Published in final edited form as:

Proteins. 2009 November 15; 77(3): 670–684. doi:10.1002/prot.22481.

Effective approach for calculations of absolute stability of proteins using focused dielectric constants

Spyridon Vicatos¹, Maite Roca^{1,2}, and Arie Warshel^{1,*}

¹Department of Chemistry, University of Southern California, Los Angeles, California 90089

²Department Química Física, Universitat de Valencia, 46100 Burjassot, Valencia, Spain

Abstract

The ability to predict the absolute stability of proteins based on their corresponding sequence and structure is a problem of great fundamental and practical importance. In this work, we report an extensive, refinement and validation of our recent approach (Roca *et al.*, FEBS Lett 2007;581:2065–2071) for predicting absolute values of protein stability G_{fold} . This approach employs the semimacroscopic protein dipole Langevin dipole method in its linear response approximation version (PDL/S-LRA) while using the best fitted values of the dielectric constants ϵ'_p and ϵ'_{eff} for the self energy and charge–charge interactions, respectively. The method is validated on a diverse set of 45 proteins. It is found that the best fitted values of both dielectric constants are around 40. However, the self energy of internal residues and the charge–charge interactions of Lys have to be treated with care, using a somewhat lower values of ϵ'_p and ϵ'_{eff} . The predictions of G_{fold} reported here, have an average error of only 1.8 kcal/mole compared to the observed values, making our method very promising for estimating protein stability. It also provides valuable insight into the complex electrostatic phenomena taking place in folded proteins.

Keywords

protein stability; folding energy; dielectric constants; electrostatics in proteins

INTRODUCTION

The ability to predict physical and chemical properties of proteins, given their sequence and folded tertiary structure, is of crucial importance for the study of enzymes.^{1–3} Computational approaches for predicting the thermal stability of proteins, the difference of free energy between their folded and unfolded state, have yet to emerge and validate. That is, despite the progress in the development of models for studying the folding of proteins^{4–12} there are still major problems in predicting protein stability by either microscopic or macroscopic models. For example, we lack a clear understanding of the magnitude of electrostatic contributions to thermal stability and to the overall folding free energy. Similarly, the values of the dielectric

*Correspondence to: Arie Warshel, Department of Chemistry, University of Southern California, 418 SGM Building, 3620 McClintock Avenue, Los Angeles, CA 90089-1062. warshel@usc.edu.

constants to be used and the contribution of the ionizable residues to the stability of a folded protein are only a few of the questions still unanswered.

Discretized continuum studies^{13,14} have suggested that charged and polar groups lead to destabilization of a folded protein. Other studies, however, have indicated that protein stability is far more complicated than originally thought, and that charged residues do not necessarily destabilize the protein core. Quite the opposite, they tend to increase protein stability.^{15–18} Even the general idea in continuum studies that internal ionizable residues tend to destabilize the protein core is now under reevaluation.¹⁹ Overall there is a growing realization that electrostatic energy is related to stability in general, and that electrostatic interactions usually stabilize the native states of proteins quite significantly. However, the exact role of electrostatic interactions in protein stability is obscured by the competition between desolvation penalties, stabilization by local protein dipoles, and charge–charge interaction.

Recently, we introduced an approach²⁰ that evaluates the electrostatic contribution to protein stability by selecting relatively high values of dielectric constants for charge–charge interactions (ϵ_{eff}) and for self energy (ϵ_p). This approach determines the absolute stability of a given protein based on (a) the use of the semimacroscopic protein dipole Langevin dipole (PDL/D/S) in its linear response approximation version (the PDL/D/S-LRA method) for self energy and (b) the usage of high values ϵ_p and ϵ_{eff} . Our preliminary studies²⁰ indicated that a major part of the absolute folding energy is reproduced by the electrostatic energy evaluated with large ϵ_{eff} and more surprisingly by a large ϵ_p . The contribution of the screened charge–charge interactions is consistent with the recent findings (e.g., Makhatadze and coworkers^{21–23} as well as Garcia Moreno and coworkers^{24,25} and Pace and coworkers^{26–28}), as far as the relative contribution of surface groups is concerned. However, our finding appeared to be more general with implications about the energetics of internal groups. The corresponding physical picture appeared to be quite intriguing but more validation is essential. Moreover, the predictive value of our approach has not been fully validated. The present work explores this model in a much more extensive and systematic way and establishes its general validity. We focus on the challenge of finding the best fitted values for the dielectric constants and on the effectiveness of the method in predicting protein stability.

METHODS

The method described in this work focuses on the electrostatic contribution to the folding free energy, assuming that this is the main contribution. The method was described previously by Roca *et al.*,²⁰ and we consider below the main points.

Our starting point is the folding paths of Figure 1, where the system can move from the unfolded state (A) to the folded state (E) in two pathways. In one of them (A) $\rightarrow$ (B) $\rightarrow$ (C) $\rightarrow$ (D) $\rightarrow$ (E), the unfolded protein becomes uncharged (transferring the charges to solution on the proper proton transfer process), folds to its neutral folded form (C), and then move to the final folded charged form (E) through an uncharged structure that is similar to that of (E). In the second path, the charged protein folds directly by moving from (A) to (E). The

electrostatic contributions to folding, $\Delta G_{\text{fold}}^{\text{elec}}$, of a given protein at a specific ionization state and at a given pH, is evaluated by²⁹:

$$\begin{aligned} \Delta G_{\text{fold}}^{\text{elec}} = & \Delta G_{\text{f}}^{\text{elec}} - \Delta G_{\text{uf}}^{\text{elec}} = \\ & -2.3RT \sum_i Q_i^{(\text{f})} \left(\text{p}K_{i,\text{int}}^{\text{p}}(\epsilon_{\text{p}}) - \text{pH} \right) + 166 \sum_{i \neq j} \frac{Q_i^{(\text{f})} Q_j^{(\text{f})}}{r_{ij}^{(\text{f})} \epsilon_{\text{eff}}(r_{ij}^{(\text{f})})} \quad (1) \\ & + 2.3RT \sum_i Q_i^{(\text{uf})} \left(\text{p}K_{a,i}^{\text{w}} - \text{pH} \right) - 166 \sum_{i \neq j} \frac{Q_i^{(\text{uf})} Q_j^{(\text{uf})}}{\epsilon_{\text{water}} r_{ij}^{(\text{uf})}} \end{aligned}$$

where uf and f designate, the unfolded and folded states, Q_i is the charge of the i th residue, in the folding (f) or the unfolding (uf) state, ϵ_{p} is the dielectric constant used in the semimacroscopic calculation of intrinsic $\text{p}K_{\text{a}}$, $\text{p}K_{i,\text{int}}^{\text{p}}(\epsilon_{\text{p}})$ is the intrinsic $\text{p}K_{\text{a}}$ of the i th residue in its given protein state when all other residues are neutral, at a given ϵ_{p} . Finally r_{ij} is the distance between residues i and j , while ϵ_{eff} is the effective dielectric for charge–charge interactions inside the protein and ϵ_{water} is the effective dielectric constant for charge–charge interactions in water.

As clarified elsewhere,^{29,30} ϵ_{p} determines the “self energy” and the corresponding intrinsic $\text{p}K_{\text{a}}$ 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 explicitly in the simulation system. Basically, ϵ_{p} reflects all the effects that are not included explicitly in the calculations of the self energy.³⁰ Similarly, ϵ_{eff} is a parameter that reflects the compensation of the gas phase charge–charge interaction by the reorganization of the solvent and the protein.²⁹

The folding free energy also includes non-electrostatic contributions such as configuration entropy and hydrophobic contributions. These contributions depend on the path used in Figure 1. For example, we can write according to Figure 1,

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

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 $\Delta G_{\text{fold}}^{\text{elec}}$.

Eq. (1) describes the electrostatic contribution to the folding energy of the given protein, and the corresponding value depends, of course, upon the choice of ϵ_{p} and ϵ_{eff} . Concerning the non-electrostatic term $\Delta G_{\text{uf} \rightarrow \text{f}}^{\text{uncharged}}$, it may well be of different value for different proteins, and there is no current computational method that can provide a reliable estimate of the quantities. However, if we use the direct (A) $\rightarrow$ (E) path of Figure 1 we get a different picture of the calculate G_{fold} , since now the entire folding process can be described in terms of the corresponding change in charge–charge interactions, using

$$\Delta G_{\text{fold}} \cong 166 \sum_{i \neq j} \frac{Q_i Q_j}{\varepsilon_{\text{eff}}(r_{ij})} \left[\frac{1}{r_{ij}^{(f)}} - \frac{1}{r_{ij}^{(\text{uf})}} \right] \quad (3)$$

That is, now the non-electrostatic effects may be absorbed in the effective dielectric for the work of bringing the charged groups from the unfolded to the folded state in the (A) → (E) path of Figure 1. In other words, the (B) → (D) step involves non-electrostatic contributions that may be significant, while the step (D) → (E) is formally a well defined electrostatic step, which includes, of course, structural reorganization. The dielectric that reflects the structural reorganization in the (D) → (E) step may be modified by considering the free energy of the (B) → (D) step and requiring that this free energy will be reproduced by Eq. (1). Now if the effect of the (B) → (D) step is not large or if it is somehow correlated with the trend in the electrostatic free energy, we will have a uniform dielectric that works for all proteins. Conversely, if this is not the case we will not be able to use such approximation. It is this philosophy, though with some assumptions, that leads us to the next important step of our method. Since our concept of a dielectric constant is somewhat complex we refer the reader to Refs. 29–31 for further clarification.

If the $\Delta G_{\text{uf} \rightarrow \text{f}}^{\text{uncharged}}$ term is somehow small, reflecting compensation between the entropy and the hydrophobic effect, we may examine the approximation:

$$\Delta G_{\text{fold}} \approx \Delta G_{\text{fold}}^{\text{elec}}(\varepsilon'_p, \varepsilon'_{\text{eff}}) \quad (4)$$

where $\varepsilon'_p, \varepsilon'_{\text{eff}}$ is the set of ε_p and ε_{eff} that successfully reproduce the folding energy. In this case we obtain:

$$\Delta G_{\text{fold}}^{\text{elec}} = \Delta G_{\text{fold}}^{\text{elec}}(\varepsilon'_p, \varepsilon'_{\text{eff}}) \approx -2.3RT \sum_i Q_i \left(\text{p}K_{i,\text{int}}^p(\varepsilon'_p) - \text{p}K_{i,w}^w \right) + 166 \sum_{i \neq j} Q_i Q_j \left[\frac{1}{r_{ij}^{(f)} \varepsilon'_{\text{eff}}} - \frac{1}{80 r_{ij}^{(\text{uf})}} \right] \quad (5)$$

Where: uf and f designate, the unfolded and folded states, Q_i is the charge of the i th residue, in the folding (f) state only, $\text{p}K_{i,\text{int}}^p(\varepsilon'_p)$ is the intrinsic $\text{p}K_a$ of the i th residue calculated from the chosen ε'_p and not the original ε_p , and $\text{p}K_{i,w}^w$ is the $\text{p}K_a$ of the i residue in water.

Rigorously, we should use $Q_i^{(f)}$ but since the contribution from the unfolded state will be neglected here we leave the notation Q_i . We also use 80 instead of the $\varepsilon_{\text{water}}$ since the difference is trivial and the effect of this term is small. Finally, r_{ij} is the distance between residues i and j , in the folding (f) and unfolding (uf) state. Here, $\varepsilon_{\text{eff}}^{(f)}$ is the chosen dielectric effective constant for the charge–charge interactions. It should be pointed out that $\varepsilon_{\text{eff}}^{(f)}$ in Eq. (5) is taken here as independent of the distance r_{ij} . In later sections, we discuss exceptions, where the choice of $\varepsilon_{\text{eff}}^{(f)}$ for certain cases is affected by both the residues i and j as well as its distance r_{ij} . At any rate, the first term in Eq. (5) represents the change of self energy upon moving a charge from water to its site in the folded protein: That is,

$$-2.3RT \sum_i Q_i \left(\text{p}K_{i,\text{int}}^p(\varepsilon'_p) - \text{p}K_{i,w}^w \right) = \Delta \Delta G_{\text{sol}}^{\text{w} \rightarrow \text{p}}(Q_i=0 \rightarrow Q_i=Q_i^0) \quad (6)$$

The second term in Eq. (5) represents the effect of charge–charge interaction.

Adopting the approximation of Eq. (5) reduces the complicated and difficult problem of calculating absolute folding energy of a protein, to the finding of the proper, best fitted values of ε'_p and $\varepsilon_{\text{eff}}^{(f)}$, which provides the best estimate of G_{fold} . This idea is based on the fact that semimacroscopic models can represent implicitly physical and chemical properties even though they lack explicit representation of the phenomena taking place. Of course, such an approach can work only if the given model retains the main physics of the real system. Overall, we expect the best fitted values of ε'_p and $\varepsilon_{\text{eff}}^{(f)}$ to be higher than the values of 4 to 8, values used for example to calculate accurate values of $\text{p}K_a$'s in proteins.³²

Equation (5) can be further simplified, by assuming that $r_{ij}^{(\text{uf})}$ is in general much larger than $r_{ij}^{(f)}$, and that $\varepsilon_{\text{eff}}^{(f)}$ is smaller than 80. Thus, Eq. (5) is finally approximated by:

$$\begin{aligned} \Delta G_{\text{fold}}^{\text{elec}}(\varepsilon'_p, \varepsilon_{\text{eff}}^{(f)}) &= -2.3RT \sum_i Q_i \left(\text{p}K_{i,\text{int}}^p(\varepsilon'_p) - \text{p}K_{i,w}^w \right) \\ &\quad + 166 \sum_{i \neq j} Q_i Q_j \left[\frac{1}{r_{ij}^{(f)} \varepsilon_{\text{eff}}^{(f)}} \right] \\ &= 2.3RT \sum_i Q_i \left(\text{p}K_{i,\text{int}}^p(\varepsilon'_p) - \text{p}K_{i,w}^w \right) \\ &\quad + 166 \sum_{i \neq j} Q_i Q_j \left[\frac{1}{r_{ij} \varepsilon_{\text{eff}}'} \right] \end{aligned} \quad (7)$$

In Eq. (7) we drop the notation (f) from the variables $r_{ij}^{(f)}$ and $\varepsilon_{\text{eff}}^{(f)}$, and we will use r_{ij} and $\varepsilon_{\text{eff}}'$ as the distance and the corresponding dielectric constant for interactions between residues i and j in the folding state. This simplification may overlook cases where the contribution from the unfolded protein is significant. This might be the case when the unfolded protein has a compact structure, as implied by the studies of Raleigh and coworkers³³ or when electrostatic interaction in the unfolded state play a significant role for folding kinetics as suggested by Pace, Trefethen, and coworkers.^{26,27} However, it is possible that the contribution from the unfolded configuration is relatively small and the best way to judge this issue is to explore the validity of Eq. (7) in a large number of test cases.

The evaluation of the intrinsic $\text{p}K_a$'s were done by using the PDLD/S-LRA method according to standard protocol using the POLARIS module and the ENZYME force field of MOLARIS program.^{20,36} Each protein studied in this work was first solvated by the surface constrained all atom solvent (SCAAS) model,^{35,36} and all the ionizable groups (Asp, Glu, Lys, Arg, but not His) at pH = 7 were assigned a charge which is 50% of their full

charge at the ionized state (this was considered as the best procedure for the initial relaxation). The resulting system (protein and waters) was equilibrated by running a 100 ps molecular dynamics simulation with 1 fs time step at 300 K. The subsequent PDL/D-S-LRA calculations of the $pK_{i,int}^p(\epsilon'_p)$ of each ionizable residue started with 10 ps equilibration run, followed by 25 2 ps runs (starting from different configurations) on both the charged and uncharged state of the given residue. The resulted calculated $pK_{i,int}$ were then averaged to evaluate the final $pK_{i,int}^p(\epsilon'_p)$. The total simulation time depends on the size and the number of ionizable groups of the given protein. For example it took 22 hours on 4 dual core nodes (Dual Intel P4 3.0 GHz, 2 GB Memory) to evaluate all the $pK_{i,int}^p(\epsilon'_p)$ of snase (protein 1EY0, size 136 residues, which contains 48 ionizable residues).

Once the intrinsic pK_a 's were evaluated, the ionization states of the protein residues at pH = 7 were determined, by using a Monte Carlo approach described previously.³⁶ This procedure was repeated at different $\epsilon'_p, \epsilon'_{eff}$ and provides the charges (the Q_i) of the ionized residues as a function of the given dielectric constants. Using the $pK_{i,int}^p(\epsilon'_p)$ and the Q_i in Eq. (7) provided $\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff})$ as a function of ϵ'_p and ϵ'_{eff} .

RESULTS

Initial estimate of the best fitted values of ϵ'_p and ϵ'_{eff}

In this work, we moved to a much more extensive benchmark than the one used in our previous work,²⁰ considering all the proteins listed in Table I. Our test set included 45 proteins with various sequence sizes, folds, and function.

We started with a rough estimate of the best fitted set of $\epsilon'_p, \epsilon'_{eff}$ for the prediction of G_{fold} . This was done first for a single protein (L-Repressor, PDB ID 1LMB) by the approach illustrated in Figure 2.

In order to clarify the nature of Figure 2, it is useful to start, for example, with the evaluation of $\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff})$ at a specific point (e.g., $\epsilon'_p=20$ and $\epsilon'_{eff}=30$). In this case, we first used $\epsilon'_p=20$ and evaluate the intrinsic pK_a values for all the ionizable residues. In the next step, we used Eq. (7) with $\epsilon'_{eff}=30$ and obtained

$\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff}) = \Delta G_{fold}^{elec}(20, 30) = -4.27 \text{ kcal/mol}$. This value was then assigned in the corresponding point in Figure 2. Now the same procedure was repeated for other values of ϵ'_p and ϵ'_{eff} and the figure was completed. Next, we identified the set of $\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff})$ that gave the best agreement with the observed G_{fold} (around -4.6 kcal/mole for L-Repressor) and identified the optimal dielectric constants. In the case of Figure 2, the best fitted values were found to be in the vicinity of $\epsilon'_p = \epsilon'_{eff} = 40$.

After applying the above approach to all the 45 proteins of our test case, we considered several alternative ways of analyzing the corresponding results. The first and simplest

analysis which is summarized in Figure 3, was based on allowing both ϵ'_p and ϵ'_{eff} to be either 35 or 40, and choosing the value of $\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff})$ that minimizes the difference $|\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff}) - \Delta G_{fold}^{obs}|$. The corresponding results were used to generate Figure 3(A). This type of treatment will be referred here as getting the “best fitted” $\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff})$. The same procedure was followed in the generation of Figure 3(B), but this time with $8 \leq \epsilon'_p \leq 20$ and $8 \leq \epsilon'_{eff} \leq 20$.

As seen from both Figures 3(A) and 3(B), there is significant disagreement between the best fitted $\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff})$ and observed G_{fold} for many proteins. However, from these initial trials it appears that with the use of high values of ϵ'_p and ϵ'_{eff} of around 40, the absolute difference between observed and calculated stability $|\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff}) - \Delta G_{fold}|$ is significantly smaller than the difference obtained with other dielectric constants (see example demonstrated in Figure 3(B), where the discrepancy is much larger at the region when $8 \leq \epsilon'_p \leq 20, 8 \leq \epsilon'_{eff} \leq 20$).

Improving the model

In the next stage of the refinement of our model, we tried to further search for the best fitted values for the dielectric constants. This was done by considering the dielectrics that minimize the function

$$F_{error}(\epsilon'_p, \epsilon'_{eff}) = \sum_i |\Delta G_{(i)fold}^{elec}(\epsilon'_p, \epsilon'_{eff}) - \Delta G_{(i)fold}^{obs}| \quad (8)$$

Where i runs on all 45 test proteins, and $F_{error}(\epsilon'_p, \epsilon'_{eff})$ is the sum of the absolute value of errors between calculated and observed stabilities, for a specific value of ϵ'_p and ϵ'_{eff} . The resulting surface is shown in Figures 4(A) and 4(B).

As seen from the figures, the best fitted values of ϵ'_p and ϵ'_{eff} follow approximately a line where $\epsilon'_p = \epsilon'_{eff}$, and the global minimum occurs approximately where $\epsilon'_p = \epsilon'_{eff} = 40$. Thus, it is concluded that the best fitted dielectrics needed for prediction of protein stability is the one around $\epsilon'_p = \epsilon'_{eff} = 40$.

To improve the performance of our model, we started with a global analysis of the reasons for the disagreement between the calculated and observed values shown in Figure 3. This analysis focused first on the location of the ionized residues in the cases that have not produced satisfactory results. This was done by evaluating the distance between the ionized groups to the closest water molecule using the grid created by MOLARIS in the initial process of generating the SCAAS water sphere (see Refs. 29 and 36 for earlier studies along this line). The use of water grid in the identification of the internal/external residues is illustrated in Figure 5.

The above grid analysis indicated that the main problem is associated with internal groups, and led us to modify our approach, with the working hypothesis that in the case of truly internal groups it is better to use lower ε'_p while still using a large $\varepsilon'_{\text{eff}}$. This was done by defining an internal residue when the following conditions occur: (a) the shortest distance between a grid point and a heavy atom of the given residue is higher or equal to the threshold value reported in Table II, depending upon the type of the residue, and (b) the given residue has a low number of water molecules within a radius of 5 Å (five or less water molecules) from its geometrical center. Condition (b) was used mainly as a verification of condition (a).

For example, a Glu residue is considered to be internal when there are five or less water molecules within a distance of 5 Å from its geometrical center, and the closest water is 5.2 Å or more from its furthest heavy atom, which in this case is one of the two oxygen atoms off its side chain.

If a single ionizable residue is identified as an internal residue, then the intrinsic pK_a for this residue is chosen to take a value between the calculated $\text{pK}_{i,\text{int}}^p$ for $\varepsilon'_p=8$ and the $\text{pK}_{i,\text{int}}^p$ for $\varepsilon'_p=20$. For example, if a tested protein has 20 ionizable residues, and residue 17 is identified as internal, then in Eq. (7), for residue $i = 17$ the pK_a value used will be:

$$\text{pK}_{17,\text{int}}^p(\varepsilon'_p=\text{fixed}) \simeq \frac{\text{pK}_{17,\text{int}}^p(\varepsilon'_p=8) + \text{pK}_{17,\text{int}}^p(\varepsilon'_p=20)}{2} \quad (9)$$

Now Eq. (7) for this case (protein with 20 ionizable residues and one internal residue) becomes:

$$\begin{aligned} \Delta G_{\text{fold}}^{\text{elec}}(\varepsilon'_p, \varepsilon'_{\text{eff}}) = & -2.3RT \sum_{\substack{i=1 \\ i \neq 17}}^{20} Q_i (\text{pK}_{i,\text{int}}^p(\varepsilon'_p) - \text{pK}_{i,w}^w) \\ & - 2.3RT Q_{17} (\text{pK}_{17,\text{int}}^p(\varepsilon'_p=\text{fixed}) - \text{pK}_{17,w}^w) \\ & + 166 \sum_{i \neq j} Q_i Q_j \left[\frac{1}{r_{ij} \varepsilon'_{\text{eff}}} \right] \end{aligned} \quad (10)$$

In some cases we identified internal ionizable residues with a very large difference between calculated $\text{pK}_{i,\text{int}}^p(\varepsilon'_p)$ and $\text{pK}_{i,w}^w$. In these cases we found it useful to use an even lower value of ε'_p . Thus we further refined our procedure and used an intrinsic pK_a subscript that corresponds to $\varepsilon'_p=8$ when the difference $\text{pK}_{i,\text{int}}^p(\varepsilon'_p) - \text{pK}_{i,w}^w$ for values of ε'_p between 4 and 8 was positive and higher than 4 for Asp and Glu, and negative and lower than -3 for Lys and Arg.

Assigning lower ϵ'_p values and using the corresponding $pK_{i,int}^p(\epsilon'_p)$ for the internal ionizable residues, resulted in a significant increase in the accuracy of the calculated G_{fold} for our protein test set. Now the majority of the proteins were showing excellent G_{fold} predictions for dielectrics of $\epsilon'_p = \epsilon'_{eff} \approx 40$. However, we also observed that a significant fraction of the tested proteins give more accurate results for $35 \leq \epsilon'_p \leq 40$ and $20 \leq \epsilon'_{eff} \leq 25$. The result for both cases are shown in Figures 6(A) and 6(B), respectively.

Some of the results obtained after the specialized treatment of the internal residues were still puzzling. That is, although the majority of the proteins tests were giving excellent results for $\epsilon'_p = \epsilon'_{eff} = 40$, there was still a significant number of proteins where $\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff})$ was substantially lower than G_{fold} , (see Fig. 6B). However, the fact that in both cases the value of ϵ'_p was around 40, indicated that the problem is due to ϵ'_{eff} , that produced — in certain cases—the underestimation of G_{fold} . To decipher this non obvious behavior of ϵ'_{eff} , we concentrated on proteins cases where point mutation resulted in “shifting” the best fitted values of ϵ'_p and ϵ'_{eff} from the $35 \leq \epsilon'_p \leq 40$ and $35 \leq \epsilon'_{eff} \leq 40$ range to the $35 \leq \epsilon'_p \leq 40$ and $20 \leq \epsilon'_{eff} \leq 25$ range, and vice versa. This was found in the case of the CSP where we examined three proteins with high sequence similarities and only a few point mutations. In the case of bs-CSP we obtained the best results in the $35 \leq \epsilon'_p \leq 40$ and $20 \leq \epsilon'_{eff} \leq 25$ range, in the case of bc-CSP we obtained the best results in the $35 \leq \epsilon'_p \leq 40$ and $25 \leq \epsilon'_{eff} \leq 30$ range, while in the case of tm-CSP we obtained the best results in the $35 \leq \epsilon'_p \leq 40$ and $35 \leq \epsilon'_{eff} \leq 40$ range. The analysis of this data indicated that the interaction of Lys residue with charged residues at 10 Å or less leads to the above shift. Thus, we modified the ϵ'_{eff} for the interaction of Lys with negatively charged residues by using:

$$(\epsilon'_{eff})_{shifted} = 0.8\epsilon'_{eff} \quad (11)$$

for a distance $r_{ij} \leq 10$ Å.

This special treatment is rationalized by the fact that Lys residues can move their center quite easily and thus can adjust their interaction with counter ions (see analysis in our previous work⁶¹).

Using both the modified ϵ'_p and ϵ'_{eff} for Lys allowed us finally to obtain much better results, as shown in Figure 7, and in the corresponding columns in Table III. Now the agreement between the best fitted $\Delta G_{fold}^{elec}(\epsilon'_p, \epsilon'_{eff})$ and G_{fold} is excellent.

Using the results described in Figure 7 to calculate the sum of errors $F_{error}(\epsilon'_p, \epsilon'_{eff})$ introduced in Eq. (8), we evaluated again the best fitted values for the dielectric constants by determining the region of $(\epsilon'_p, \epsilon'_{eff})$ where $F_{error}(\epsilon'_p, \epsilon'_{eff})$ is minimized. This treatment led to the surface described in Figures 8(A) and 8(B). The best fitted region follows again an

approximate line where $\varepsilon'_p = \varepsilon'_{\text{eff}}$. The main difference between Figures 4 and 8 is that the region with $\varepsilon'_p = \varepsilon'_{\text{eff}}$ corresponds to much deeper valley (indicating a minimum in the function $F_{\text{error}}(\varepsilon'_p, \varepsilon'_{\text{eff}})$).

Table III summarizes the calculated results for the best fitted dielectric constants. As seen from the Table, we obtained quite accurate results. Most of the protein tested show error of around 1 kcal/mole, and only a few have errors above 2 kcal/mole, while still not exceeding 3 kcal/mole. The average error for the best fitted values of $\Delta G_{\text{fold}}^{\text{elec}}(\varepsilon'_p, \varepsilon'_{\text{eff}})$ (shown on the forth column of Table III) is only 0.7 kcal/mole.

The G_{fold} values reported in Table III for Glucanase C is zero, since the protein is unstable according to Ref. 50. In this case, it is found experimentally that without the Cys6–Cys27 disulfide bond, the protein is unstable, thus our finding that $\Delta G_{\text{fold}}^{\text{elec}}(\varepsilon'_p, \varepsilon'_{\text{eff}})$ is positive for all dielectric constants, is consistent with the experimental observation, since our calculations do not include the disulfide bond. Protein acidic fibroblast growth factor is also unstable under physiological conditions. It is reported^{62,63} that the protein without certain polyanions, remains relatively unstable.

Concerning the two proteins of Thioredoxin^{64,37} and Ferridoxin,^{3,65} our method underestimated the observed stability by 5 kcal/mole. However, those two proteins contain one disulfide bond, and since a disulfide bond stabilizes a protein with an amount of approximately 5 kcal/mole,^{66,67} we consider our method to be successful in predicting their absolute stabilities.

Predicting absolute stabilities of proteins

This work so far has shown that the calculated absolute stability of a protein can be correlated quite accurately with the corresponding observed value, by choosing best fitted values of ε'_p and $\varepsilon'_{\text{eff}}$ in the region $35 \leq \varepsilon'_p \leq 40; 35 \leq \varepsilon'_{\text{eff}} \leq 40$.

This however is not fully unbiased since the “best fitted” values are taken as the values that give the best result. Thus, it is important to select a well defined scheme for actually predicting the relevant stability. After considering several options, we concluded that a reasonable prediction can be obtained by using:

$$\overline{\Delta G}_{\text{fold}}^{\text{calc}} = \frac{\Delta G_{\text{fold}}^{\text{elec}}(35, 35) + \Delta G_{\text{fold}}^{\text{elec}}(35, 40)}{4} + \frac{\Delta G_{\text{fold}}^{\text{elec}}(40, 35) + \Delta G_{\text{fold}}^{\text{elec}}(40, 40)}{4} \quad (12)$$

The corresponding result for $\overline{\Delta G}_{\text{fold}}^{\text{calc}}$ are shown on the right columns of Table III and in Figure 9. As seen from the figure, the agreement is reasonable (an average error of 1.8 kcal/mole) although less impressive than that of Figure 7. However, in this case, we are not selecting any best fitted values and have a robust way of predicting protein stability.

Dependence of absolute stabilities on the size and fold of the protein

During the validation of our model we examined the possible influence/dependence of factors such as protein size and fold on absolute stability. To investigate such dependence, we chose the protein test set that included proteins of largely different size and fold. Comparisons of both the observed and the calculated stability with size and fold of proteins didn't show any particular correlation. For example, there are many small proteins in the test set shown in Table I, which have relatively large values of observed and calculated absolute stabilities (for example SSO7D, Histidine Phosphocarrier and Sac7d) but at the same time there are many that show low values (for example Barstar, PFRD-X64, and Bs CSP). The same trend is observed for large proteins (for example staphylococcal nuclease, isomerase, Tm DHFR, and aspartate aminotransferase show large values of absolute stability, while GDH Domain II, staphylococcal nuclease I92K, and Ec DHFR have low values for absolute stability). Comparison of fold and absolute stability also has not showed any correlation, i.e., whether the structures were mainly alpha, mainly beta, amorphous, or alpha-beta, didn't show an important difference in absolute stability.

Finally, no correlation between best fitted values of ϵ'_p and ϵ'_{eff} and the size or fold has been observed.

A practical and powerful approach for stabilizing proteins

The present approach should provide a powerful new way of stabilizing proteins by systematic mutations and this is expected to help augmenting existing strategies (e.g., Reetz *et al.*⁶⁸). We can of course just use Eq. (7) and examine systematically different mutant. This is done as a demonstration in Table IV. However, a more effective approach is obtained by differentiating Eq. (7) and evaluating the gradient of the free energy with regards to the charges of the protein residues.

$$\frac{\partial \Delta G_{fold}^{elec}}{\partial Q_i} \approx -2.3RT(pK_{i,int}^p - pK_{i,w}^w) + 332 \sum_{j \neq i} Q_j \left[\frac{1}{r_{ij} \epsilon'_{ij}} \right] \quad (13)$$

The gradient can then be used in predicting the change of charges that will stabilize the protein, using

$$\Delta Q_i = -\alpha \left(\frac{\partial \Delta G_{fold}^{elec}}{\partial Q_i} \right) \quad (14)$$

(where α is a scaling factor) and repeating the procedure iteratively we can predict the charge configurations that will increase the stability. Our preliminary validations of the above approach included a demonstration that we can reproduce the energetics of known mutants starting from the native protein, or from other mutants. Experimental validation of our approach should be extremely useful.

DISCUSSION

This work attempted to develop and validate a method for predicting absolute folding free energies. The main idea behind our approach is the hypothesis that the folding free energy is mainly determined by electrostatic free energies. Thus, the working hypothesis is that fitting the relevant dielectric constants will give a highly predicted model. The specific strategy involved evaluation of the intrinsic pK_a by the PDLD/S-LRA method with a given ϵ'_p and evaluating charge–charge interactions with a dielectric ϵ'_{eff} .

The overall best fitted dielectric constants were found to be in the range of 35–40 for both ϵ'_p and ϵ'_{eff} . However, a significant improvement was obtained by using a lower value of ϵ'_p for internal groups and by reducing ϵ'_{eff} for Lys groups.

As seen from Figure 9, the agreement between calculated and observed stability is reasonable although slightly less impressive than those of Figure 7. However, in this case, we are not selecting any best fitted values of ϵ'_p and ϵ'_{eff} , and we have a robust way of predicting protein stability. In fact, it is quite possible that the present method can be further improved by, for example, reducing ϵ'_{eff} for internal ion pairs, or by other physically based refinements.

The validity of our method can be best judged from its performance on the diverse test case used. To the best of our knowledge, such a performance in prediction of absolute folding energies has not been obtained by other approaches. Here, one can argue that the method is empirical and thus may reflect effects that are not related to the electrostatic free energy. We believe, however, that the method captures consistently electrostatic effects and discusses these issues in our previous work (Ref. 20). That is, using Figure 1 we rationalize the high ϵ'_p even for internal groups by pointing out that going on A $\rightarrow$ E path corresponds to a folding of polyelectrolyte where the energetics of buried ionized group in the protein interior is reflected in the effective dielectric for charge–charge interactions. We also note that using a large ϵ'_{eff} is fully consistent with our previous concepts.²⁹ Furthermore, we addressed the puzzling finding that ϵ_p for calculating pK_a 's is small (when evaluated consistently while considering the local protein relaxation³⁰) while ϵ'_p for folding calculations is relatively large. This issue was explored in the specific case of ubiquitin where it was concluded that ϵ_p in pK_a calculations reflects only the relaxation in the D $\rightarrow$ E step (the protein is near its structure in the charged configuration). On the other hand, ϵ'_p should reflect implicitly the compensating protein pre-organization in the C $\rightarrow$ E process (see Figure 5 in Ref. 20). Apparently, since this effect is not included explicitly, we have to use a large value of ϵ'_p .

The present work involved much more extensive benchmark than the one used in Ref. 20 and thus allowed us to refine our ideas about ϵ'_p and ϵ'_{eff} . We found that the idea of using large dielectrics is confirmed as much as surface groups are concerned. However, for

internal groups, it appears that a compromise that involves $\epsilon'_p \approx 20$ (which is still quite large) is required.

One may argue that the identification of our calculated folding free energy with electrostatic energy is not fully justified since the effective dielectric reflects non-electrostatic contributions. However, this argument overlooks the logical definition of energy contributions in biophysics. That is, if a given effect scales according to the charges of the system, it can and should be defined as an electrostatic contribution. For example, the free energy change associated with turning a charge on or off (e.g., in redox and pK_a processes) is rigorously defined as electrostatic free energy, despite the fact that the corresponding charging process is associated with a response of the system that reflects all the forces and interactions in the system. The same is true for the charging of an ion in water, which reflects the reorientation of the water molecules and even the water–water Van der Waals interactions. On the other hand, if the folding energy is due to the size of different residues, rather than their charges, it would be logical to state that the folding energy reflects steric repulsion. Alternatively, if we will be able to correlate the absolute folding free energy with the hydrophobicity of the residues, we will be able to say that the folding free energy is determined by the hydrophobic effect. The same is true with regards to configurational entropy effects. This does not mean that hydrophobic and entropic effects do not contribute to the folding free energy. However, the present study indicates that these contributions may cancel each other in some way. More specifically, it has been long argued that hydrophobic effects contribute in a major way to protein stability (e.g., Refs. 70, 71). While there are overwhelming evidences that such effects are a crucial part of the folding free energy, we are not aware of any clear correlation between the number or position of the hydrophobic residues and the observed absolute protein stability. Thus, although the reason for the lack of correlation is unclear (it may be due to a compensation by configurational entropy) it seems to us that (until proved otherwise) we should accept the fact that the absolute protein stability is not correlated with the markers of hydrophobic free energy.

One may also claim that our effective dielectric reflect non-electrostatic contributions such as the hydrophobic contributions in the (B) $\rightarrow$ (D) step of our cycle. This argument is reasonable as much as the cycle (A) $\rightarrow$ (E) is concerned but less reasonable when Eq. (3) captures the trend in the observed folding energy. At any rate, as long as we cannot establish correlation between hydrophobic markers and absolute stability it would be hard to attribute the folding energy to such effects.

It must be stated here that the present work has not established that the correlation with the electrostatic free energy is perfect. Obvious examples are the effects of SS bonds in the proteins thioredoxin, ferredoxin, and glucanase C. Cases of interaction with the ligand were not explored here carefully by considering the relevant electrostatic contributions. Another interesting case is observed for example in the case of residues of staphylococcal nuclease,⁴³ where the deviation between calculated and observed folding energy is due most probably to the hydrophobic contribution. However, the overall trend is clearly best and overwhelmingly correlated with the electrostatic free energy.

We would also like to mention that it is very likely that we will identify cases where the dielectric constant will not follow our idealized roles. For example, there could be cases where we will have to use higher ϵ'_{eff} when we have a large number of neighboring charge residues, that will create the equivalent of a high ionic strength. There could be also cases where ϵ'_{eff} at close distance will have to be decreased (this may be handled by employing a distance dependent dielectric³⁰). To explore and analyze such cases, we will have to focus on more comparative mutational studies, in particular between thermophile and mesophile enzymes. Obviously, further experiments and theoretical studies are clearly needed. Fortunately, recent experimental studies provide remarkable benchmark for such detailed studies.¹⁹ At any rate, regardless of the formal justification of our approach, its power can be exploited in predicting and analyzing protein stabilities.

As mentioned in the results-section, we have not found a correlation between the G_{fold} and the size of the protein or other characteristic properties. We did observe, however, that regardless of the size or the fold of a protein, a single mutation that turns a charge on or off may have a large impact on stability. This is clearly shown in the case of the staphylococcal nuclease mutations,⁴³ where a single mutation of a non-charged residue into Lysine, reduced the protein's stability by the enormous amount of around 9 kcal/mole. Therefore, we consider the charges of a protein as a more important factor in determining the overall stability, than the size or the fold. Obviously, various protein sizes and folds may reflect the amount and type of certain ionizable residues within a protein, but in the end it is the charges of those ionizable residues that define the stability.

It may be useful to point out here that Gurd and coworkers⁷² invoked the idea that the stability of proteins could be evaluated by considering the G_{ij} obtained with ϵ'_{eff} of 40 or larger. However, this was an almost obvious conclusion for those who accepted the Tanford Kirkwood model⁷³ which overlooked the self energy term and the destabilization expected from this term in non polar regions of the protein (see discussion in Warshel *et al.*⁷⁴). The interesting finding that seems to emerge from the present and other works is the observation that despite the fact that charges should not be stable in non polar environments the effective environment around charges in proteins behave as a relatively polar environment.

It is significant, however, to note that Eq. (3) does not produce perfect results [and the same is true for Eq. (7)] when ϵ'_p and ϵ'_{eff} are equal. In some cases, we had to use ϵ'_p , which is significantly smaller than 40 and this reflects the effect of the self energy. Yet this effect is much smaller than what would be deduced from the assumption that the charged groups are in partially non polar environment. In fact, as established in our previous work (Ref. 20) and the present work the best fitted ϵ'_p is much larger than the one that is required to obtain the observed pK_a 's. The reasons for this have been discussed in great length in Ref. 20 and it is associated with the compensation in the (D) $\rightarrow$ (E) step in Figure 1.

Another interesting issue is the acid unfolding effect.⁷⁵ Here, again we are dealing with a full electrostatic effect exactly as is the case in redox processes since the unfolding upon charging of ionizable residues reflects exactly the free energy of the charging process, which always involves structural rearrangements. In other words, it does not matter that the

unfolding process is opposed by hydrophobic forces, the overall energetics is rigorously the charging free energy and thus electrostatic free energy. This is very different than the (B) → (D) free energy which is a non-electrostatic energy as it does not involve any change in charge. At any rate, predicting the energetics of the acid unfolding is clearly a task that should be accomplished by our formulation.

The present approach focuses in absolute folding energy and is likely to work well on the less challenging problem of exploring mutational effects. The limited test cases of mutational effects examined in this present work provides encouraging results, but further studies are needed. Here, we have a very extensive set of relevant experiments (e.g., Ref. 43–46, 58).

ACKNOWLEDGMENTS

The authors thank USC's High Performance Computing and Communication Center (HPCC) for computer time.

Grant sponsor: NIH; Grant number: GM24492

REFERENCES

1. Jaenicke R. Protein stability and molecular adaptation to extreme conditions. *Eur J Biochem.* 1991; 202:715–728. [PubMed: 1765088]
2. Luke KA, Higgins CL, Wittung-Stafshede P. Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS J.* 2007; 274:4023–4033. [PubMed: 17683332]
3. Razvi A, Scholtz JM. Lessons in stability from thermophilic proteins. *Protein Sci.* 2006; 15:1569–1578. [PubMed: 16815912]
4. Clark NS, Dodd I, Mossakowska DE, Smith RA, Gore MG. Folding and conformational studies on SCR1-3 domains of human complement receptor 1. *Protein Eng.* 1996; 9:877–884. [PubMed: 8931127]
5. Dill KA. Dominant forces in protein folding. *Biochemistry.* 1990; 29:7133–7155. [PubMed: 2207096]
6. Fan ZZ, Hwang JK, Warshel A. Using simplified protein representation as a reference potential for all-atom calculations of folding free energy. *Theor Chem Acc.* 1999; 103:77–80.
7. Go N. Theoretical-studies of protein folding. *Annu Rev Biophys Bioeng.* 1983; 12:183–210. [PubMed: 6347038]
8. Karanikolas J, Brooks CL. 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.* 2003; 100:3954–3959. [PubMed: 12655041]
9. Khalili M, Liwo A, Scheraga HA. 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.* 2006; 355:536–547. [PubMed: 16324712]
10. Levitt M, Warshel A. Computer-simulation of protein folding. *Nature.* 1975; 253:694–698. [PubMed: 1167625]
11. Onuchic JN, Wolynes PG, Lutheyschulten Z, Socci ND. Toward an outline of the topography of a realistic protein-folding funnel. *Proc Natl Acad Sci USA.* 1995; 92:3626–3630. [PubMed: 7724609]
12. Snow CD, Nguyen N, Pande VS, Gruebele M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature.* 2002; 420:102–106. [PubMed: 12422224]
13. Hendsch ZS, Tidor B. Do salt bridges stabilize proteins—a continuum electrostatic analysis. *Protein Sci.* 1994; 3:211–226. [PubMed: 8003958]
14. Honig B, Yang AS. Free-energy balance in protein-folding. *Adv Protein Chem.* 1995; 46:27–58. [PubMed: 7771321]

15. Schwehm JM, Fitch CA, Dang BN, Garcia-Moreno EB, Stites WE. Changes in stability upon charge reversal and neutralization substitution in staphylococcal nuclease are dominated by favorable electrostatic effects. *Biochemistry*. 2003; 42:1118–1128. [PubMed: 12549934]
16. Whitten ST, Garcia-Moreno EB. pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. *Biochemistry*. 2000; 39:14292–14304. [PubMed: 11087378]
17. Giletto A, Pace CN. Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T-I. *Biochemistry*. 1999; 38:13379–13384. [PubMed: 10529213]
18. Xiao L, Honig B. Electrostatic contributions to the stability of hyperthermophilic proteins. *J Mol Biol*. 1999; 289:1435–1444. [PubMed: 10373377]
19. Isom DG, Cannon BR, Castaneda CA, Robinson A, Garcia-Moreno EB. High tolerance for ionizable residues in the hydrophobic interior of proteins. *Proc Natl Acad Sci USA*. 2008; 105:17784–17788. [PubMed: 19004768]
20. Roca M, Messer B, Warshel A. Electrostatic contributions to protein stability and folding energy. *FEBS Lett*. 2007; 581:2065–2071. [PubMed: 17466986]
21. Strickler SS, Gribenko AV, Gribenko AV, Keiffer TR, Tomlinson J, Reihle T, Loladze VV, Makhatadze GI. Protein stability and surface electrostatics: a charged relationship. *Biochemistry*. 2006; 45:2761–2766. [PubMed: 16503630]
22. Gribenko AV, Makhatadze GI. Role of the charge-charge interactions in defining stability and halophilicity of the CspB proteins. *J Mol Biol*. 2007; 366:842–856. [PubMed: 17188709]
23. Schweiker KL, Zarrine-Afsar A, Davidson AR, Makhatadze GI. Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge charge interactions. *Protein Sci*. 2007; 16:2694–2702. [PubMed: 18029422]
24. Baran KL, Chimenti MS, Schlessman JL, Fitch CA, Herbst KJ, Garcia-Moreno BE. Electrostatic effects in a network of polar and ionizable groups in staphylococcal nuclease. *J Mol Biol*. 2008; 379:1045–1062. [PubMed: 18499123]
25. Lee KK, Fitch CA, Garcia-Moreno EB. Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. *Protein Sci*. 2002; 11:1004–1016. [PubMed: 11967358]
26. Pace CN. Single surface stabilizer. *Nat Struct Biol*. 2000; 7:345–346. [PubMed: 10802723]
27. Trefethen JM, Pace CN, Scholtz JM, Brems DN. Charge-charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa. *Protein Sci*. 2005; 14:1934–1938. [PubMed: 15937282]
28. Trevino SR, Gokulan K, Newsom S, Thurlkill RL, Shaw KL, Mitkevich VA, Makarov AA, Sacchettini JC, Scholtz JM, Pace CN. Asp79 makes a large, unfavorable contribution to the stability of RNase Sa. *J Mol Biol*. 2005; 354:967–978. [PubMed: 16288913]
29. Warshel A, Sharma PK, Kato M, Parson WW. Modeling electrostatic effects in proteins. *Biochim Biophys Acta*. 2006; 1764:1647–1676. [PubMed: 17049320]
30. Schutz CN, Warshel A. What are the dielectric ‘constants’ of proteins and how to validate electrostatic models. *Proteins*. 2001; 44:400–417. [PubMed: 11484218]
31. Sham YY, Muegge I, Warshel A. The effect of protein relaxation on charge-charge interactions and dielectric constants of proteins. *Biophys J*. 1998; 74:1744–1753. [PubMed: 9545037]
32. Sham YY, Chu ZT, Warshel A. Consistent calculations of pKa’s of ionizable residues in proteins: semi-microscopic and microscopic approaches. *J Phys Chem B*. 1997; 101:4458–4472.
33. Shan B, Bhattacharya S, Eliezer D, Raleigh DP. The low-pH unfolded state of the C-terminal domain of the ribosomal protein L9 contains significant secondary structure in the absence of denaturant but is no more compact than the low-pH urea unfolded state. *Biochemistry*. 2008; 47:9565–9573. [PubMed: 18707127]
34. Wong KB, Lee CF, Chan SH, Leung TY, Chen YW, Bycroft M. Solution structure and thermal stability of ribosomal protein L30e from hyperthermophilic archaeon *Thermococcus celer*. *Protein Sci*. 2003; 12:1483–1495. [PubMed: 12824494]
35. King G, Warshel A. A surface constrained all-atom solvent model for effective simulations of polar solutions. *J Chem Phys*. 1989; 91:3647–3661.

36. Lee FS, Chu ZT, Warshel A. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIK programs. *J Comp Chem*. 1993; 14:161–185.
37. Kumar S, Tsai CJ, Nussinov R. Maximal stabilities of reversible two-state proteins. *Biochemistry*. 2002; 41:5359–5374. [PubMed: 11969396]
38. Feng YQ, Sligar SG. Effect of heme binding on the structure and stability of *Escherichia coli* apocytochrome b562. *Biochemistry*. 1991; 30:10150–10155. [PubMed: 1931945]
39. Bowie JU, Sauer RT. Equilibrium dissociation and unfolding of the Arc repressor dimer. *Biochemistry*. 1989; 28:7139–7143. [PubMed: 2819054]
40. Hollien J, Marqusee S. A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry*. 1999; 38:3831–3836. [PubMed: 10090773]
41. Li WT, Grayling RA, Sandman K, Edmondson S, Shriver JW, Reeve JN. Thermodynamic stability of archaeal histones. *Biochemistry*. 1998; 37:10563–10572. [PubMed: 9692945]
42. Strop P, Mayo SL. Contribution of surface salt bridges to protein stability. *Biochemistry*. 2000; 39:1251–1255. [PubMed: 10684603]
43. Nguyen DM, Leila Reynald R, Gittis AG, Lattman EE. X-ray and thermodynamic studies of staphylococcal nuclease variants I92E and I92K: insights into polarity of the protein interior. *J Mol Biol*. 2004; 341:565–574. [PubMed: 15276844]
44. Jacob M, Holtermann G, Perl D, Reinsteiner J, Schindler T, Geeves MA, Schmid FX. Microsecond folding of the cold shock protein measured by a pressure-jump technique. *Biochemistry*. 1999; 38:2882–2891. [PubMed: 10074340]
45. Mueller U, Perl D, Schmid FX, Heinemann U. Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein. *J Mol Biol*. 2000; 297:975–988. [PubMed: 10736231]
46. Went HM, Jackson SE. Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel*. 2005; 18:229–237. [PubMed: 15857839]
47. Khorasanizadeh S, Peters ID, Butt TR, Roder H. Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry*. 1993; 32:7054–7063. [PubMed: 8392867]
48. Dams T, Jaenicke R. Stability and folding of dihydrofolate reductase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biochemistry*. 1999; 38:9169–9178. [PubMed: 10413491]
49. Mukaiyama A, Takano K, Haruki M, Morikawa M, Kanaya S. Kinetically robust monomeric protein from a hyperthermophile. *Biochemistry*. 2004; 43:13859–13866. [PubMed: 15504048]
50. Creagh AL, Koska J, Johnson PE, Tomme P, Joshi MD, McIntosh LP, Kilburn DG, Haynes CA. Stability and oligosaccharide binding of the N1 cellulose-binding domain of *Cellulomonas fimi* endoglucanase CenC. *Biochemistry*. 1998; 37:3529–3537. [PubMed: 9521674]
51. Janssen MJ, van de Wiel WA, Beiboer SH, van Kampen MD, Verheij HM, Slotboom AJ, Egmond MR. Catalytic role of the active site histidine of porcine pancreatic phospholipase A2 probed by the variants H48Q, H48N and H48K. *Protein Eng*. 1999; 12:497–503. [PubMed: 10388847]
52. Jackson SE, Moracci M, elMasry N, Johnson CM, Fersht AR. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry*. 1993; 32:11259–11269. [PubMed: 8218191]
53. Chrnyk BA, Evans J, Lillquist J, Young P, Wetzel R. Inclusion body formation and protein stability in sequence variants of interleukin-1 beta. *J Biol Chem*. 1993; 268:18053–18061. [PubMed: 8394358]
54. Zhou Z, Feng H, Bai Y. Detection of a hidden folding intermediate in the focal adhesion target domain: Implications for its function and folding. *Proteins*. 2006; 65:259–265. [PubMed: 16909417]
55. Cowley AB, Altuve A, Kuchment O, Terzyan S, Zhang X, Rivera M, Benson DR. Toward engineering the stability and heme-binding properties of microsomal cytochromes b5 into rat outer mitochondrial membrane cytochrome b5: examining the influence of residues 25 and 71. *Biochemistry*. 2002; 41:11566–11581. [PubMed: 12269800]
56. Sandberg WS, Terwilliger TC. Engineering multiple properties of a protein by combinatorial mutagenesis. *Proc Natl Acad Sci USA*. 1993; 90:8367–8371. [PubMed: 8378307]

57. Hornby JA, Luo JK, Stevens JM, Wallace LA, Kaplan W, Armstrong RN, Dirr HW. Equilibrium folding of dimeric class mu glutathione transferases involves a stable monomeric intermediate. *Biochemistry*. 2000; 39:12336–12344. [PubMed: 11015213]
58. Mach H, Ryan JA, Burke CJ, Volkin DB, Middaugh CR. Partially structured self-associating states of acidic fibroblast growth factor. *Biochemistry*. 1993; 32:7703–7711. [PubMed: 7688566]
59. Mainfroid V, Mande SC, Hol WG, Martial JA, Goraj K. Stabilization of human triosephosphate isomerase by improvement of the stability of individual alpha-helices in dimeric as well as monomeric forms of the protein. *Biochemistry*. 1996; 35:4110–4117. [PubMed: 8672446]
60. Chen L, Cabrita GJ, Otzen DE, Melo EP. Stabilization of the ribosomal protein S6 by trehalose is counterbalanced by the formation of a putative off-pathway species. *J Mol Biol*. 2005; 351:402–416. [PubMed: 16002092]
61. Cutler RL, Davies AM, Creighton S, Warshel A, Moore GR, Smith M, Mauk AG. Role of arginine-38 in regulation of the cytochrome c oxidation-reduction equilibrium. *Biochemistry*. 1989; 28:3188–3197. [PubMed: 2545252]
62. Copeland RA, Ji H, Halfpenny AJ, Williams RW, Thompson KC, Herber WK, Thomas KA, Bruner MW, Ryan JA, Marquis-Omer D, Sanyal G, Sitrin RD, Yamazaki S, Middaugh CR. The structure of human acidic fibroblast growth factor and its interaction with heparin. *Arch Biochem Biophys*. 1991; 289:53–61. [PubMed: 1716876]
63. Gospodarowicz D, Cheng J. Heparin protects basic and acidic FGF from inactivation. *J Cell Physiol*. 1986; 128:475–484. [PubMed: 3528177]
64. Katti SK, LeMaster DM, Eklund H. Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *J Mol Biol*. 1990; 212:167–184. [PubMed: 2181145]
65. Macedo-Ribeiro S, Darimont B, Sterner R, Huber R. Small structural changes account for the high thermostability of [4Fe-4S] ferredoxin from the hyperthermophilic bacterium *Thermotoga maritima*. *Structure*. 1996; 4:1291–1301. [PubMed: 8939753]
66. Matsumura M, Matthews BW. Stabilization of functional proteins by introduction of multiple disulfide bonds. *Methods Enzymol*. 1991; 202:336–356. [PubMed: 1784181]
67. Zavodszky P, Kardos J, Svingor A, Petsko GA. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc Natl Acad Sci USA*. 1998; 95:7406–7411. [PubMed: 9636162]
68. Reetz MT, Carballeira JD, Vogel A. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew Chem Int Ed Engl*. 2006; 45:7745–7751. [PubMed: 17075931]
69. Jaeger KE, Dijkstra BW, Reetz MT. Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases. *Annu Rev Microbiol*. 1999; 53:315–351. [PubMed: 10547694]
70. Privalov PL. Stability of proteins: small globular proteins. *Adv Protein Chem*. 1979; 33:167–241. [PubMed: 44431]
71. Creighton TE. Protein folding. *Biochem J*. 1990; 270:1–16. [PubMed: 2204340]
72. Matthew JB, Gurd FR, Garcia-Moreno B, Flanagan MA, March KL, Shire SJ. pH-dependent processes in proteins. *CRC Crit Rev Biochem*. 1985; 18:91–197. [PubMed: 3899508]
73. Tanford C, Kirkwood JG. Theory of protein titration curves. I. General equations for impenetrable spheres. *J Am Chem Soc*. 1957; 79:5333.
74. Warshel A, Russell ST, Churg AK. Macroscopic models for studies of electrostatic interactions in proteins: limitations and applicability. *Proc Natl Acad Sci USA*. 1984; 81:4785–4789. [PubMed: 6589625]
75. Perutz MF. Electrostatic effects in proteins. *Science*. 1978; 201:1187–1191. [PubMed: 694508]

Figure 1.

The thermodynamic cycle for the folding of a charged protein. Steps $A \rightarrow B$ and $B \rightarrow C$ correspond, respectively, to the uncharging of the unfolded protein and the folding of the uncharged protein. Step $C \rightarrow D$ corresponds to the change of the unfolded protein from its equilibrium structure to the structure of the folded charged protein. Step $A \rightarrow E$ corresponds to a direct folding of the ionized protein. The figure defines the free energy terms used in this work.

	e_eff	2.00	4.00	8.00	20.00	25.00	30.00	35.00	40.00	45.00	50.00	60.00	70.00	80.00
e_p														
2.0		-76.21	-32.23	-10.29	5.54	7.22	8.14	8.69	9.04	9.28	9.45	9.64	9.73	9.75
4.0		-134.64	-50.93	-13.42	4.92	7.23	8.84	9.92	10.65	11.16	11.53	12.02	12.32	12.52
8.0		-179.00	-84.34	-32.64	-5.77	-2.04	0.50	2.21	3.38	4.21	4.84	5.73	6.33	6.77
20.0		-184.38	-90.03	-38.78	-10.41	-6.69	-4.27	-2.71	-1.65	-0.89	-0.30	0.54	1.12	1.54
25.0		-185.36	-91.06	-39.94	-11.25	-7.53	-5.13	-3.58	-2.53	-1.77	-1.19	-0.35	0.22	0.65
30.0		-186.02	-91.77	-40.76	-11.85	-8.13	-5.72	-4.18	-3.13	-2.37	-1.79	-0.96	-0.38	0.04
35.0		-186.43	-92.23	-41.29	-12.21	-8.50	-6.10	-4.57	-3.52	-2.77	-2.19	-1.36	-0.78	-0.35
40.0		-186.77	-92.61	-41.72	-12.55	-8.83	-6.44	-4.91	-3.87	-3.11	-2.53	-1.70	-1.12	-0.70
45.0		-187.04	-92.88	-42.01	-12.78	-9.06	-6.68	-5.15	-4.11	-3.35	-2.78	-1.95	-1.37	-0.95
50.0		-187.19	-93.08	-42.25	-12.95	-9.24	-6.85	-5.32	-4.28	-3.52	-2.95	-2.11	-1.54	-1.11
60.0		-187.54	-93.43	-42.64	-13.25	-9.54	-7.16	-5.63	-4.59	-3.84	-3.26	-2.43	-1.86	-1.43
70.0		-187.70	-93.63	-42.87	-13.42	-9.71	-7.33	-5.80	-4.77	-4.02	-3.44	-2.61	-2.03	-1.61
80.0		-187.89	-93.83	-43.10	-13.59	-9.88	-7.50	-5.97	-4.94	-4.18	-3.61	-2.78	-2.20	-1.78

The optimal region

Figure 2.
The calculated values of $\Delta G_{\text{fold}}^{\text{elec}}$ as a function of ϵ_p' and ϵ_{eff}' for L-Repressor. The best fitted values of ϵ_p' and ϵ_{eff}' are taken as the values that give the best agreement between $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon_p', \epsilon_{\text{eff}}')$ and the observed G_{fold} . Observed stability for this protein is -4.6 kcal/mole. As seen from the figure, these values are around $35 \leq \epsilon_p' \leq 45$ and $35 \leq \epsilon_{\text{eff}}' \leq 45$.

Figure 3.

The correlation between the best fitted $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ and G_{fold} for best fitted values of ϵ'_p and ϵ'_{eff} in a restricted range. The calculations considered all the proteins in the benchmark of Table I and take the value of $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ that gives best fitted results for the two allowed values of ϵ'_p and ϵ'_{eff} (see text for the procedure used for selecting the best fitted values). The two values considered in Figure 3(A) are $\epsilon_p, \epsilon_{\text{eff}} = 35, 40$ whereas in Figure 3(B) we consider $\epsilon_p, \epsilon_{\text{eff}} = 8, 20$. As discussed in the text, the calculated best fitted values of

$\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ diverge greatly from the corresponding observed values. The average error between observed G_{fold} and best fitted $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ is 3.5 kcal/mole, and 7.3 kcal/mole for Figure 3(A) and 3(B), respectively. Some points of Figure 3(B) are outside the scale of this figure.

Figure 4.

The surface (A) and contour plots (B) of $F_{\text{error}}(\epsilon'_p, \epsilon'_{\text{eff}})$. In these plots it is clearly shown that the best fitted values for the dielectric constants follow approximately the line $\epsilon'_p = \epsilon'_{\text{eff}}$, and that for $\epsilon'_p = \epsilon'_{\text{eff}} = 40$ and above, the error reaches its minimum plateau.

Figure 5. Illustrating our procedure for defining internal groups. The protein is surrounded by a cubic grid of water molecules and the distance between the water oxygen and the given group is used to determine whether the group is internal or external.

Figure 6.

The correlation between $\Delta G_{\text{fold}}^{\text{elec}}$ and $\Delta G_{\text{fold}}^{\text{obs}}$ after using lower ϵ'_p for the internal residues. Both Figures 6(A) and 6(B) use the same representation as in Figure 3. In this case, the test proteins fall into two main categories: (a) the one shown in Figure 6(A) where the best fitted $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ is in the range of $35 \leq \epsilon'_p \leq 40$, $35 \leq \epsilon'_{\text{eff}} \leq 40$, and (b) the one shown in Figure 6(B) where the best fitted $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ is in the range $35 \leq \epsilon'_p \leq 40$, $20 \leq \epsilon'_{\text{eff}} \leq 25$. The majority of the test proteins fall into the first category.

Figure 7.

The best fitted $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ and observed G_{fold} for the tested proteins, after the use of lower ϵ'_p for the internal residues and lower ϵ'_{eff} for the Lys with $r_{ij} > 10$ Å. In this case, all proteins show best fitted values of $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ when dielectric constants are in the range of $35 \leq \epsilon'_p \leq 40$, $35 \leq \epsilon'_{\text{eff}} \leq 40$. The difference between best fitted $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ and observed G_{fold} has significantly decreased, compared to the previous cases of Figures 3 and 5. (For 45 proteins, the average difference is reduced to 0.7 kcal/mole.)

Figure 8.

The surface (A) and contour plots (B) of $F_{\text{error}}(\epsilon'_p, \epsilon'_{\text{eff}})$ obtained after refinement of ϵ'_p for internal groups and ϵ'_{eff} for Lys. The plots show clearly that the best fitted values for the dielectric constants follow approximately the line $\epsilon'_p = \epsilon'_{\text{eff}}$, with a global minimum of around $(\epsilon'_p, \epsilon'_{\text{eff}}) = (40, 40)$. This means that values of $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$ proximal to G_{fold} can be acquired when $\epsilon'_p, \epsilon'_{\text{eff}}$ have high values, and $\epsilon'_p = \epsilon'_{\text{eff}}$.

Figure 9. Calculated and observed folding energies for all our benchmark. The calculations are done using Eq. (12), rather than by choosing the best fitted values of $\Delta G_{\text{fold}}^{\text{elec}}(\epsilon'_p, \epsilon'_{\text{eff}})$, as was in Figure 7.

Table I

The benchmark considered in this work

No.	Name	PDB ID	Number of residues	Type	Source
1	SSO7d	ISSO	62	monomer	37
2	Thioredoxin	2TRX	108	monomer	37
3	Barstar	1A19	89	monomer	37
4	Apoflavodoxin	IFTG	168	monomer	37
5	λ -Repressor	ILMB	87	monomer	37
6	Snase	1EY0	136	monomer	37
7	BsHpr, Phosphotransferase	2HID	87	monomer	37
8	FeCyt b ₅₆₂	IQPU	106	monomer	37,38
9	Arc Repressor	1ARQ	53	dimer	37,39
10	Aspartate Amino transferase	1VPE	398	monomer	3
11	Chey	1TMY	118	monomer	3
12	GDH Domain II	1B26 (A02)	234	monomer	2,3,37
13	Histidine Phosphocarrier	1Y4Y	87	monomer	3
14	Phosphotransferase, Histidine containing protein	2HPR	87	monomer	3
15	Ribosomal	1H7M	99	monomer	34
16	RNase H*	1JXB	147	monomer	3,40
17	Ferridoxin	1VJW	59	monomer	3
18	O-Methyl Guanine DNA methyltransferase	1MGT	169	monomer	3
19	Sac7d	1WD0	66	monomer	2,3
20	^a Histone	1BFM	134	dimer	2,41
21	PFRD-XC4	1QCV	53	monomer	42
22	^a Staphylococcal Nuclease I92K	1TR5	130	monomer	43
23	Staphylococcal Nuclease I92E	1TR5	130	monomer	43
24	^a Staphylococcal Nuclease	1TR5	130	monomer	43
25	Bs CSP	1CSP	67	monomer	20,44
26	Bc CSP	1C9O	66	monomer	20,45
27	Tm CSP	1G6P	61	monomer	2,20

No.	Name	PDB ID	Number of residues	Type	Source
28	^a Ubiquitin D2IN	1AAR	76	monomer	20,46
29	Ubiquitin F45W	1AAR	76	monomer	20,47
30	^a Ubiquitin K27A	1AAR	76	monomer	20,46
31	Tm DHFR	1CZ3	164	dimer	20,48
32	Ec DHFR	1RX2	159	monomer	20
33	Ribonuclease	1XIP	212	monomer	2,49
34	Glucanase C	1CX1	153	monomer	50
35	Phospholipid A2	1P2P	124	monomer	51
36	Trypsin Proteinase	2CI2	65	monomer	52
37	Interleukin	1IOB	153	monomer	53
38	^a Adhesion transferase Y92E	1K40	126	monomer	54
39	Complement	1GKG	136	monomer	4
40	Cytochrome b5 Rat	1B5M	84	monomer	55
41	Gene V DNA Binding	1VQB	86	monomer	56
42	Glu Transferase	1GSD	208	dimer	57
43	Growth Factor	1FGA	124	monomer	58
44	Isomerase	1HTI	248	dimer	59
45	Ribosomal S6	1RIS	97	dimer	60

^aThe structure used to calculate $\Delta G_{\text{fold}}^{\text{elec}}$ was obtained by mutating the corresponding PDB structure reported in the third column of this table.

Table II

The Threshold Distances Between the Terminal Heavy Atom of the Given Amino Acid and the Closest Grid Point (The Closest Water Molecule)

Residue type	Threshold (Å)
ASP	5.0
GLU	5.2
HIS	5.2
LYS	5.5
ARG	5.7

Table III

Predicted G_{fold} for the Series of Proteins Considered in this Work^a

No.	Protein	$-\Delta G_{\text{fold}}^{\text{elec}}$	$-\Delta G_{\text{fold}}^{\text{obs}}$	Error	Best ϵ_p	Best ϵ_{eff}	$-\Delta G_{\text{fold}}^{\text{calc}}$	Error
1	SSO7d	7.9	8.0	0.1	40	40	8.5	0.5
2	Thioredoxin ^b	4.4	9.0	4.6	40	35	3.1	5.9
3	Barstar	5.9	5.7	0.2	35	35	4.9	0.8
4	Apoflavodoxin	4.3	4.3	0.0	35	40	5.0	0.7
5	λ -Repressor	4.5	4.6	0.1	40	40	4.8	0.2
6	Snase	6.3	6.2	0.1	35	40	8.0	1.8
7	BsHpr, Phosphotransferase	3.1	4.0	0.9	40	35	2.1	1.9
8	FeCyt b ₅₆₂	7.7	5.2	2.5	35	40	9.7	4.5
9	Arc Repressor	4.3	4.6	0.3	40	35	1.7	2.9
10	Aspartate Amino transferase	26.8	28.9	2.1	40	40	31.0	2.1
11	Chey	11.6	9.5	2.1	35	40	13.5	4.0
12	GDH Domain II	5.6	4.9	0.7	35	35	4.2	0.7
13	Histidine Phosphocarrier	5.4	8.2	2.8	40	35	4.5	3.7
14	Phosphotransferase, Histidine containing protein	5.3	5.2	0.1	40	35	4.0	1.2
15	Ribosomal	7.4	10.7	3.3	40	35	6.4	4.3
16	RNase H*	7.1	7.5	0.4	40	35	6.5	1.0
17	Ferridoxin ^b	3.8	9.3	5.5	40	35	3.2	6.1
18	O-Methyl Guanine DNA methyltransferase	10.2	10.2	0.0	35	35	9.1	1.1
19	Sac7d	7.8	7.4	0.4	40	40	9.0	1.6
20	Histone	7.8	7.2	0.6	40	35	7.6	0.4
21	PFRD-XC4	2.1	3.2	1.1	40	35	1.2	2.0
22	Staphylococcal Nuclease I92K	1.5	2.3	0.8	40	40	2.8	0.5
23	Staphylococcal Nuclease I92E	3.8	4.2	0.4	40	35	2.2	2
24	Staphylococcal Nuclease	11.5	11.5	0.0	35	35	10.3	1.2
25	Bs CSP	1.1	1.2	0.1	35	35	0.8	0.4
26	Bc CSP	5.2	5.0	0.2	40	35	4.2	0.8
27	Tm CSP	6.8	6.5	0.3	35	40	8.0	1.5

No.	Protein	$-\Delta G_{\text{fold}}^{\text{elec}}$	$-\Delta G_{\text{fold}}^{\text{obs}}$	Error	Best ϵ_p	Best ϵ_{eff}	$-\Delta G_{\text{fold}}^{\text{calc}}$	Error
28	Ubiquitin D21N	5.1	6.1	1.0	40	35	3.9	2.2
29	Ubiquitin F45W	6.9	7.4	0.5	40	35	5.9	1.5
30	Ubiquitin K27A	4.3	4.4	0.1	40	40	4.8	0.4
31	Tm DHFR	30.4	30.1	0.3	40	35	25.9	4.2
32	Ec DHFR	6.5	6.0	0.5	35	35	4.9	1.1
33	Ribonuclease	11.2	10.5	0.7	40	40	13.1	2.6
34	Glucanase C ^b	0.0	0.0	0.0	–	–	0.0	0.0
35	Phospholipid A2	7.8	6.5	1.3	35	40	9.1	2.6
36	Trypsin Proteinase	7.7	7.5	0.2	35	35	6.8	0.7
37	Interleukin	9.7	9.1	0.6	35	35	8.9	0.2
38	Adhesion transferase Y92E	7.5	8.0	0.5	40	40	8.4	0.4
39	Complement	2.6	2.7	0.1	35	35	2.5	0.2
40	Cytochrome b5 Rat	3.5	3.2	0.3	35	35	2.3	0.9
41	Gene V DNA Binding	7.6	9.0	1.4	40	35	6.6	2.4
42	Glu Transferase	13.6	14.4	0.8	35	35	10.2	4.2
43	Growth Factor	0.0	0.0	–	–	–	0.0	0.0
44	Isomerase	19.2	19.3	0.1	40	35	13.7	5.6
45	Ribosomal S6	11.1	8.0	3.1	35	40	12.9	4.9

^a All free energies reported are in kcal/mol. The third column reports the experimental values of the absolute stabilities of the tested proteins. The last two columns give the predicted stability according to Eq. (12) and the corresponding error. The sources of the observed values are given in Table I.

^b Proteins containing a disulfide bond (which is not treated by the present method) where the corresponding “missing” energy is around 5 kcal/mole.

Table IV

Calculated $\Delta G_{\text{fold}}^{\text{elec}}$ for the Set of Dielectrics $\epsilon'_{\text{p}}=40$ and $\epsilon'_{\text{eff}}=40$ for Different Proposed Mutations in the Wild Type Lip A^a

Ionized to non polar mutations of the wild type Lip A	
Structures	$\Delta G_{\text{fold}}^{\text{elec}}$
Wild type	-7.62 ^b
Arg142Ala	-9.27
Lys23Ala	-9.61
Lys70Ala	-9.65
Lys88Ala	-9.80
Arg107Ala	-10.15
Arg33Ala	-10.31
Lys23Ala//Arg33Ala	-10.64
Lys23Ala//Arg33Ala//Arg107Ala	-12.43
Non polar to ionized mutations of the wild type Lip A	
Structures	$\Delta G_{\text{fold}}^{\text{elec}}$
Wild type	-7.62 ^b
Ala38Asp	-8.53
Ala132Asp	-8.96
Ala105Asp	-9.39
Ala15Asp	-9.52
Ala75Asp	-9.77
Ala81Asp	-9.97
Ala20Asp	-10.78
Ala68Asp	-10.89
Ala146Asp	-11.03
Ala97Asp	-11.21
Ala113Asp	-11.63
Charged to non polar mutations of the most stable mutant (variant XI) of Lip A	
Structures	$\Delta G_{\text{fold}}^{\text{elec}}$
Variant XI	-10.30 ^b
Lys32Ala	-10.71
Asp133Ala	-11.08
Lys70Ala	-11.16
Lys23Ala//Lys70Ala	-11.32

^aEnergies in kcal/mol.

^bTaken from Ref. 69.