Electrostatic Contributions to Protein Stability and Folding Energy

Maite Roca, Benjamin Messer and Arieh Warshel*

Department of Chemistry, University of Southern California, 418 SGM Building, 3620 McClintock Avenue, Los Angeles, CA, 90089-1062, USA

* To whom correspondence should be address. E-mail: warshel@usc.edu

Phone: +1 213 740 4114; Fax +1 213 740 2701

Abstract

The ability to predict the thermal stability of proteins based on their corresponding sequence is a problem of great fundamental and practical importance. Here we report an approach for calculating the electrostatic contribution to protein stability based on the use of the semimacroscopic protein dipole Langevin dipole (PDLD/S) in its linear response approximation version for self energy with a dielectric constant, (ε_p) and an effective dielectric for charge-charge interactions (ε_{eff}). The method is applied to the test cases of ubiquitin, lipase, dihydrofolate reductase and cold shock proteins with series of ε_p and ε_{eff}. It is found that the optimal values of these dielectric constants lead to very promising results, both for the relative stability and the absolute folding energy. Consideration of the specific values of the optimal dielectric constants leads to an exciting conceptual description of the reorganization effect during the folding process. Although this description should be examined by further microscopic studies, the practical use of the current approach seems to offer a powerful tool for protein design and for studies of the energetics of protein folding.

Keywords: Protein stability; Folding Energy; Dielectric constants; Electrostatics in proteins.

Abbreviations: PDLD/S-LRA, protein dipole Langevin dipole/semimacroscopic with the linear response approximation; Lip A, lipase; WT, wild type; EcDHFR, dihydrofolate reductase from *Escherichia coli*; TmDHFR, dihydrofolate reductase from *Thermotoga maritime*; Bs-Csp, cold shock protein from *Bacillus subtilis*; Bc-Csp, cold shock protein from *Bacillus caldolyticus*; Tm-Csp, cold shock protein from *Thermotoga maritime*.

I. Introduction

Understanding the factors that determine the thermal stability of proteins presents a fundamental and practical challenge (e.g. [1]). One of the outcomes of a better understanding of protein stability should be the ability to predict the trend in stability within related proteins or between different mutants of the same protein. Unfortunately, despite the progress in the development of models for studying the folding of proteins [2-10] we still have major problems in predicting protein stability by either microscopic or macroscopic models. More specifically, despite the impressive progress in studies of protein folding, we still lack a clear understanding of the contributions of electrostatic effects to thermal stability and to the overall folding free energy. Experimental studies of mesophilic, thermophilic, and hyperthermophilic proteins have provided an excellent benchmark for studies in this area [11]. In general, the number of ionizable residues increases in hyperthermophiles, indicating that charged residues can be considered to be a stabilizing factor. However, some continuum studies have suggested that charged and polar groups lead to destabilization [12,13]. It was also suggested [12] that internal salt bridges tend to destabilize proteins, although as discussed in ref [14], this study did not reproduce the relevant observed energies. Other studies (e.g.[15,16]) appear to support the idea that charged residues can help to optimize protein stability.

The difficulty in reaching clear conclusions on the role of electrostatic interactions in protein stability is associated with the fact that we have a competition between desolvation penalties, stabilization by local protein dipoles, and charge-charge interactions. In the original Tanford-Kirkwood (TK) model of a non polar protein [17], both isolated ions and ion pairs should become unstable in the “protein” interior [18].

However, the situation is much more complex in real proteins, where charges are stabilized by polar groups [18]. Here the balance between charge-charge interactions and self-energy can depend drastically on the assumed ε_p (e.g. [14]).

This work attempts to quantify the electrostatic contribution to protein stability and to determine its relationship to the assumed dielectric constants. It is found that we can obtain promising results while using optimal effective dielectric constants for charge-charge interactions (ε_{eff}) and for self energy (ε_p). Furthermore, examining the physical basis for the particular values of ε_{eff} and ε_p to point toward an exciting new picture.

II. Methods

Our starting point is the (A')→(C) folding process of Fig. 1. In this process we start by folding the uncharged protein and then continue by moving the charges from reference groups in water to the same groups in the protein. Assuming that the folding energy for the uncharged protein is similar for all mutants that involve changes of charged groups, we can focus on the electrostatic contribution to folding ΔG^{elec}_{fold}. In this case we can use the general expression for the electrostatic energy of different ionization states of a given protein at a given pH [19], and obtain [14],

$$\Delta G^{elec}_{fold} = \Delta G^{elec}_{f} - \Delta G^{elec}_{uf} = -2.3RT\sum_{i} Q_i^{(f)}\left(pK^{p}_{i,int}(\varepsilon_p) - pH\right) + 166\sum_{i\neq j}\frac{Q_i^{(f)}Q_j^{(f)}}{r_{ij}^{(f)}\varepsilon_{eff}(r_{ij}^{(f)})}$$

$$-\sum_{i} Q_i^{(uf)}\left(pK^{w}_{a,i} - pH\right) - 166\sum_{i\neq j}\frac{Q_i^{(uf)}Q_j^{(uf)}}{80 r_{ij}^{(uf)}} \qquad (1)$$

where *uf* and *f* designate, respectively, the unfolded and folded states. Here, Q_i is the charge of the i^{th} residue, $pK^{p}_{i,int}$ is the intrinsic pK_a ($pK_{a,int}$) of the i^{th} residue in its given

protein state when all other residues are neutral, ε_p is the dielectric constant used in the semimacroscopic calculation of $pK_{a,int}$, r_{ij} is the distance between residues i and j, and ε_{eff} is the effective dielectric for charge-charge interactions. The nature of ε_p and ε_{eff} is far from trivial and cannot be fully clarified in this short communication. Thus, it is recommended that readers who are unfamiliar with these dielectrics read the discussion in ref. [14,20]. Overall, ε_p determines the intrinsic pK_a and represents the "self energy" of each charged group. This parameter is not related to the response of the protein to electric field but to the method used in the calculations and to the elements included in the simulation system. Basically, ε_p reflects all the effects that are not included explicitly in the given model [20]. On the other hand, ε_{eff} is a phenomenological parameter that represents the free energy of charge-charge interactions. This parameter reflects the compensation of the gas phase charge-charge interaction by the reorganization of the solvent and the protein [14].

Of course the folding free energy includes nonelectrostatic contributions such as configurational entropy and hydrophobic contributions and these contributions depend on the path used in Fig. 1. For example, we can write according to Fig. 1,

$$\Delta G_{fold} = \Delta G_{f}^{elec} - \Delta G_{uf}^{elec} + \Delta G_{uf\rightarrow f}^{uncharged} = \Delta G_{fold}^{elec} + \Delta G_{uf\rightarrow f}^{uncharged} \qquad (2)$$

In this description the electrostatic terms represent the charging process in the folded state and the uncharging in the unfolded state, while the non electrostatic is entirely associated with the folding of the fully uncharged protein. The use of this equation for mutations of ionized residues of a given protein allows us to focus only on ΔG_{fold}^{elec}.

The implementation of Eq (2) requires one to define the structure of the uncharged folded protein (B). That is, in order to eliminate the nonelectrostatic contributions in studies of mutations of the same protein we need to use the same structure (B in Fig. 1) for the uncharged folded state of all mutants (typically a structure near that of the wild type (WT) protein). This can be done by imposing a small constraint that would force all the mutants to stay near the WT structure in the charging step. However, more consistent electrostatic calculations would start in each case from the structure near that of the charged mutant (B' in Fig. 1). Apparently, the ε_p needed to reproduce the observed pK_a may be different for B and B' (see below).

Now, if we force Eq. (1) to reproduce the *change* in folding energy of different mutants by focusing on the B to C part of the cycle we can use the approximation:

$$\Delta G^{elec}_{fold} \approx -2.3RT\sum_i Q_i\left((pK^{p}_{i,int}(\varepsilon_p'))'-pK^{w}_{i,w}\right)+166\sum_{i\neq j} Q_iQ_j\left[\frac{1}{r_{ij}^{(f)}\varepsilon_{eff}^{(f)}}-\frac{1}{80r_{ij}^{(uf)}}\right] \qquad (3)$$

The first term in Eq. 3 represents the change of self-energy upon moving a charge from water to its site in the folded protein, and the second term represents the effect of charge-charge interaction. Here we use ε_p' to designate the fact that the calculations of the intrinsic pK_a of the different residues are done at different structures for the WT and mutant systems (B' in Fig 1). We chose this treatment since it seems to give the best results. However, now we have $(pK_{a.int})$' instead of $pK_{a,int}$ designating the fact that now the hypothetical $(pK_{a.int})$' should include implicitly the $\Delta G^{uncharged}_{f\rightarrow f'}$ contribution. We also assume that the last term in Eq. (3) is neglected since $r_{ij}^{(uf)}$ is usually larger than $r_{ij}^{(f)}$ and $\varepsilon_{eff}^{(f)}$ is smaller than 80. Furthermore, we assumed for simplicity that the same groups are ionized in the folded state and unfolded state.

It is useful to clarify for the benefit of the subsequent discussion that the first term is given by [18],

$$-2.3\mathrm{RT}\sum_{\mathrm{i}}\mathrm{Q}_i^{(f)}\left(\mathrm{pK}_{i,\mathrm{int}}^{p}(\varepsilon_p)-\mathrm{pK}_{\mathrm{a,w}}^{\mathrm{w}}\right)=\Delta\Delta\mathrm{G}_{sol}^{w\to p}(\mathrm{Q}_i=0\to\mathrm{Q}_i=\mathrm{Q}_i^0) \tag{4}$$

The solvation energy follows the trend of the Born´s energy in the simple case of a charge in the center of an hypothetical non polar protein (see discussion in e.g [18,20]).

Obviously, the non electrostatic term may well be different for different proteins. However, if we use the (A) → (C) path of Fig. 1 we can get a different picture since now the entire folding process can be described as an electrostatic process where,

$$\Delta\mathrm{G}_{fold}\cong\Delta\mathrm{G}_{uf\to f}^{charged}\cong 166\sum_{i\neq j}\frac{Q_iQ_j}{\varepsilon_{eff}}\left[\frac{1}{r_{ij}^{(f)}}-\frac{1}{r_{ij}^{(uf)}}\right] \tag{5}$$

That is, now the nonelectrostatic effects may be absorbed in the effective dielectric for bringing the charged groups from the unfolded to the folded state in the (A) → (C) path of Fig. 1. Of course, this assumption might be a poor approximation, but it clearly deserves serious examination.

The first step in the evaluations of Eq. (3) is the calculations of the $pK_{a,int}$'s. This is done by using the semimacroscopic Protein Dipole Langevin Dipole approach with the linear response approximation (PDLD/S-LRA) according to standard protocol using the POLARIS module of MOLARIS program [21] (see also [14] and references therein). After evaluating the $pK_{a,int}$'s we determine the ionization states of the different residues at the given pH (here we perform the calculations at pH=7) using the Monte Carlo approach described elsewhere (e.g.[21]). This procedure allows us to evaluate the $pK_{a,int}$ and Q for

different values of ε_p and ε_{eff} and then to use Eq. 2 to calculate ΔG^{elec}_{fold} as a function of these dielectric constants.

The proteins studied here were first solvated by the surface constrained all atom solvent (SCAAS) model [21] and all the groups that become ionized at pH=7 were assigned a charge which is 50% of their full charge at the ionized state (this was considered as the optimal procedure for the initial relaxation). The resulting system was equilibrated by running a 100 ps molecular dynamics simulation with 1fs time step at 300 K. Next we equilibrated each system (e.g. each mutant) by running an additional 10 ps simulation. After that, we evaluated the $pK_{a,int}$ and the average charge using the PDLD/S-LRA approach by averaging the corresponding values over the results obtained for 25 protein configurations (for the charged and uncharged state) each averaged over 2 ps of simulation. This procedure could require very extensive calculations where in principle we have to calculate the $pK_{a,int}$ for each mutant. However, here we found it reasonable to simplify the protocol and evaluate the $pK_{a,int}$ for all residues in a sphere of 10 Å centered around the mutated residues, while keeping the $pK_{a,int}$ for the rest of the residues at their value in the WT enzyme. The resulting $pK_{a,int}$ at the given ε_p were used with the help of Eq. (3) to evaluate the folding energy for each ε_{eff}.

III. Results

In this work we explored two issues; (i) our ability to predict the effect of mutations of a specific protein by using Eq. (3) with a universal set of ε_p' and ε_{eff} and (ii) the ad hoc assumption that Eq. (3) also describes the difference in folding energy between different

proteins. This is equivalent to the assumption that the non-electrostatic term in Eq. (4) is either constant or small, or to the more likely possibility that Eq. (5) is valid.

As a starting system we chose the protein ubiquitin, whose folding has been studied extensively (for a review see [22]). The structure of the WT ubiquitin is depicted in Fig. 2. The residues of interest to this study have been explicitly depicted. In order to compare our results with the experimental data [22], we started from a pseudo WT protein where residue 45, a Phe, has already been mutated to Trp. Using Eq. (3) we explored the dependence of ΔG_{fold}^{elec} on ε_p' and ε_{eff}, looking for values of the parameters that best reproduce the observed difference in stability of some mutants [23]. The corresponding analysis is given in Table 1, and as seen from the table we obtain the trend with $\varepsilon_{eff} = 40$ and $\varepsilon_p' = 20$ or 40.

In the second step we examined the performance of the dielectric constants found by considering as a benchmark a set mutants of lipase from *Bacillus subtilis* (Lip A) [24]. Lip A is a mesophilic lipase composed of 181 amino acids and the X-Ray crystallographic structure that we have chosen contains a single independent molecule in the asymmetric unit. This system was chosen recently in an exciting predication procedure where enhancing the thermostability of mesophilic enzymes should be made possible by increasing the rigidity at appropriate sites [25]. The chosen strategy by Reetz et al. [25] has been based on iterative saturation mutagenesis on the amino acids that show the highest B factors (or B values, atomic displacements parameters obtained from X-ray data) [25] where the rigidity is to be increased. Figure 3 depicts the Lip A system and the residues that were involved in the mutational study. The studied mutations from the WT structure with their observed T_{50}[15] values (the temperature required to reduce the

initial enzymatic activity by 50% within 15 min of heat treatment, which is often used to quantitatively characterize thermostability and is close to the critical temperature of denaturation) are specified in Figure 4A.

Here again we evaluated ΔG_{fold}^{elec} as a function of ε_p' and ε_{eff}. The corresponding results are summarized in Table 2. Assuming that T_{50}^{15} is directly correlated with ΔG_{fold}^{elec}, we obtain the best agreement with the experimental data by using ε_p' =40 and ε_{eff} =40. The resulting relationship between the calculated ΔG_{fold}^{elec} and the observed T_{50}^{15} is depicted in Figure 4A and 4B. The success of the present model and the success of Reetz´s approach indicate that the rigidity and electrostatic effect are correlated.

In the next step we tried to explore the ad hoc assumption that ΔG_{fold}^{elec}, with the set of ε_p and ε_{eff} found in studies of relative stabilities of mutants of the same protein, somehow reproduces the trend in stability between different proteins. We start this exploration by considering the folding energy of the mesophilic dihydrofolate reductase from *Escherichia coli* (EcDHFR) [26] and the thermophilic dihydrofolate reductase from *Thermotoga maritime* (TmDHFR) [27]. Assuming that Eq. (3) represents the largest contribution to Eq. (4) we tried to establish the range of ε_p' and ε_{eff} that reproduces the best agreement between the calculated and observed folding energies.

The simulated ΔG_{fold}^{elec} are given in Table 3 for the different ε_p' and ε_{eff}. If we select ε_{eff} =40 and ε_p' =40 or 80 we obtain a good agreement with the observed folding energies of -6 kcal/mol and -34 kcal/mol for the monomer and dimer respectively [28,29].

Obviously, the "leap of faith" made in assuming that the electrostatic contribution to folding determines the trend in the overall folding energy cannot be established by two

proteins and thus we examined three more proteins with the same approach . This was done for the cold shock proteins (Csp) from the mesophilic bacterium *Bacillus subtilis* (Bs-Csp), thermophilic bacterium *Bacillus caldolyticus* (Bc-Csp) and hyperthermophilic bacterium *Thermotoga maritime* (Tm-Csp). Here again we obtained (see supplementary material) the trend in the observed folding energy by using Eq. (3) with ε_{eff} =40 and ε_p' =40. In Table 4 we show the melting temperature, the observed folding energies and the calculated ΔG^{elec}_{fold} for the set of ε_{eff} =40 and ε_p' =40.

Although the trend obtained in the above studies is very promising, the values of the optimal dielectric constants are far from obvious and in some respect puzzeling. That is, consistent analysis of the dielectric constants in proteins indicated that the ε_p' obtained with the PDLD/S-LRA approximation should be between 4 to 6 and ε_{eff} should be around 40. While here we found ε_p' =40, 80 and ε_{eff} =40.

In a preliminary attempt to explore this dielectric trend we turned back to the benchmark of the ubiquitin, whose pK_a's have also been subjected to pK_a studies in the WT and some of its mutants [30]. In this case we know the observed apparent pK_a for some acidic residues of the protein and we can examine the optimal ε_p and ε_{eff} for the pK_a calculations. The results of our study (Table 5) appear to give a ε_p between 6 and 8 which is significantly smaller than the values obtained for ε_p' from the folding studies. The origin of this difference will be analyzed below.

IV. Discussion

This work examined the electrostatic contributions to protein folding by using the semimacroscopic PDLD/S-LRA approach for exploring the relationship between the

folding energy and ε_p' and ε_{eff} . The ΔG_{fold}^{elec} obtained with ε_p'=40, 80 and ε_{eff} =40 are very promising and seem to offer a practical way for predicting the general trend in protein stability.

As stated in the previous section values of the optimal dielectric constants are puzzeling. In particular the high value of ε_p' seems to be inconsistent with previous considerations. We reestablish the fact that the ε_p used in semimacoscopic approaches with the LRA treatment should be around 6 [14] by evaluating the pK_a's of acidic groups in ubiquitin. However, the ε_p' that accounted best for the folding energy was found to be around 40. The origin of this discrepancy it is likely to be due to the fact that we calculate the intrinsic pK_a's for each mutant from an initial structure that includes the relaxation under the influence of the charges of the mutants (B' in Fig. 1). To account for the missing reorganization of the B $\rightarrow$ B' relaxation we need to use a larger ε_p (ε_p' in Eq. 3). That is, using Eq. (1-3) as well as Fig. 1, we can write,

$$\Delta\Delta G_{fold}^{WT\rightarrow M} \cong -2.3RT\sum_i Q_i \Delta pK_{i,\mathrm{int}}^{WT\rightarrow M}(\varepsilon_p) + \Delta\Delta G_{QQ}^{WT\rightarrow M} + \Delta\Delta G_{f\rightarrow f'}^{WT\rightarrow M} \cong$$
$$-2.3RT\sum_i Q_i (\Delta pK_{i,\mathrm{int}}^{WT\rightarrow M}(\varepsilon_p'))' + \Delta\Delta G_{QQ}^{WT\rightarrow M} \qquad (6)$$

where WT designates wild type, M mutant, $\Delta\Delta G_{QQ}^{WT\rightarrow M}$ is the last term in Eq. (3) and $\Delta\Delta G_{f\rightarrow f'}^{WT\rightarrow M}$ is the energy difference for the B$\rightarrow$B' step. Apparently, $\left|\sum_i Q_i \Delta\Delta pK_{i,\mathrm{int}}^{WT\rightarrow M}(\varepsilon_p)\right|$ with the ε_p that reproduces pK_a's (i.e. ε_p= 4-8) is much larger than the change in folding upon mutations, so that $\left|\sum_i Q_i \Delta(\Delta pK_{i,\mathrm{int}}^{WT\rightarrow M}(\varepsilon_p'))'\right|$ must involve larger ε_p' to account for the compensating effect of $\Delta\Delta G_{f\rightarrow f'}^{WT\rightarrow M}$. Note that $\Delta pK_{a,\mathrm{int}}^{WT\rightarrow M}(\varepsilon_p)$ corresponds to the actual pK_a.

Another way to see this point is to realize that the (ΔpK_a)' involves the use of a constraint or partially fixed structure and as such requires a large ε_p to compensate for the missing reorganization.

Although the above discussion is instructive, it may be sufficient in this stage to simplify. Note that the optimal values of ε_p' and ε_{eff} provide a very useful estimate of the relative stability of different mutants of the same protein and that other values do not seem to give reasonable results. Interestingly, we can also gain additional insight about the optimal ε_p by considering the (A)→(C) path in Fig 1. In the cases where we compare different proteins, it was found that $\Delta G^{charged}_{uf\rightarrow f}$ with large ε_p' and ε_{eff} gives a very reasonable estimate of the total folding energy. The origin of this surprising finding is not fully clear, but it is consistent with the view that the electrostatic energy in this path can be represented by the ε_{eff} of Eq. (5). It also indicates that the folding process in the (A)→(C) path is similar to the process considered implicitly in studies of polyelectrolyte with a "simple" Coulomb's law type dielectric. In this case ε_{eff} reflects all the compensating effects, including the effect of changing the self energy by changing the environment of each ionized group. It must be stated at this point that we are not violating any electrostatic principal by the above view, since ε_{eff} is a parameter that describes the work of bringing charges from one distance to another and it includes all the reorganization effect (see discussion in [18]).

At present, it is not clear whether the dielectric ε_{eff} can or cannot be described by a uniform general function. Similarly, it is not clear if we can describe the interaction between the charges in the folded protein by the $Q_iQ_j/r_{ij}\varepsilon_{eff}$ term of Eq. (3), while using

a uniform ε_{eff}. In fact, the use of a uniform ε_{eff} in the calculations of the ΔG^{elec}_{fold} of Eq. (3) will be most physical if we start with the partially folded protein when each charge is already fixed in its local environment and bring the charges together (see *uf, 1→ folded* in Fig. 5). However, it is more likely that in the general case of protein folding we have a situation with *uf, 2→ folded* in Fig. 5, which would justify the use of Eq. (5). Thus the issue of the nature of ε_{eff} boils down to the nature of the electrostatic reorganization during the folding process and remains an open question.

Obviously, a determination of the relative merit of considering path 1 or path 2 cannot be determined by simple phenomenological analysis. Here it will be useful to have experimental and theoretical analysis of the pK_a in the unfolded state and of $\Delta G^{uncharged}_{uf\rightarrow f}$ as well as careful microscopic studies such as the free energy perturbation studies of charging and mutations (e.g. see reviews in ref. [14]). However, at present we feel that using the present model with the optimal ε_p' and ε_{eff} should provide a powerful tool for predicting protein stability. In particular, it will be interesting to see if we can predict new extra stable mutants in the series considered in the exciting approach of Reetz and coworkers [25].

Acknowledgments

This work was supported by NIH grant GM24492. We thank USC´s High Performance Computing and Communication Center (HPCC) for computer time. M.R. thanks the Generalitat Valenciana from Spain for the postdoctoral fellowship.

References

[1] Jaenicke, R. (1991) Protein Stability and Molecular Adaptation to Extreme Conditions. Eur. J. Biochem. 202, 715-728.

[2] Levitt, M. and Warshel, A. (1975) Computer-Simulation of Protein Folding. Nature 253, 694-698.

[3] Go, N. (1983) Theoretical-Studies of Protein Folding. Annu. Rev. Biophys. Bioeng. 12, 183-210.

[4] Onuchic, J.N., Wolynes, P.G., Lutheyschulten, Z. and Socci, N.D. (1995) Toward an Outline of the Topography of a Realistic Protein-Folding Funnel. Proc. Natl. Acad. Sci. USA 92, 3626-3630.

[5] Karanicolas, J. and Brooks, C.L. (2003) The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: Lessons for protein design? Proc. Natl. Acad. Sci. USA 100, 3954-3959.

[6] Dill, K.A. (1990) Dominant Forces in Protein Folding. Biochemistry 29, 7133-7155.

[7] Snow, C.D., Nguyen, N., Pande, V.S. and Gruebele, M. (2002) Absolute comparison of simulated and experimental protein-folding dynamics. Nature 420, 102-106.

[8] Fan, Z.Z., Hwang, J.K. and Warshel, A. (1999) Using simplified protein representation as a reference potential for all-atom calculations of folding free energy. Theor. Chem. Acc. 103, 77-80.

[9] Smith, L.J., Jones, R.M. and van Gunsteren, W.F. (2005) Characterization of the denaturation of human alpha-lactalbumin in urea by molecular dynamics simulations. Proteins 58, 439-449.

[10] Khalili, M., Liwo, A. and Scheraga, H.A. (2006) Kinetic studies of folding of the B-domain of staphylococcal protein A with molecular dynamics and a united-residue (UNRES) model of polypeptide chains. J. Mol. Biol. 355, 536-547.

[11] Hollien, J. and Marqusee, S. (1999) Structural distribution of stability in a thermophilic enzyme. Proc. Natl. Acad. Sci. USA 96, 13674-13678.

[12] Hendsch, Z.S. and Tidor, B. (1994) Do Salt Bridges Stabilize Proteins - a Continuum Electrostatic Analysis. Protein Sci. 3, 211-226.

[13] Honig, B. and Yang, A.S. (1995) Free-Energy Balance in Protein-Folding. Adv. Protein Chem. 46, 27-58.

[14] Warshel, A., Sharma, P.K., Kato, M. and Parson, W.W. (2006) Modeling electrostatic effects in proteins. BBA-Proteins Proteom. 1764, 1647-1676.

[15] Xiao, L. and Honig, B. (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. J. Mol. Biol. 289, 1435-1444.

[16] Giletto, A. and Pace, C.N. (1999) Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T-1. Biochemistry 38, 13379-13384.

[17] Tanford, C. and Kirkwood, J.G. (1957) Theory of protein titration curves. I. General equations for impenetrable spheres. J. Am. Chem. Soc. 79, 5333.

[18] Warshel, A., Russell, S.T. and Churg, A.K. (1984) Macroscopic Models for Studies of Electrostatic Interactions in Proteins - Limitations and Applicability. Proc. Natl. Acad. Sci. USA 81, 4785-4789.
[19] Warshel, A. (1979) Conversion of Light Energy to Electrostatic Energy in the Proton Pump of Halobacterium halobium. Photochem. Photobiol. 30, 285-290.
[20] Schutz, C.N. and Warshel, A. (2001) What are the dielectric 'constants' of proteins and how to validate electrostatic models. Proteins 44, 400-417.
[21] Lee, F.S., Chu, Z.T. and Warshel, A. (1993) Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the Polaris and Enzymix Programs. J. Comput. Chem. 14, 161-185.
[22] Jackson, S.E. (2006) Ubiquitin: a small protein folding paradigm. Org. Biomol. Chem. 4, 1845-1853.
[23] Went, H.M. and Jackson, S.E. (2005) Ubiquitin folds through a highly polarized transition state. Protein Eng. Des. Sel. 18, 229-237.
[24] Jaeger, K.E., Dijkstra, B.W. and Reetz, M.T. (1999) Bacterial biocatalysts: Molecular biology, three-dimensional structures, and biotechnological applications of lipases. Annu. Rev. Microbiol. 53, 315-351.
[25] Reetz, M.T., Carballeira, J.D. and Vogel, A. (2006) Interactive Saturation Mutagenesis on the Basis of B Factors as a Strategy for Increasing Protein Themostability. Angew. Chem. 45, 7745-7751.
[26] Reyes, V.M., Sawaya, M.R., Brown, K.A. and Kraut, J. (1995) Isomorphous Crystal-Structures of Escherichia-Coli Dihydrofolate-Reductase Complexed with Folate, 5-Deazafolate, and 5,10-Dideazatetrahydrofolate - Mechanistic Implications. Biochemistry 34, 2710-2723.
[27] Dams, T., Auerbach, G., Bader, G., Jacob, U., Ploom, T., Huber, R. and Jaenicke, R. (2000) The crystal structure of dihydrofolate reductase from Thermotoga maritima: Molecular features of thermostability. J. Mol. Biol. 297, 659-672.
[28] Dams, T. and Jaenicke, R. (1999) Stability and Folding of Dihydrofolate Reductase from the Hyperthermophilic Bacterium Thermotoga maritima. Biochemistry 38, 9169-9178.
[29] Ionescu, R.M., Smith, V.F., O´Neill, J.C.J. and Matthews, C.R. (2000) Multistate Equilibrium Unfolding od Escherichia coli Dihydrofolate Reductase: Thermodynamic and Spectroscopic Description of the Native, Intermediate, and Unfolded Ensembles. Biochemistry 39, 9540-9550.
[30] Sundd, M. and Robertson, A.D. (2003) Rearrangement of charge-charge interactions in variant ubiquitins as detected by double-mutant cycles and NMR. J. Mol. Biol. 332, 927-936.

Figure Legends

Figure 1. Two alternative descriptions of the folding of a charged protein. The process (A')→(C) involves a folding of the uncharged protein and subsequent charging, while (A)→(C) involves an initial charging of the unfolded protein and then folding of the charged protein. Structure B is already folded and has the same structure for all mutants, while B', it is the relaxed structure for each mutant.

Figure 2. X-Ray structure of wild type ubiquitin. The residues that are involved in the mutational study are represented in sphere model.

Figure 3. X-Ray structure of lipase from *Bacillus subtilis*. The residues that are involved in the mutational study are represented in sphere model.

Figure 4. (A) Thermostability diagram for the wild type (WT) lipase structure and some of its mutants with their observed T_{50}[15] values in Celsius. (B) Calculated thermostability diagram for the wild type (WT) lipase structure and some of its mutants with their calculated folding energy ΔG^{elec}_{fold} (in kcal/mol). The mutations are expressed in their simplified nomenclature and are cumulative along the arrows.

Figure 5. A schematic description of the nature of the dielectric effect in the two extremes. In the first case (1) the unfolded protein already stabilizes the separated ionized groups, while in the second case (2) the ionized groups are not surrounded by protein

dipoles in the unfolded protein. The dielectric effect reflects in both cases the change in solvation (by the protein and the solvent) during the charge separation process. This solvation effect compensates the gas phase energy (-332/R kcal/mol) and the net effect (-(332/R) + $\Delta G_{sol}(R)$ - ΔG^{∞}_{sol}) can be considered as (-(332/Rε_{eff})). The figure also includes the short range steric repulsion (V_{steric}) between the ions in addition to the 1/R term.

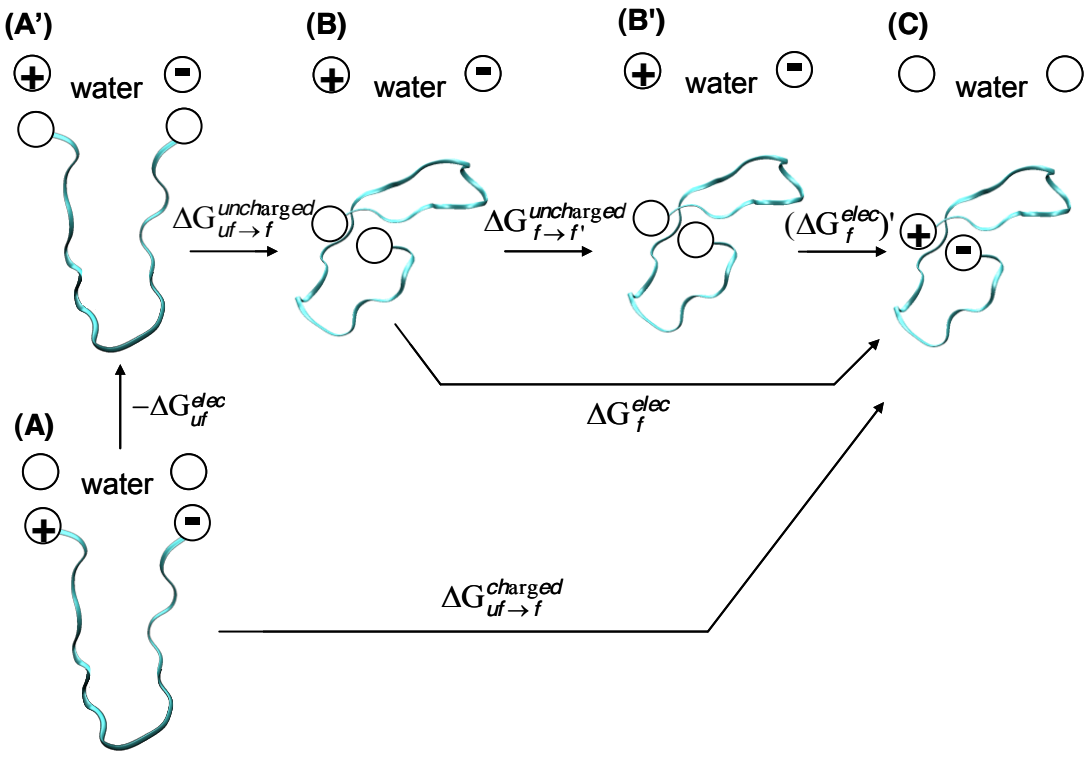
(A')
water
(B)
water
(B')
water
(C)
water
$\Delta G^{uncharged}_{uf \to f}$
$\Delta G^{uncharged}_{f \to f'}$
$(\Delta G^{elec}_{f})'$
$-\Delta G^{elec}_{uf}$
(A)
water
ΔG^{elec}_{f}
$\Delta G^{charged}_{uf \to f}$

Figure 1

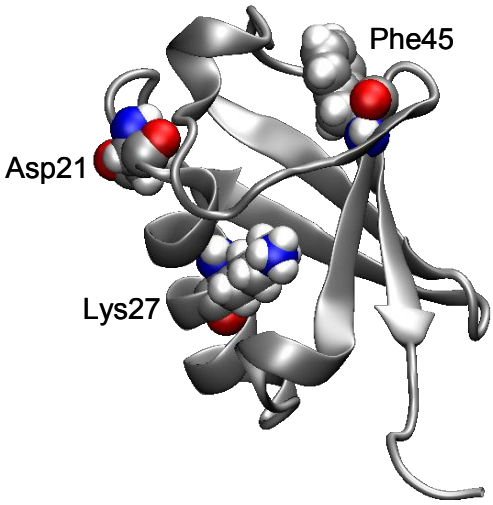

Figure 2

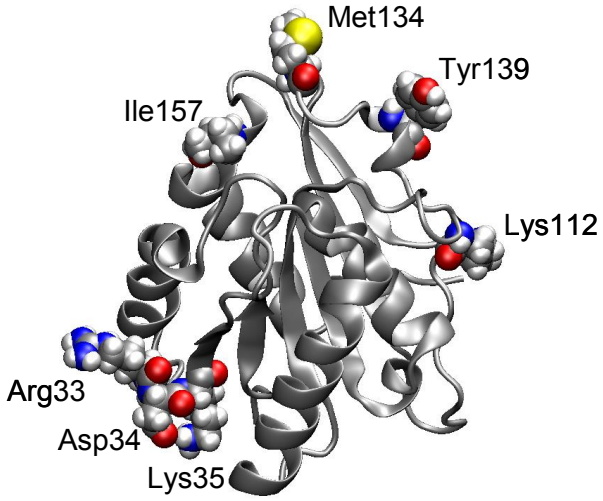

Figure 3

(A)

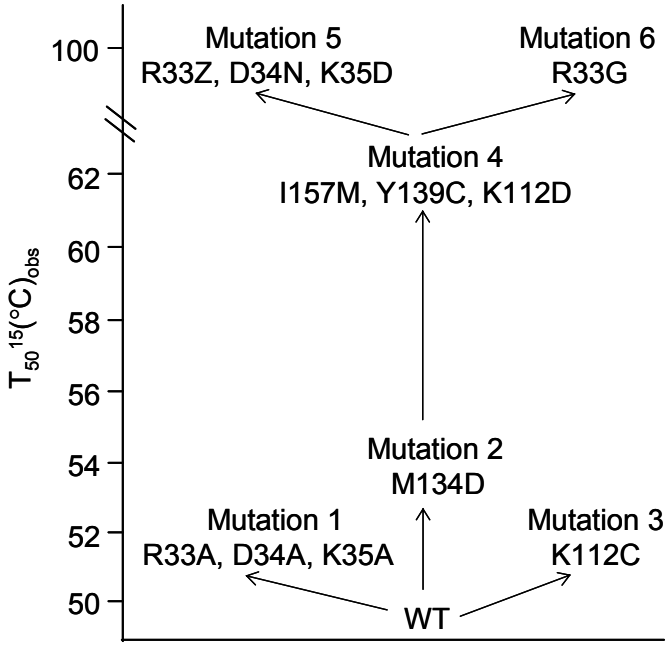

Figure 4A

(B)

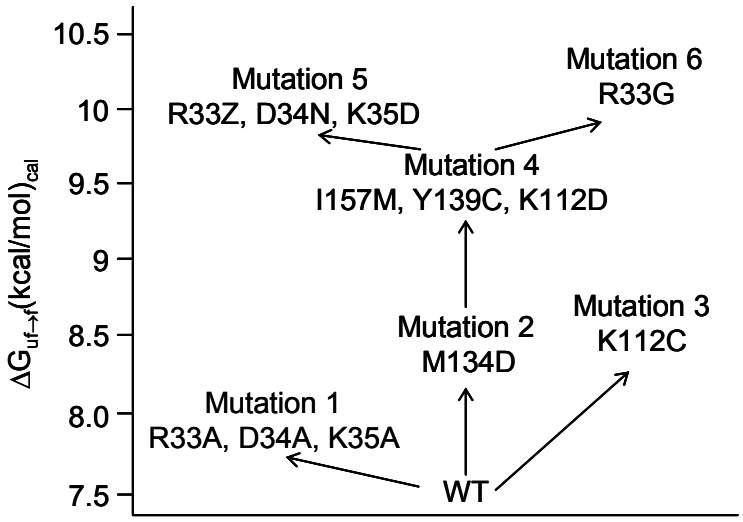

Figure 4B

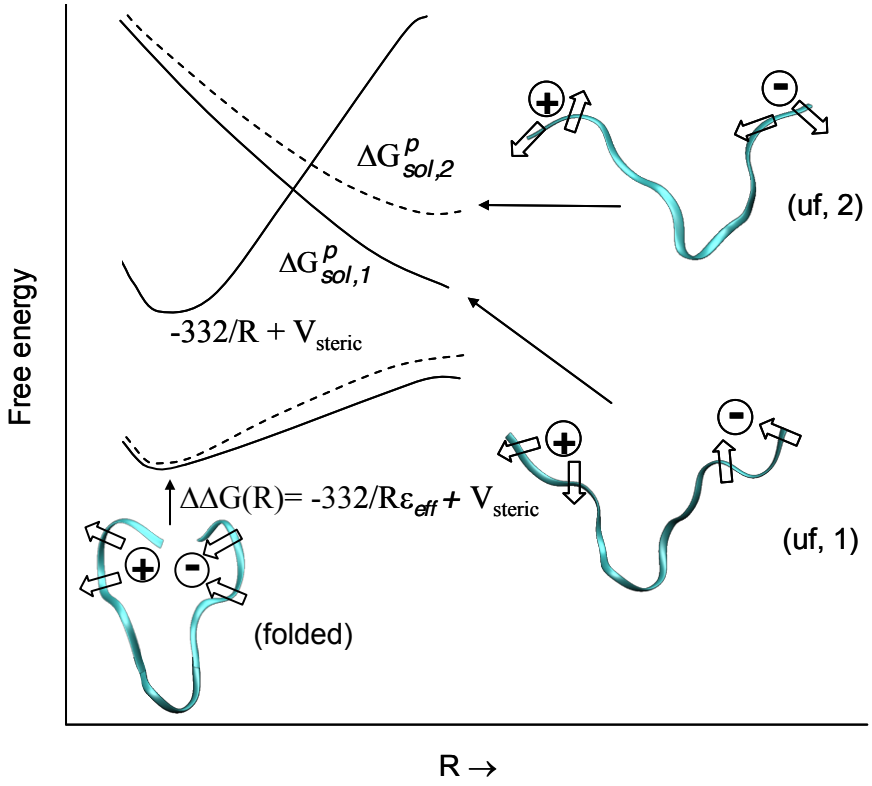
Free energy
R →
$\Delta G^{p}_{sol,2}$
$\Delta G^{p}_{sol,1}$
-332/R + V_{steric}
$\Delta\Delta G(R) = -332/R\varepsilon_{eff} + V_{steric}$
(uf, 2)
(uf, 1)
(folded)

Figure 5

Table 1. The dependence of the calculated ΔG_{fold}^{elec} on ε_p' and ε_{eff} for the ubiquitin and two of its mutants[(a)]

Pseudo Wild Type ubiquitin (Phe45Trp) $\Delta G_{fold,\,obs}$= -7.4				
ε_p' \ ε_{eff}	12	16	40	80
6	-30.99	-23.46	-8.02	-3.62
8	-31.34	-23.62	-8.21	-3.8
20	-31.87	-23.78	**-8.42**	-4.02
40	-32.26	-24.05	**-8.55**	-4.14
80	-32.55	-24.29	**-8.72**	-4.3

Asp21Asn				$\Delta G_{fold,\,obs}$= -6.1
ε_p' \ ε_{eff}	12	16	40	80
6	-15.71	-7.6	6.52	11.15
8	-19.58	-11.54	2.58	7.26
20	-26.07	-18.09	-3.98	0.71
40	-28.73	-20.79	**-6.69**	-2
80	-30.02	-22.1	-8	-3.31

Lys27Ala				$\Delta G_{fold,\,obs}$= -4.4
ε_p' \ ε_{eff}	12	16	40	80
6	-22.01	-14.63	-1.68	1.38
8	-24	-16.46	-2.84	0.29
20	-26.94	-19.35	**-4.89**	-1.4
40	-28.41	-20.81	**-5.98**	-2.18
80	-29.18	-21.58	-6.65	-2.63

(a) Energies in kcal/mol. In bold we indicate the calculated folding energies that are in good agreement with the observed folding energies.

Table 2. The dependence of the calculated ΔG_{fold}^{elec} on ε_p' and ε_{eff} for the Lip A and some of its mutants[a]

Wild Type Lip A	T_{50}^{15}=50ºC		
ε_p' \ ε_{eff}	**16**	**40**	**80**
6	-11.91	14.3	21.84
8	-18.21	7.77	14.91
20	-28.24	-2.74	4.03
40	-32.86	**-7.62**	-1.01
80	-35.22	-10.14	-3.59

Mutation 1	T_{50}^{15}=52ºC		
ε_p' \ ε_{eff}	**16**	**40**	**80**
8	-18.17	0.08	7.65
20	-23.71	-5.12	1.36
40	-26.43	**-7.94**	-2.03
80	-27.81	-9.42	-3.67

Mutation 2	T_{50}^{15}=54.3ºC		
ε_p' \ ε_{eff}	**16**	**40**	**80**
8	-31.37	5.1	11.66
20	-40.12	-4.12	2.28
40	-44.23	**-8.48**	-2.12
80	-46.32	-10.68	-4.37

Mutation 3	T_{50}^{15}=52ºC		
ε_p' \ ε_{eff}	**16**	**40**	**80**
8	-17.85	6.69	13.57
20	-27.91	-3.85	2.99
40	-32.43	**-8.63**	-1.82
80	-34.75	-11.09	-4.29

Mutation 4	T_{50}^{15}=62.8ºC		
ε_p' \ ε_{eff}	**16**	**40**	**80**
8	-17.22	5	12.18
20	-27.22	-5.19	2
40	-31.66	**-9.74**	-2.56
80	-33.92	-12.06	-4.86

Mutation 5		T_{50}^{15}=100ºC	
ε_p' \ ε_{eff}	**16**	**40**	**80**
8	-18.63	2.96	9.27
20	-24.92	-4.51	1.65
40	-28.33	**-9.93**	-2.2
80	-30.07	-11.27	-4.14

Mutation 6		T_{50}^{15}=100ºC	
ε_p' \ ε_{eff}	**16**	**40**	**80**
8	-21.98	1.31	8.67
20	-29.78	-6.98	0.47
40	-33.61	**-10.3**	-3.45
80	-35.55	-12.87	-5.43

(a) Energies in kcal/mol. In bold we indicated the calculated folding energies that present a correlation similar to the one that presents the observed T_{50}^{15} values.

Table 3. The dependence of the calculated ΔG_{fold}^{elec} on ε_p' and ε_{eff} for EcDHFR and TmDHFR[(a)]

Mesophile monomer	**EcDHFR**		$\Delta G_{fold, obs}$ **= -6**
ε_p' \ ε_{eff}	**16**	**40**	**80**
8	25.47	5.99	11.82
20	17.73	-2.01	4.34
40	13.86	**-5.87**	0.52
80	11.65	**-8.06**	-1.73
Hyperthermophile dimer	**TmDHFR**		$\Delta G_{fold, obs}$ **= -34**
ε_p' \ ε_{eff}	**16**	**40**	**80**
8	-41.28	36.54	59.01
20	-85.2	-3.81	19.19
40	-105.12	**-24.28**	-1.47
80	-115.08	**-34.55**	-11.79

(a) Energies in kcal/mol. In bold we indicate the calculated folding energies that are in good agreement with the observed folding energies.

Table 4. The melting temperature (T_m), the observed folding energy ($\Delta G_{fold,obs}$) and the calculated ΔG^{elec}_{fold} for ε_p'=40 and ε_{eff} =40 for three proteins from the Cold shock protein (Csp) family [(a)]

	T_m (°C)	$\Delta G_{fold, obs}$	ΔG^{elec}_{fold}
Mesophilic Bs-Csp	54	-1.2	-1.5
Thermophilic Bc-Csp	77	-5	-5.7
Hyperthermophilic Tm-Csp	83	-6.5	-7.9

(b) Folding energies are in kcal/mol.

Table 5. Calculated and observed pK_a's for acidic residues of ubiquitin for ε_{eff} 20 and 40[(a)]

Wild Type ubiquitin												
pK_a cal	Glu18			Asp21			Glu24			Asp32		
ε_p	$pK_{a,int}$	$pK_{a,app}$ ε_{eff}=20	$pK_{a,app}$ ε_{eff}=40	$pK_{a,int}$	$pK_{a,app}$ ε_{eff}=20	$pK_{a,app}$ ε_{eff}=40	$pK_{a,int}$	$pK_{a,app}$ ε_{eff}=20	$pK_{a,app}$ ε_{eff}=40	$pK_{a,int}$	$pK_{a,app}$ ε_{eff}=20	$pK_{a,app}$ ε_{eff}=40
4	3.92	6.10	5.16	1.28	3.46	2.52	2.84	**4.34**	3.68	5.24	5.70	5.54
6	4.05	**5.73**	**4.91**	2.16	**3.82**	**3.01**	3.32	**4.21**	3.78	4.79	**3.19**	**4.01**
8	4.11	5.78	4.97	2.59	4.23	**3.44**	3.57	**4.45**	**4.02**	4.57	2.96	**3.78**
20	4.17	5.82	5.02	3.36	4.96	4.2	3.94	4.79	**4.39**	4.24	2.61	3.44
40	4.23	5.86	5.08	3.63	5.20	4.47	4.12	4.95	4.57	4.07	2.42	3.27
80	4.27	5.88	5.12	3.76	5.31	4.6	4.21	5.03	4.65	3.98	2.31	3.18
pK_a obs	**4.3**			**3.1**			**4.3**			**3.8**		

(a) The intrinsic and apparent pK_a's for ε_{eff} =20 and 40 are given above for four acidic residues. In bold we indicate the apparent pK_a's that are in good agreement with the observed pK_a's. The observed pK_a's are taken from [30]