sup> (LYS116)	-88.1	35	59	+1.1	.5	1.6



Table 6: Enthalpic contribution of charged groups in ALACTA. Nonbonded energy of group  $g$ ,  $\langle E_{nb}^g \rangle$ , (in kcal/mol); time average SASA,  $\langle SA \rangle$  (in  $\text{\AA}^2$ ); exposure to solvent,  $SE$ , (in %); enthalpic gain upon folding from charged groups,  $\Delta H_g^U$ , solvent reorganization,  $\Delta H_h^U$ , and the net change in enthalpy,  $\Delta H^U$  (in kcal/mol).

Group	$\langle E_{nb}^g \rangle$	$\langle SA \rangle$	SE	$\Delta H_g^U$	$\Delta H_h^U$	$\Delta H^U$
COO <sup>-</sup> (ASP15)	-210.2	63	64	+4.6	.9	5.5
COO <sup>-</sup> (ASP38)	-203.5	39	40	+1.3	1.5	2.8
COO <sup>-</sup> (ASP47)	-217.5	34	34	+8.3	1.6	9.9
COO <sup>-</sup> (ASP65)	-189.0	84	86	-6.0	.3	-5.7
COO <sup>-</sup> (ASP79)	-204.0	32	33	+1.5	1.6	3.1
COO <sup>-</sup> (ASP83)	-218.4	13	13	+8.7	2.1	10.8
COO <sup>-</sup> (ASP84)	-189.4	63	64	-5.8	.9	-4.9
COO <sup>-</sup> (ASP85)	-192.2	44	64	-4.4	.9	-3.5
COO <sup>-</sup> (ASP88)	-218.0	16	17	+8.5	2.0	10.5
COO <sup>-</sup> (ASP89)	-213.3	0	0	+6.2	2.4	8.6
COO <sup>-</sup> (ASP117)	-201.4	52	53	+.2	1.1	1.3
COO <sup>-</sup> (GLU2)	-181.0	96	98	-1.5	0.0	-1.5
COO <sup>-</sup> (GLU8)	-182.9	44	45	-.6	.9	.3
COO <sup>-</sup> (GLU12)	-180.8	31	32	-1.6	1.2	-.4
COO <sup>-</sup> (GLU26)	-188.6	36	36	+2.3	1.1	3.3
COO <sup>-</sup> (GLU50)	-189.5	50	51	+2.8	.8	3.6
COO <sup>-</sup> (GLU114)	-181.4	84	86	-1.3	.2	-1.1
COO <sup>-</sup> (GLU122)	-178	83	85	-3.0	.3	-2.7
Gu(ARG11)	-82.4	141	100	+5.2	0.0	5.2
NH <sub>3</sub> <sup>+</sup> (LYS6)	-82.7	40	68	-1.7	.4	-1.3
NH <sub>3</sub> <sup>+</sup> (LYS14)	-91.5	37	62	+2.8	.4	3.2
NH <sub>3</sub> <sup>+</sup> (LYS17)	-89.2	47	80	+1.6	.2	1.8
NH <sub>3</sub> <sup>+</sup> (LYS59)	-89.8	33	56	+1.9	.5	2.4
NH <sub>3</sub> <sup>+</sup> (LYS63)	-84.0	2.6	4	-1.0	1.1	.1
NH <sub>3</sub> <sup>+</sup> (LYS80)	-88.0	40	68	+2.0	.4	2.4
NH <sub>3</sub> <sup>+</sup> (LYS94)	-82.6	36	60	-1.7	.4	-1.3
NH <sub>3</sub> <sup>+</sup> (LYS95)	-78	16	27	-4.0	.8	-3.2
NH <sub>3</sub> <sup>+</sup> (LYS109)	-81.4	50	85	-2.3	.2	-2.1
NH <sub>3</sub> <sup>+</sup> (LYS115)	-82.6	33	55	-1.7	.5	-1.2
NH <sub>3</sub> <sup>+</sup> (LYS123)	-80.1	37	62	-3.0	.4	-2.6

Table 7: Contribution of solvent “reorganization” to the change in system enthalpy upon protein unfolding: the average number of water molecules in the first, second and third hydration shells,  $N_w^1$ ,  $N_w^2$ ,  $N_w^3$ ; the average nonbonded energy for each,  $E_{nb}^1$ ,  $E_{nb}^2$ ,  $E_{nb}^3$ ; the change in enthalpy of a water molecule for each hydration shell,  $\Delta H_w^1$ ,  $\Delta H_w^2$ ,  $\Delta H_w^3$ ; and the maximum change in solvent enthalpy upon complete charged group burial,  $\Delta H_h^{max}$ . Energies are in kcal/mol. For each shell,  $\Delta H_w^{sh}$  is calculated as the difference between  $E_{nb}^{sh}$  and the average nonbonded energy of a water molecule in the dipeptide solution, -9.2 kcal/mol.

Group	$N_w^1$	$E_{nb}^1$	$\Delta H_w^1$	$N_w^2$	$E_{nb}^2$	$\Delta H_w^2$	$N_w^3$	$E_{nb}^3$	$\Delta H_w^3$	$\Delta H_h^{max}$
glud	4.5	-12.3	-3.1	17.0	-8.6	.6	27	-9.0	.2	1.7
aspd	4.1	-11.7	-2.5	13.4	-8.7	.5	27	-9.0	.2	2.4
lysd	4.1	-11.5	-2.3	15.6	-8.7	.5	28	-9.1	.1	1.1
argd	4.9	-10.3	-1.1	18.1	-8.8	.4	27.8	-9.2	0.0	1.9
vald	5.7	-8.7	.5	9.8	-9.7	-.5	22	-9.0	.2	2.4

Table 8: Enthalpy of unfolding - change in local environment of charged group upon protein unfolding,  $\Delta H_g^U$  (kcal/mol), summed over charged groups of a given type as a function of exposure to solvent, SE, (in %). Group averages,  $\overline{\Delta H_g^U}$ , are averages over each charged group type in the indicated SE range, from 4 proteins. Global averages,  $\widetilde{\Delta H_g^U}$ , are averages over all charged groups in 4 proteins in the indicated SE range.

Group	Protein	$\Delta H_g^U$					
		0-100	0-20	20-40	40-60	60-80	80-100
g= COO <sup>-</sup> (ASP)	TRPS	+20.6	+7.6	0	+13.3	0	-.3
	RBNA	+25.2	+4.1	+12.2	0	+8.9	0
	HLYSO	+23.0	+7.5	+3.5	+12.0	0	0
	ALACTA	+23.1	+23.4	+11.1	+2	-5.6	-6
	$\Sigma \Delta H_g^U$	+91.9	+42.6	+26.8	+25.5	+3.3	-6.3
	$\Sigma N^g$	29	10	7	4	6	2
	$\overline{\Delta H_g^U}$	+3.2	+4.3	+3.8	+6.4	.6	-3.2
g= COO <sup>-</sup> (GLU)	TRPS	+25.4	+21	0	+3.5	0	+9
	RBNA	+14.6	0	+16.8	-3.1	-2.6	+3.5
	HLYSO	-1.4	0	0	0	-1.4	0
	ALACTA	-2.9	0	+7	+2.2	0	-5.8
	$\Sigma \Delta H_g^U$	+35.7	+21.0	+17.5	+2.6	-4.0	-1.4
	$\Sigma N^g$	17	2	4	4	2	5
	$\overline{\Delta H_g^U}$	+2.1	+10.5	+4.4	+7	-2.0	-3
g= Gu(ARG)	TRPS	+1	0	-1.3	0	+1.4	0
	RBNA	+4.3	+9	+1.9	+1.7	-.2	0
	HLYSO	+3.1	+3.1	-1.9	+4.3	-.5	-1.9
	ALACTA	+5.2	0	0	0	0	+5.2
	$\Sigma \Delta H_g^U$	+12.7	+4.0	-1.3	+6	+7	+4.3
	$\Sigma N^g$	18	2	3	4	5	4
	$\overline{\Delta H_g^U}$	+7	+2	-.4	+1.5	+1	+1.1
g= NH <sub>3</sub> <sup>+</sup> (LYS)	TRPS	+2.4	+1.5	-1.8	+4.0	-2.7	+1.4
	RBNA	-3.8	-2.2	-4.1	+3.6	-1.3	+3
	HLYSO	-.3	0	0	+1.1	-1.4	0
	ALACTA	-7.1	-1.0	-4.0	-1.5	+1.7	-2.3
	$\Sigma \Delta H_g^U$	-8.8	-1.7	-9.9	+6.6	-3.7	-.1
	$\Sigma N^g$	40	3	3	11	17	6
	$\overline{\Delta H_g^U}$	-.2	-.6	-3.3	+7	-.2	-.1
All	$\Sigma \Sigma \Delta H_g^U$	131.5	65.9	33.1	38.1	-3.7	-3.5
All	$\Sigma \Sigma N^g$	104	117	17	23	30	17
Ave All	$\widetilde{\Delta H_g^U}$	1.26	+3.9	1.9	+1.7	-.1	-.2



Table 9: Enthalpy of unfolding - contribution from solvent reorganization and structural collapse,  $\Delta H_h^U$  (kcal/mol), summed over charged groups of a given type as a function of exposure to solvent, SE, (in %). Group averages,  $\overline{\Delta H_h^U}$ , are averages over each charged group type in the indicated SE range, from 4 proteins. Global averages,  $\widetilde{\Delta H_h^U}$ , are averages over all charged groups in 4 proteins in the indicated SE range.

Group	Protein	$\Delta H_h^U$					
		0-100	0-20	20-40	40-60	60-80	80-100
g= COO <sup>-</sup> (ASP)	TRPS	10.1	8.4	0	1.2	0	.5
	RBNA	7.9	4.3	1.8	.9	.9	0
	HLYSO	10.8	2.4	5.2	2.3	0.9	0
	ALACTA	15.3	6.5	4.7	1.1	2.7	.3
	$\Sigma \Delta H_h^U$	43.1	21.6	11.7	5.5	+3.5	.8
	$\Sigma N^g$	29	10	7	4	6	2
	$\overline{\Delta H_h^U}$	1.48	2.2	1.7	1.4	.6	.4
g= COO <sup>-</sup> (GLU)	TRPS	4.3	3.4	0	.9	0	0
	RBNA	3.9	0	2.3	1.0	.3	.3
	HLYSO	.6	0	0	0	.6	0
	ALACTA	4.5	0	2.3	1.7	0	.5
	$\Sigma \Delta H_h^U$	13.1	3.4	4.6	3.6	1.7	.8
	$\Sigma N^g$	17	2	4	4	2	5
	$\overline{\Delta H_h^U}$	.8	1.7	1.2	.8	.9	.2
g= Gu(ARG)	TRPS	2.0	0	1.4	0	.6	0
	RBNA	4.7	1.6	1.5	1.1	.5	0
	HLYSO	9.5	1.9	1.6	3.5	2.2	.3
	ALACTA	0	0	0	0	0	0
	$\Sigma \Delta H_h^U$	16.2	3.5	4.5	4.6	3.3	.3
	$\Sigma N^g$	18	2	3	4	5	4
	$\overline{\Delta H_h^U}$	.9	1.8	1.5	1.2	.7	.1
g= NH <sub>3</sub> <sup>+</sup> (LYS)	TRPS	6.3	.9	.5	2.3	2.5	.3
	RBNA	4.4	.9	.8	2.1	.3	+ .3
	HLYSO	2.0	0	0	.5	1.5	0
	ALACTA	5.3	1.1	.8	1.4	1.8	.2
	$\Sigma \Delta H_h^U$	18.2	2.9	2.1	6.3	5.3	.8
	$\Sigma N^g$	40	3	3	11	17	6
	$\overline{\Delta H_h^U}$	.5	1.0	.7	.6	.3	.1
All	$\Sigma \Delta H_h^U$	90.6	31.4	22.9	20.0	13.8	2.7
All	$\Sigma N^g$	104	17	17	23	30	17
All Ave	$\widetilde{\Delta H^U}$	.9	1.8	1.3	.9	.03	.16