Prediction of Protein Relative Enthalpic Stability from Molecular Dynamics Simulations of the Folded and Unfolded States

Voichita M. Dadarlat, Lev A. Gorenstein, and Carol Beth Post*
Markey Center for Structural Biology, Department of Medicinal Chemistry, Purdue University, West Lafayette, Indiana

ABSTRACT For proteins of known structure, the relative enthalpic stability with respect to wild-type, $\Delta\Delta H^U$, can be estimated by direct computation of the folded and unfolded state energies. We propose a model by which the change in stability upon mutation can be predicted from all-atom molecular dynamics simulations for the folded state and a peptide-based model for the unfolded state. The unfolding enthalpies are expressed in terms of environmental and hydration-solvent reorganization contributions that readily allow a residue-specific analysis of $\Delta\Delta H^U$. The method is applied to estimate the relative enthalpic stability of variants with buried charged groups in T4 lysozyme. The predicted relative stabilities are in good agreement with experimental data. Environmental factors are observed to contribute more than hydration to the overall $\Delta\Delta H^U$. The residue-specific analysis finds that the effects of burying charge are both localized and long-range. The enthalpy for hydration-solvent reorganization varies considerably among different amino-acid types, but because the variant folded state structures are similar to those of the wild-type, the hydration-solvent reorganization contribution to $\Delta\Delta H^U$ is localized at the mutation site, in contrast to environmental contributions. Overall, mutation of apolar and polar amino acids to charged amino acids are destabilizing, but the reasons are complex and differ from site to site.

INTRODUCTION

Genetic variation due to changes in the amino-acid composition of wild-type (WT) proteins can lead to abnormal functioning of biological molecules and outright disease. Reliable estimates of the folded-state relative stability of variant proteins with respect to WT are important for understanding monogenic diseases, protein misfolding diseases, and the evolutionary pathway to resistance to medication (1–3). Experimental techniques can measure the overall changes in protein structure and folded state stability but cannot give a detailed picture of amino-acid-specific effects on protein stability.

Studies of protein stability can be complemented and guided by well-designed, time-saving, and less-expensive computational studies. In recent years, computational models based on statistical descriptions of the folded state ensembles have been successful in reproducing native-state hydrogen exchange and model the pH and temperature dependence of protein stability by introducing and exploiting the so-called residue-based energetic profiling of proteins (4–6), and references therein. Despite this notable progress, recent comparisons of various methods to quantify contributions of individual amino acids to WT protein and variant folded state stability (7,8) have concluded that a complete understanding and accurate prediction of the change in stability and structure associated with specific mutations has not yet been accomplished ((9) and references therein). These studies have revealed that existing prediction methods for protein relative stability are particularly prone to error in cases where target mutations involved buried polar-polar, polar-charged, and charged-charged interactions, and the introduction of buried, unsatisfied hydrogen-bonding partners. The authors identified two areas for improvement of existing energy functions:

1. better description of amino-acid desolvation and formation of favorable buried polar interactions upon protein folding, and
2. better modeling of the unfolded state.

For each WT and naturally occurring or engineered variant protein, the change in enthalpy upon unfolding is a measure of structural stability. In this work, we report what to our knowledge is a new method to predict relative enthalpic stabilities for variants with respect to WT proteins. To estimate the difference in enthalpy between the folded and unfolded state ensembles from computational studies, good models that represent these states are needed for both. While the three-dimensional structure of the folded state is usually obtained from x-ray or NMR structural studies, a model for the unfolded state is not usually well defined and reliable assessment of the energetics of the unfolded state is an outstanding problem that still hinders theoretical predictions of protein folding and stability (10). Many studies that use models for the unfolded state adopt the random or statistical coil models. In the random coil model, free rotations can take place around every bond, similar to those occurring in a small molecule (11,12). Other workers have defined the random coil state as a well-defined reference state in which no side-chain-to-side-chain interactions are present (13) or have assumed that the Φ- and Ψ-backbone torsional angles of each residue in a random coil are independent of the (Φ, Ψ) angles of every other residue (14). Here we use a simplified model

Submitted May 31, 2012, and accepted for publication August 17, 2012.

*Correspondence: cbp@purdue.edu

Editor: Michael Feig.

0006-3495/12/10/1762/12 $2.00

http://dx.doi.org/10.1016/j.bpj.2012.08.048

of the unfolded state that combines the two previously described characteristics, in that the unfolded state is a collection of peptides defined by the specific amino-acid composition of the protein.

The model uses a peptide-cocktail normalization for the unfolded state that allows for comprehensive estimates of relative enthalpy changes in the intrasolute, solute-solvent, and intrasolvent interaction energies upon protein folding/unfolding. It differs conceptually from transfer-based models (15) where effective hydration protein-to-solvent transfer coefficients are obtained assuming a homogeneous protein environment and group additivity. This new, to our knowledge, model, based on direct energy calculations from all-atom molecular dynamics (MD) simulations of the folded and unfolded states, does not make any assumptions about the homogeneity of the protein interior because it considers explicitly each amino acid's individual local environment in both folded and unfolded states and uses explicit solvent to model protein-solvent interactions.

For more than two decades, researchers have successfully used computer simulations with explicit solvent to separately estimate enthalpic and entropic contributions to free energy (16–22) and make progress in understanding underlying physical principles governing protein unfolding. Following their lead, with the model proposed here applied to T4 lysozyme and its variants, we examine the thermodynamic consequences of burying charged groups in the interior of the protein by assessing the relative change in the enthalpic stability for the whole variant, and individually, for each component amino acid.

For the test systems investigated here, we find that while overall structure is well preserved upon mutation, mutations at a given site have both local and long-range effects on the unfolding enthalpies. The results for the relative change in the enthalpy of unfolding, $\Delta\Delta H^U$, are in good agreement with experimental data. This model for the prediction of relative enthalpic stability upon mutation overcomes deficiencies identified for other methods as it includes physical models for the description of amino-acid desolvation and the formation of favorable buried polar interactions as well as reasonable modeling of the unfolded state. The general method presented here for assessing protein-relative enthalpic stability and residue-specific contributions to the relative stability could prove valuable for providing basic rules for rational protein design and engineering.

MODEL AND THEORY

The unfolded state

The unfolded state model is a collection of amino acids defined by the protein-specific sequence. The conformational space of each amino acid in the unfolded state is sampled through MD simulations of the corresponding amino-acid dipeptide in explicit solvent and salt. Each amino-acid dipeptide, AAd, is CH_3–CO–NH–C_αR–CO–NH–CH_3, where R is the specific amino-acid side chain. This dipeptide model for the unfolded state captures differences in the environment between the solvated and folded state of a residue, as illustrated with the comparison in conformational distribution shown in Section S1 in the Supporting Material. While the model is attractive because of its simplicity and robustness, it may not constitute an ideal model for all proteins in the unfolded state, in that it cannot account for transiently stabilized protein conformations.

Indeed, the suitability of the random coil model to represent the unfolded state is still an open question and subject of debate since the early work of Tanford et al. (23) and the more recent literature (13,24–28). However, recent results from small angle x-ray scattering (29,30) show that under certain solution conditions, such as neutral pH and large protein total charge, unfolded proteins are well described by an excluded volume random coil ensemble. T4 lysozyme is a highly charged protein, with a total charge of $+9e$ at pH 7. We work under the assumption that the T4 lysozyme-specific collection of 162 dipeptides serves as a reasonable approximation for the unfolded-state ensemble of T4 lysozyme.

Model for assessing protein enthalpic stability

Protein enthalpic stability is the change in enthalpy upon protein unfolding, ΔH^U, defined as the difference between the enthalpy of the unfolded (H^u) and folded (H^f) states:

$$\Delta H^U = H^u - H^f. \tag{1}$$

When the folded state is more enthalpically stable than the unfolded state, $\Delta H^U > 0$. At the infinite dilution limit, the total enthalpy of a solvated protein solution can be partitioned as a sum of intraprotein $E(p,p)$, protein-solvent $E(p,s)$, and solvent-solvent $E(s,s)$ interactions,

$$H_{tot} = E(p,p) + E(p,s) + E(s,s), \tag{2}$$

where $E(p,p)$, $E(p,s)$, and $E(s,s)$ are ensemble averages of respective interaction energies. This decomposition is valid for both folded and unfolded states. The change in enthalpy upon protein unfolding is a sum of the corresponding differences in enthalpy between the unfolded and folded states:

$$\Delta H^U = \Delta E^U(p,p) + \Delta E^U(p,s) + \Delta E^U(s,s). \tag{3}$$

The difference in the change in unfolding enthalpy between the WT and variant proteins, $\Delta\Delta H^U$, is a measure of relative stabilization/destabilization of the folded state upon mutation, and is defined as

$$\Delta\Delta H^U = \Delta H^U_{mut} - \Delta H^U_{WT}, \tag{4}$$

where ΔH^U_{mut} and ΔH^U_{WT} are the enthalpies of unfolding of the variant and WT proteins. If $\Delta\Delta H^U < 0$, the variant is less stable than WT, and when $\Delta\Delta H^U > 0$, the variant is more stable than WT.

To gain insight into the roles of individual amino acids in protein destabilization and a more detailed understanding of how each mutation affects protein stability, it is useful to decompose the relative enthalpic stability by estimating contributions from each individual amino acid:

$$\Delta\Delta H^U = \sum_{i=1}^{N_{AA}} \Delta\Delta H^{U,i}. \quad (5)$$

To this end, it is convenient to consider two separate contributions to enthalpy: a component that is related to the protein and a second component related to solvent. The second term in Eq. 3, $\Delta E^U(p,s)$, reflects the difference in the strength of the protein-solvent interactions in the folded and unfolded states. Because of the chosen decomposition of enthalpy into protein and solvent related contributions, the protein-solvent interaction energy is here divided equally between terms representing protein contributions and solvent contributions to ΔH^U. Therefore, the total change in the enthalpy of the system upon protein unfolding, ΔH^U, is separated into two components:

1. the change in enthalpy of component amino acids due to changes in the local neighborhood, or context-dependent environmental changes contributed by the protein, ΔH^U_{env}; and
2. the hydration-solvent reorganization contribution associated with the changes in solvent-solvent and solute-solvent interactions caused by the insertion of the solute, ΔH^U_{hyd-sr}.

With these, we have

$$\Delta H^U = \Delta H^U_{env} + \Delta H^U_{hyd-sr}, \quad (6)$$

where

$$\Delta H^U_{env} = \Delta E^U(p,p) + \frac{1}{2}\Delta E^U(p,s) \quad (7)$$

and

$$\Delta H^U_{hyd-sr} = \Delta E^U(s,s) + \frac{1}{2}\Delta E^U(p,s). \quad (8)$$

The factor $1/2$ in Eqs. 7 and 8 corresponds to equally partitioning the nonbonded protein-solvent interaction between protein and solvent contributions. The total change in the enthalpy of unfolding can then be calculated by summing specific environmental and hydration-solvent reorganization contributions for all amino acids in the protein:

$$\Delta H^U = \sum_{i=1}^{N_{AA}} \left(\Delta H^{U,i}_{env} + \Delta H^{U,i}_{hyd-sr}\right). \quad (9)$$

The environmental component, $\Delta H^{U,i}_{env}$

The change in enthalpy upon protein unfolding is in part due to the change in the local environment of individual amino acids in the folded and unfolded states. To probe the local, context-dependent environment, the environmental component of the enthalpy for each amino acid i in the folded (f) or unfolded (u) state is estimated as

$$H^{x,i}_{env} = E^{x,i}(i,i) + \frac{1}{2}\left(E^{x,i}(i,j) + E^{x,i}(i,s)\right), \quad (10)$$

where x denotes the state of the system (f or u), the first term represents the intra-amino-acid interaction, and the second term includes interactions of amino acid i with all other amino acids in the protein, j, excluding itself, and the solvent. Please note that in this decomposition, all other amino acids in the protein ($i \neq j$) and all solvent molecules contribute to the environment of amino acid i.

For the folded state, the sum over all the amino acids in the protein, N_{AA},

$$H^f_{env} = \sum_{i=1}^{N_{AA}} \left(E^{f,i}(i,i) + \frac{1}{2}\left(E^{f,i}(i,j) + E^{f,i}(i,s)\right)\right), \quad (11)$$

is exactly the sum of intraprotein, (p,p), and half the protein-solvent, (p,s), interactions:

$$H^f_{env} = E^f(p,p) + \frac{1}{2}E^f(p,s). \quad (12)$$

For the unfolded state, u, the total environmental component can be calculated as

$$H^u_{env} = \sum_{i=1}^{N_{AA}} \left(E^{u,i}(i,i) + \frac{1}{2}\left(E^{u,i}(i,d) + E^{u,i}(i,s)\right)\right). \quad (13)$$

$E^{u,i}(i,d)$ denotes the time average of the interaction energy of amino acid i with the rest of the dipeptide and $E^{u,i}(i,s)$ is the interaction of the amino acid i in the corresponding dipeptide with the solvent. Ultimately, the environment specific, amino-acid-based transfer enthalpies, from the unfolded to folded state, are

$$\Delta H^{U,i}_{env} = \langle H^{u,i}_{env}\rangle - H^{f,i}_{env}, \quad (14)$$

where $\langle\ldots\rangle$ denotes time averages. It is worth reemphasizing here that these amino-acid-based transfer enthalpies, as defined above, are context-/environment-specific and do not assume a homogeneous environment in either folded or unfolded states.

The hydration-solvent reorganization component, $\Delta H^{U,i}_{hyd-sr}$

Properties of solvent molecules in contact with a solute differ from bulk solvent properties. Traditionally, the protein-solvent interaction energy has been associated with the so-called hydration energy (16). In addition, the change in solvent-solvent interaction energy upon solute insertion is often referred to as the solvent reorganization enthalpy and is denoted as λ_s. The solvation energy, H_{solv}, is defined as the sum of the hydration and solvent reorganization energies (19,20). Hence, because of its historical connections to both the traditional hydration energy and solvent reorganization energy, we refer to the second term in the proposed decomposition of ΔH^U that incorporates solvent-related effects as the hydration-solvent reorganization, or hyd-sr, term.

In this work, the change in hydration-solvent reorganization enthalpy upon protein unfolding is the change in the enthalpy of the solvent (including half of the protein-solvent interaction) when the protein undergoes the unfolding transition,

$$\Delta H^{U}_{hyd-sr} = H^{u}_{hyd-sr} - H^{f}_{hyd-sr}.$$

When the protein folds, part or all of each amino-acid hydration shell (i.e., all solvent molecules that are different from bulk solvent) in the unfolded state is lost and there is a net transfer of solvent molecules back to bulk. For solvent molecules in the hydration shell of amino acid i, the change in hydration-solvent reorganization enthalpy upon protein unfolding is

$$\Delta H^{U,i}_{hyd-sr} = \Delta E^{U,i}(s,s) + \frac{1}{2}\Delta E^{U}(i,s), \quad (15)$$

and more specifically,

$$\Delta H^{U,i}_{hyd-sr} = E^{u,i}(s,s) - E^{f,i}(s,s) + \frac{1}{2}\left(E^{u,i}(i,s) - E^{f,i}(i,s)\right). \quad (16)$$

Let us consider Eq. 16 for the ideal case where residue i becomes deeply buried in the protein folded state and no longer interacts with the solvent. In this case, all solvent molecules in the hydration shell of this specific residue in the unfolded state are released to bulk solvent when the protein folds. For this particular case, the change in the hydration-solvent reorganization enthalpy for the specific amino acid that becomes deeply buried upon folding will be a maximum possible change, $\Delta H^{U,i,max}_{hyd-sr}$. This is because the last term in Eq. 16, $E^{f,i}(i,s)$, reduces to zero as the amino acid no longer interacts with the solvent, and the second term, $E^{f,i}(s,s)$, takes on bulklike values, i.e., $E^{f,i}(s,s) = E^{Bulk}(s,s)$, as all the solvent molecules previously interacting with this particular amino acid have returned to bulk. Therefore, for a deeply buried amino acid, Eq. 16 describes the maximum possible change in hydration-solvent reorganization enthalpy and this quantity can be evaluated as

$$\Delta H^{U,i,max}_{hyd-sr} = \langle E^{u,i}(s,s)\rangle + \frac{1}{2}\langle E^{u,i}(i,s)\rangle - \langle E^{Bulk}(s,s)\rangle. \quad (17)$$

Note that the difference between the first and last terms in Eq. 17, $\langle E^{u,i}(s,s)\rangle$ and $\langle E^{Bulk}(s,s)\rangle$, is exactly the change in solvent-solvent interaction energy upon solute insertion, namely the solvent reorganization energy, λ_s. If the whole $\langle E^{u,i}(i,s)\rangle$ term was counted toward the hydration-solvent reorganization enthalpy, $\Delta H^{U,i,max}_{hyd-sr}$ would be the traditional solvation energy, i.e., $\langle E(i,s)\rangle + \lambda_s$. With our specific decomposition of the total enthalpy in environmental and hydration-solvent reorganization components,

$$\Delta H^{U,i,max}_{hyd-sr} = \frac{1}{2}\langle E(i,s)\rangle + \lambda_s.$$

$\Delta H^{U,i,max}_{hyd-sr}$ can be calculated directly from bulk solvent and dipeptide solutions MD simulations.

To estimate the change in the hydration-solvent reorganization enthalpy for partially buried or solvent-exposed residues, $\Delta H^{U,i}_{hyd-sr}$, one can use a measure of the remaining hydration or solvent exposure for each amino acid in the protein and the maximum possible changes, $\Delta H^{U,i,max}_{hyd-sr}$, determined as described above. One measure of solvent exposure for each amino acid in the folded state, SE^i, can be defined as the ratio between the solvent-accessible surface areas (SASA),

$$SE^i = \frac{\langle SASA^{f,i}\rangle}{\langle SASA^{u,i}\rangle},$$

where $\langle SASA^{f,i}\rangle$ and $\langle SASA^{u,i}\rangle$ are the amino-acid solvent-accessible solvent areas in the folded and unfolded states, respectively. Using these estimates of amino-acid solvent exposure in the folded state (normalized with respect to the unfolded state exposures) and the estimated maximum possible changes in $\Delta H^{U,i,max}_{hyd-sr}$, the change in hydration-solvent reorganization enthalpy upon unfolding for solvent molecules associated with each component amino acid i can be approximated as

$$\Delta H^{U,i}_{hyd-sr} = \left(1 - SE^i\right)\Delta H^{U,i,max}_{hyd-sr}. \quad (18)$$

Practical estimate of $\Delta H^{U,i,max}_{hyd-sr}$

In practice, to estimate $\Delta H^{U,i,max}_{hyd-sr}$ for each amino acid i, we use Eq. 17 and time averages for the intrasolvent $E^{u,i}(s,s)$, dipeptide-solvent, $E^{u,i}(d,s)$, and bulk solvent interactions, $E^{Bulk}(s,s)$, respectively, from MD simulations of dipeptide solutions and bulk solvent. Note that $E^{u,i}(p,s)$ in Eq. 17 is replaced with $E^{u,i}(d_i,s)$ (where d_i stands for the dipeptide

corresponding to amino acid i) to denote the specific dipeptide-solvent interaction used in this particular application. However, the straightforward application of Eq. 17 gives $\Delta H_{hyd-sr}^{U,d_i,max}$ for the whole dipeptide molecule, including the central amino acid, the flanking –NH– and –CO– groups, and the terminal blocking methyl groups. To separate the contributions of various amino-acid side chains, $\Delta H_{hyd-sr}^{U,SC_i,max}$, we subtract the maximum hydration contribution for Glyd, $\Delta H_{hyd-sr}^{U,Glyd,max}$, from each corresponding $\Delta H_{hyd-sr}^{U,d_i,max}$, as

$$\Delta H_{hyd-sr}^{U,SC_i,max} = \Delta H_{hyd-sr}^{U,d_i,max} - \Delta H_{hyd-sr}^{U,Glyd,max}. \quad (19)$$

And finally, to obtain an estimate for the maximum hydration contribution for the amino-acid backbone (*bb*, i.e., $-NH-C_\alpha H-CO-$), $\Delta H_{hyd-sr}^{U,bb,max}$, we assume that interactions of Glyd (i.e., CH_3–CO–NH–$C_\alpha H_2$–CO–NH–CH_3) approximate those of two amino-acid backbones plus an Ala side chain (i.e., one methyl group), or, alternatively, that interactions of Alad approximate those of two amino-acid backbones plus two Ala side chains. These values for maximum contributions from amino-acid side chains and backbone, together with the respective solvent exposures, are then utilized in Eq. 18 to estimate $\Delta H_{hyd-sr}^{U,i}$ for each amino-acid side chain and backbone. Total hydration-solvent reorganization contributions for each component amino acid i are obtained by summing over enthalpy estimates for side chain and backbone.

Simulation systems and protocols

MD simulations were calculated for 25 amino-acid dipeptides corresponding to all naturally occurring amino acids in their protonated and unprotonated forms (Table 1). Each dipeptide solution contains 1094 water molecules as well as counterions, Na^+ and Cl^-, to neutralize charge and give a salt concentration of 0.2 M. In addition, bulk solvent—i.e., pure water and 0.2 M NaCl solutions—was also simulated. Each of the dipeptide and bulk solvent solutions was simulated for 101 ns.

A modified T4 lysozyme was engineered by Dao-pin et al. (31) to promote two-state folding behavior by removing a disulfide bond with the substitutions C54T and C97A. Crystallographic coordinates for this modified T4 lysozyme (PDB:1L63) were the basis for simulations of wild-type T4 lysozyme with His^{31} protonated, T4(H31^p), or neutral, T4(H31^n). Two mutations, Met^{102} to Lys (M102K) and Leu^{133} to Asp (L133D) were engineered by Dao-pin et al. (31) for folding studies, and the crystallographic structure determined for the M102K variant (PDB:1L54) was the basis for simulations of T4(M102K,H31^p). MD simulations for the folded states of T4 lysozyme wild-type with H31 protonated, T4(H31^p), or H31 unprotonated, T4(H31^n), and two variants, T4(L133D,H31^n) and T4(M102K,H31^p), were calculated for a 74-ns time period. The protein simulations include 13,000 water molecules or more, with a salt concentration of 0.1 M (Table 1). Trajectories were calculated with the CHARMM22 force-field parameters (32) and CMAP (33) using the CHARMM program (34) and an NPT ensemble, at 300 K, under periodic boundary conditions with particle-mesh Ewald method (35). Other details of the simulation protocol appear in the Supporting Material.

Solvent-accessible solvent areas in the folded and unfolded states are calculated from MD trajectories using CHARMM and a probe radius of 1.4 Å. Interaction energies ($E(p, p)$, $E(p, s)$, and $E(s, s)$) used here to estimate the changes in enthalpy were calculated from postprocessing MD trajectories for proteins and dipeptide solutions, using the CHARMM22 force field with a switching function from 12 to 14 Å applied to the van der Waals and Coulomb's electrostatic interactions.

TABLE 1 Details of the MD simulations in the folded and unfolded states

System	Solute	Time (ns)	N_w*	N_{Na+}	N_{Cl-}
Dipeptides†	Apolar:				
	alad, glyd, iled, leud, phed	101	1094	3	3
	prod, trpd, vald	101	1094	3	3
	Polar:				
	asnd, asppd, cysd, glnd	101	1094	3	3
	glupd, hsdd, hsed	101	1094	3	3
	lsnd, metd, serd, thrd, tyrd	101	1094	3	3
	Basic:				
	argd, hspd, lysd	101	1094	2	3
	Acidic:				
	glud, aspd	101	1094	3	2
Proteins	T4(H31^p)	74	13,757	9	18
	T4(H31^n)	74	13,753	10	18
	T4(L133D,H31^n)	74	13,724	11	18
	T4(M102K,H31^p)	74	13,000	8	18
Pure water	TIP3P	101	1094	0	0
Salty water	TIP3P	101	1094	3	3

*N_w, N_{Na+}, and N_{Cl-} are the numbers of water, Na^+, and Cl^- molecules.
†Dipeptides representing amino acids in alternative protonation states were simulated for completeness: protonated GLU, glupd; protonated ASP, asppd; the deprotonated LYS, lsnd; histidine protonated at C_ε, hsed; histidine protonated at C_δ, hsdd; and protonated HIS, hspd.

RESULTS

Relative changes in unfolding enthalpies of variants of T4 lysozyme with respect to WT, $\Delta\Delta H^U$, are predicted as outlined in Model and Theory, above, and compared with experimental data measured by Dao-pin et al. ((31); see their Table II therein). The detailed information of the simulations is then exploited to obtain a molecular interpretation of the thermodynamic effect of mutating buried amino acids to charged amino acids.

To start to validate our new (to our knowledge) method for predicting protein relative stability, we evaluated the effect of protonation of Histidine 31 (H31), which has

a highly elevated pK_a of 9.1 (36,37); the protonated form, T4(H31^p), is stabilized relative to the neutral form, T4(H31^n), by formation of a salt bridge with D70 (Fig. 1) (36). In addition, two variants are examined: an apolar to charged amino-acid substitution with H31 deprotonated, T4(L133D,H31^n), and a polar to charged amino-acid substitution with protonated H31, T4(M102K,H31^p). Both mutations lead to buried charged residues with altered pKa values; the pKa of K102 is 6.5 and that of D133 is 6.2 (31,36). Our computations evaluate the effect of burying charge on the relative enthalpic stability of these variants with respect to the WT protein, T4(H31^p).

Good correlation with experimental relative change in enthalpic stability

For WT and variants of T4 lysozyme, the environmental and hydration-solvent reorganization contributions to the enthalpy of unfolding were calculated according to Eqs. 14 and 18, and $\Delta\Delta H^U$, the relative enthalpic stability with respect to the reference WT T4 lysozyme T4(H31^p), was calculated according to Eq. 4. $\Delta\Delta H^U_{calc}$ values are determined from the 101-ns MD simulations for each dipeptide solution and 74 ns of each folded protein. The first 1 ns and 14 ns, were dedicated to equilibration of dipeptide and protein solutions, respectively. Interaction energies for dipeptide solutions and bulk solvent were calculated as 100-ns time averages and used in the calculation of $\Delta\Delta H^{max}_{hyd}$. The error in mean values of $\Delta\Delta H^U_{calc}$ was estimated from block averaging by separating the folded protein trajectories into five blocks of 12 ns. The estimated standard errors for the calculated $\Delta\Delta H^{calc}$ for the three variants with respect to T4(H31^p) are as follows: $\pm$3.2 kcal/mol for T4(H31^n); $\pm$3.2 kcal/mol for T4(L133D, H31^n); and $\pm$4.3 kcal/mol for T4(M102K,H31^p).

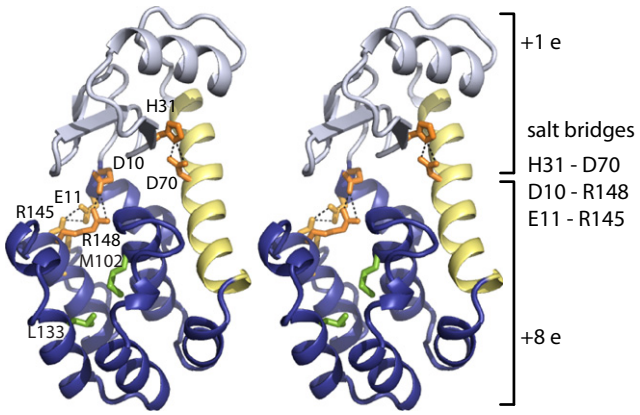

FIGURE 1 Structure of T4 lysozyme (PDB:1L63). Salt bridges between H31^p-D70, D10-R148, and E11-R145 are indicated in the figure (*dotted lines*). N- and C-terminal domains (*gray* (*top*) and *blue* (*bottom*), respectively) are connected through an interdomain linker (vertical helix on right). Side chains of residues at the mutation sites L133D and M102K are shown (*green*).

To compare the values predicted from simulations with experimental data, the enthalpy of unfolding at 300 K was estimated from the enthalpy change measured at the temperature of unfolding, T_m (Dao-pin et al. (31); see their Table II therein) using $\Delta C_p = 2.5$ kcal/(mol K). The experimental values of ΔH for T4(H31^p) and T4(M102K,H31^p) are those measured at pH = 5.3 where the wild-type protein is most stable (39.3 and 8.0 kcal/mol at 300 K, respectively). Values for T4(H31^n) and T4(L133D,H31^n) correspond to experimental ΔH values at pH = 10.4, above the pK_a of H31 (27.3 and 42.4 kcal/mol at 300 K, respectively). $\Delta\Delta H^{exp}$ for T4(L133D,H31^n) relative to T4(H31^p) includes enthalpic contributions for both the substitution of L133 for D133 ($-$15.1 kcal/mol) and protonation of H31 ($-$12.0 kcal/mol).

Table 2 shows the calculated relative enthalpic stabilities, $\Delta\Delta H^U_{calc}$, and the experimental values for $\Delta\Delta H^U_{exp}$ at 300 K. Relative enthalpies of unfolding for T4(H31^n), T4(L133D,H31^n), and T4(M102K,H31^p) are with respect to the reference WT, T4(H31^p). The calculated $\Delta\Delta H^U_{calc}$ for the relative change in enthalpy of T4(H31^n), corresponding to the deprotonation of H31 is $-$6.3 kcal/mol. T4(L133D,H31^n) and T4(M102K,H31^p) are destabilized by $-$12.7 and $-$15.6 kcal/mol, respectively. The experimental values for $\Delta\Delta H^{exp}$ estimated as described above are: $-$12.0, $-$27.7, and $-$31.3 kcal/mol for T4(H31^n), T4(L133D,H31^n), and T4(M102K,H31^p), respectively.

As seen in Fig. 2 and Table 2, the relative ranking of the predicted enthalpic stabilities, T4(H31^p) > T4(H31^n) > T4(L133D,H31^n) > T4(M102K,H31^p), matches remarkably well with the experimental ranking. It is reassuring to note that, for T4 lysozyme and its variants, the same relative ranking in stability is indicated by the experimental changes in the free energy upon mutation as shown in Table 2, last column. While we obtain a very good correlation between the experimental and calculated relative enthalpies, the magnitude of the calculated change in enthalpy is roughly half that of the experimental values.

Lower values of ΔC_p, such as $\Delta C_p = 1.8$ kcal/(mol K), lead to larger differences between the experimental and

TABLE 2 Relative enthalpic stability of T4 lysozyme variants with respect to WT calculated from 74-ns simulations and from experiment

	Calculated*			Experimental†	
Protein	$\Delta\Delta H^U_{env}$	$\Delta\Delta H^U_{hyd}$	$\Delta\Delta H^U_{calc}$	$\Delta\Delta H^U_{exp}$	$\Delta\Delta G^U_{exp}$
T4(H^{p}31)	0.0	0.0	0.0	0.0	0.0
T4(H^{n}31)	$-$4.6	$-$1.7	$-$6.3	$-$12.0	$-$1.7
T4(L133DH^{n}31)	$-$7.8	$-$4.9	$-$12.7	$-$24.2	$-$5.7
T4(M102KH^{p}31)	$-$14.3	$-$1.3	$-$15.6	$-$31.3	$-$6.9

*Experimental and calculated enthalpies of unfolding and the change in enthalpies and free energies of unfolding are reported in kcal/mol. $\Delta\Delta H^U_{calc}$ is the sum of environmental and hydration-solvent reorganization contributions.

†$\Delta\Delta H^U_{exp}$ is from the data of Dao-pin et al. (31) extrapolated to 300 K with $\Delta C_p = 2.5$ kcal/(mol K). See text.

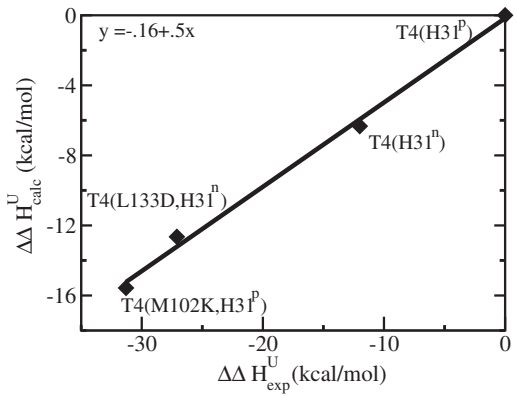

FIGURE 2 Relative enthalpic stability of the WT and variant T4 lysozyme proteins showing the calculated values plotted against the experimental values.

calculated ΔH^U, but the ranking of the relative enthalpies remains the same. More-precise experimental determination of ΔC_p for each individual protein is desirable for a more accurate comparison of the absolute values of experimental and calculated enthalpies, but these measurements are not currently available.

Impact of change in protein local and total charge upon mutation

Somewhat surprisingly, the variant T4(L133D,H31^n) is more stable than T4(M102K,H31^p). Because L133D substitutes an apolar amino acid for a charged one, whereas M102K is a polar to charged mutation, from a simplistic point of view one would expect that a more drastic change of introducing a charged group into a chemical environment that naturally accommodates an apolar group would be more damaging to protein stability than a polar-to-charged mutation. Consideration of the net charge on the protein domain provides a partial explanation for this behavior.

The total charge, Q_{tot}, of T4(H31^p) is $+9e$; however, the charge is not uniformly distributed between the N and C termini domains. Residues 1–74 of the N-terminal domain carry a net charge of $Q_{tot}^{N_{term}} = +1e$ from 23 charged groups, while residues 75–162 of the C-terminal domain carry a total charge of $Q_{tot}^{C_{term}} = +8e$ from 22 charged groups. That is, while the total number of charged groups on the N and C domains are similar, the net charge is overall balanced in the N domain but not for the C domain. For the more destabilizing mutation, M102K, the introduction of an additional positive charge by K102 into the C-terminal domain, already rich in basic residues, raises not only the total charge of the protein but also the local charge of the C-terminal domain from $+8e$ to $+9e$, making the local environment in this domain even more repulsive. By contrast, the L133D mutation in the C-terminal domain lowers both the local charge of the C-terminal domain from $+8e$ to $+7e$, and the total charge of the protein. We suggest that because the change of M102 to K increases the local charge, this substitution is more disruptive to the overall stability of the protein than the mutation of L133 to D, which decreases the total charge both locally, in the C-terminal domain, and for the whole protein.

That overall electrostatics of the domain may contribute to the effects of introducing buried charge into the C domain is illustrated by a change in a distant salt-bridge interaction between D10 and R148, one of the pillars of T4 lysozyme tertiary structure (Fig. 1). Using the average distance between the side chains as a reflection of the interaction strength, we find that the distance of the center of mass of the carboxylate group of D10 to the guanidinium group of R148 varies among WT T4 lysozyme and the two variants. It is shortest in T4(L133D,H31^n) and equal to 3.9 ± 0.2 Å. By comparison, the average distances for the WT proteins T4(H31^n) and T4(H31^p) are 4.0 ± 0.3 Å and 4.1 ± 0.3 Å. The distance between these partners is 4.6 Å in T4(M102K,H31^p). It is clear that in T4(L133D,H31^n) the D10-R148 salt bridge is well maintained while this charge-charge interaction is weakened in the least stable variant, T4(M102K,H31^p).

Residue-specific relative enthalpic stability profiles

Encouraged by the excellent agreement with experiment shown in Fig. 2, we exploit the molecular detail of MD simulations to examine the contribution to the relative enthalpic stability from each individual residue. Our model has the advantage of a straightforward analysis of the residue-specific contribution to the unfolding enthalpy and thus allows inspection of the specific consequences of any individual mutation, including buried charge. The environmental component is determined for each amino acid i from Eqs. 11 and 13, and using Eq. 14. The residue-based hydration-solvent reorganization component is calculated as the sum over amino-acid backbone and side-chain contributions, using the corresponding time-averaged solvent exposures (SE^i) and $\Delta H_{hyd-sr}^{max,i}$ for side chains from Table S1 (*fifth column*) in the Supporting Material, and the estimated maximum value for the backbone, −1.56 kcal/mol, with Eq. 18. The amino-acid-based relative change in enthalpy, $\Delta\Delta H^{U,i}$, is then calculated according to Eq. 4. Figs. 3–5 show the results of these residue-based, relative enthalpies of unfolding calculations for the unprotonated T4(H31^n) and two variants with respect to T4(H31^p). Although a large number of residues have near-zero values, several residues, in addition to the mutated residue, are perturbed by the change in buried charge. Both contributions that increase (positive $\Delta\Delta H^{U,i}$) or decrease (negative $\Delta\Delta H^{U,i}$) enthalpic stability are observed, and these nonzero values are contributed mostly by $\Delta H_{env}^{U,i}$. In contrast, $\Delta\Delta H_{hyd-sr}^{U,i}$ are close to zero for most residues except for those at the mutated sites. Specific

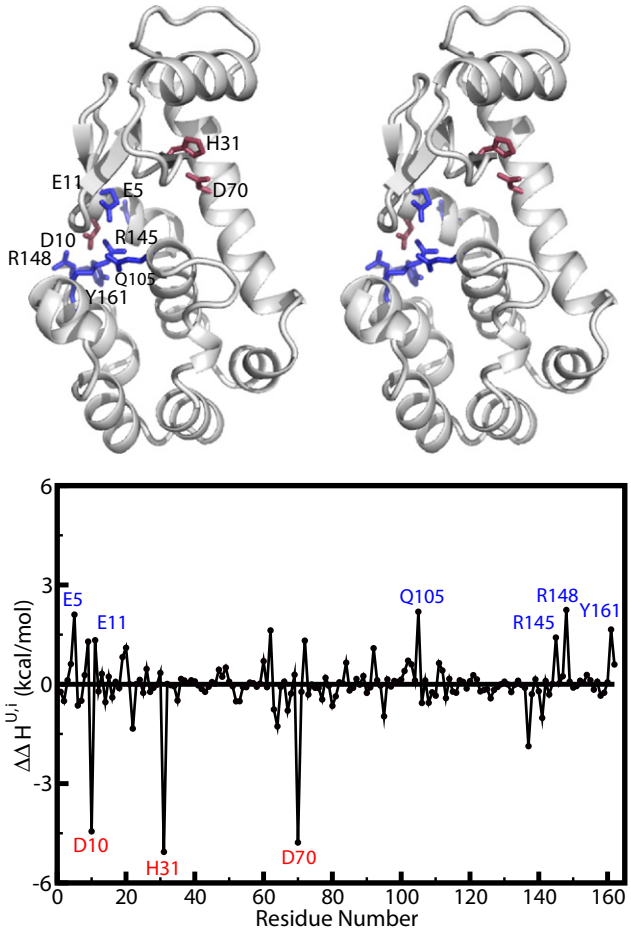

FIGURE 3 Residue-based relative enthalpic stability profile of a deprotonated T4 lysozyme (at H31), T4(H31^n), with respect to WT T4(H31^p) (*lower*). Amino acids with favorable (positive) or unfavorable (negative) contributions to the relative enthalpic stability are specifically indicated (*blue* and *red*, respectively). The structure of the deprotonated WT T4(H31^n), is shown in cartoon representation (*silver*, *upper*). Amino acid with positive contributions to $\Delta\Delta H^U$ are mapped in blue; those with negative contributions are mapped in red.

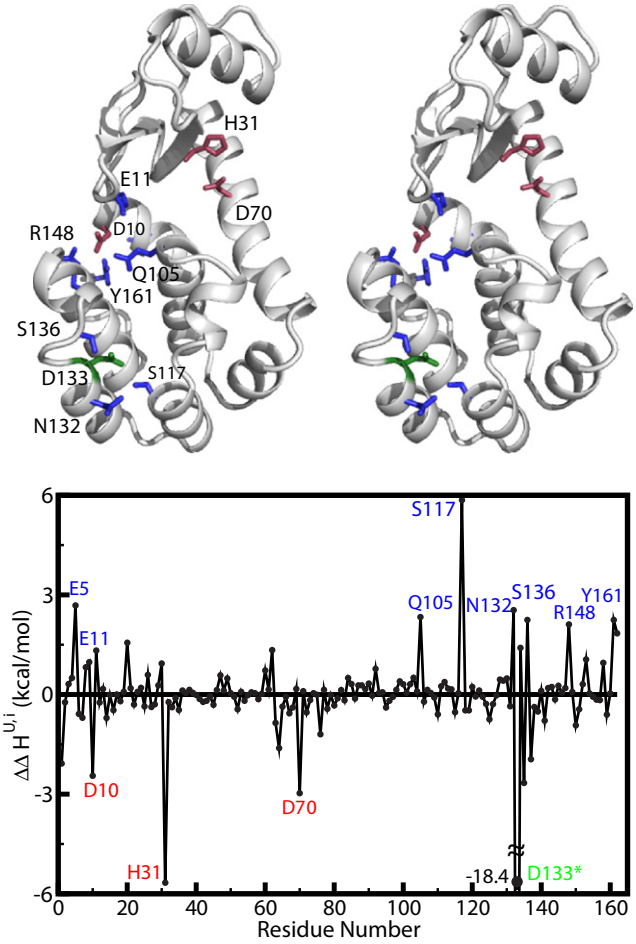

FIGURE 4 Residue-based relative enthalpic stability profile of the variant T4(L133D,H31^n), with respect to WT T4(H31^p) (*lower*). Amino acids with large positive or negative contributions to the relative enthalpic stability are specifically indicated (*blue* and *red*, respectively). The mutated site is shown with an asterisk. (*Upper*) Structure of T4(L133D,H31^n) is shown in cartoon representation, with residues labeled in the bottom plot shown in stick representation (*silver*). The mutated site is shown in green.

amino-acid contributions to $\Delta\Delta H^{U,i}_{hyd-sr}$ are shown in Fig. S14.

We focus the following discussion on residues which contribute over the full extent of the MD trajectories either more or less favorably to enthalpic stability in the variant than in the reference protein T4(H31^p). The behavior is determined from the change in environmental and hydration-solvent reorganization enthalpies, $\Delta H^{U,i}_{env}$ and $\Delta H^{U,i}_{hyd-sr}$, as a function of time. Example time profiles for the total by residue enthalpy change and its components for several amino acids from trajectories of WT and variant proteins are shown in Fig. S7, Fig. S8, Fig. S9, Fig. S10, Fig. S11, Fig. S12, and Fig. S13. The majority of $\Delta H^{U,i}_{env}$ and $\Delta H^{U,i}_{hyd-sr}$ values are converged; however, some of the surface-exposed amino acids are more mobile and exhibit side-chain dihedral transitions so that the corresponding $\Delta H^{U,i}_{env}$ values are not as well converged after 74 ns of MD simulations.

Because the fluctuations in any component of the total energy are much larger than that for the system energy, the errors in $\Delta\Delta H^{U,i}$ per residue are necessarily larger than those for the whole protein and therefore the exact quantitative value of $\Delta\Delta H^{U,i}$ is not well known. Accordingly, we remark on only the residues labeled in Figs. 3–5, which are those that have $\Delta\Delta H^{U,i}$ values that are greater or less than the reference values from T4(H31^p) over the course of the simulation analysis period, and thus exhibit consistent enthalpic differences. These residues indicated in the figures are mapped on the corresponding variant structure (*upper panels* in Figs. 3–5); residues with positive, favorable contributions to the stability are shown in blue and those with negative, destabilizing contributions are

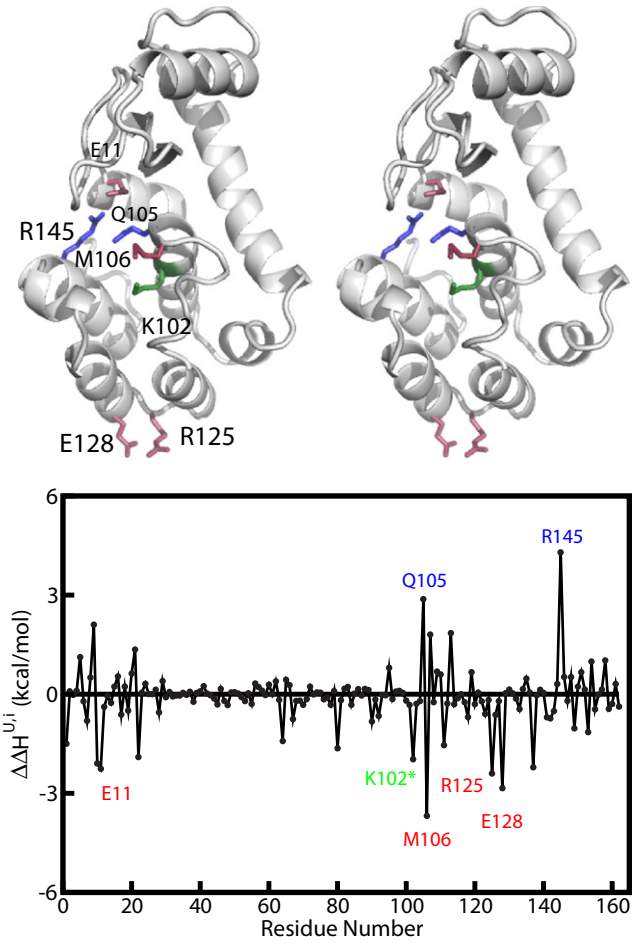

FIGURE 5 Residue-based relative enthalpic stability profile of mutant T4(M102K,H31^P), with respect to WT, T4(H31^P) (*lower*). Amino acids with larger favorable (positive) or unfavorable (negative) contributions to the relative enthalpic stability are specifically indicated (*blue* and *red*, respectively). The mutated site is shown with an asterisk. (*Upper*) Structure of T4(M102K,H31^P) is shown in cartoon representation, with residues labeled in the bottom plot shown in stick representation (*silver*). The mutated site is shown in green.

shown in red. In all cases, both stabilizing and destabilizing contributions to the relative change in enthalpy are observed upon mutation. The total change in enthalpic stability is the net sum of the positive and negative contributions.

The residue profile in Fig. 3, bottom panel, clearly shows that deprotonation of H31^P destabilizes the direct interaction of H31 with D70. Several slightly stabilizing, compensating contributions are also observed for E5, E11, Q105, R145, R148, and Y161. It is evident from these results that the main source of the folded state destabilization of T4(H31^n) is the destruction of the H31-D70 salt bridge upon H31 deprotonation, with each salt-bridge partner destabilized by $\approx$5 kcal/mol for a total destabilizing contribution of $-$9.9 kcal/mol from these partners. Nonetheless, significant longer-range effects from deprotonation of H31 (Fig. 3, *top panel*) are also observed. The effects on the D10-R148 salt bridge (Fig. 1), are overall unfavorable for folding; the result of counterbalancing changes between D10 and R148 gives a net destabilization of $-$2.0 kcal/mol. By contrast, the E11-R145 salt bridge is stabilized as indicated by the +2.8 kcal/mol with positive contributions from both E11 and R145. The total $\Delta\Delta H^U$ value of T4(H31^n) relative to T4(H31^P) is estimated to be $-$6.3 kcal/mol (Table 2). This value is close to estimates for the energy of a single salt bridge (38); however, the results presented here in Figs. 3–5, show that contributions to the total enthalpy from deprotonation and other mutations are dispersed over long distances, not confined to residues in direct contact with the mutated site. This observation highlights the difficulties of interpreting experimental data on unfolding from changes in protein structure alone.

For T4(L133D,H31^n), the substitution L133D leads to a large unfavorable change in the relative enthalpy of unfolding of $-$18.4 kcal/mol at the mutation site (Fig. 4, *bottom panel*); however, many neighboring and distant polar residues surrounding D133 in the C-terminal domain (Q105, S117, N132, S136, R148, and Y161) gain in stability by as much as 6 kcal/mol for S117, to offset the energetic cost of burying a charged amino acid. We note that the other destabilizing effects in Fig. 4 from H31 and D70 are mainly due to the neutral form of H31 used in simulation of T4(L133D,H31^n) as can be seen by comparison with Fig. 3. The polar amino acids in the neighborhood of site 133, most notably S117, S136, and N132, welcome Asp in this position and are involved in favorable interactions with its charged group (Fig. 4, *upper panel*). Thus, several residues in the C-terminal domain are stabilized on average even though the energy of D133 is highly unfavorable. The energetic cost is due to the unfavorable protein environment as well as hydration at site L133D; $\Delta\Delta H_{env}^{U,133}$ is $\sim-$18.4 kcal/mol, and has unusually large fluctuations so that this value is not as well converged as for other residues. (See the Supporting Material for comparisons of $\Delta\Delta H_{env}^{U,i}$ of various residues.) S117 has been previously recognized as a player in T4 lysozyme folded-state stability (9).

Our calculations show that S117, N132, and S136 are all stabilized by the mutation of L133 to D (Fig. 4). In this variant, D133 forms a quadrad with S117, N132, and S136, pulling the three residues closer to each other (by 0.5 Å) and shortening the distance between its charged carboxylate group and the side chains of S117 and N132 by 1.0 and 2.1 Å compared to distances in T4(H31^P), respectively. Therefore, introduction of the carbonyl group of D133 in the neighborhood of these polar side chains leads to local structural reorganization that allows S117 and N132 to switch main interaction partners from each other to D133.

Among the comparisons of buried charge forms of T4 lysozyme studied here, the most destabilizing change is

M102K with an estimated $\Delta\Delta H^{U}_{calc} = -15.6$ kcal/mol for T4(M102K,H31^{p}) (Table 2). The by-residue values (Fig. 5, *bottom panel*) show the contribution from K102 to be ~−2.0 kcal/mol, much less of a penalty than that for burying D133 in T4(L133D,H31^{n}) (Fig. 4); however, the introduction of a basic side chain at position 102 in the overall positively charged C-terminal domain is not compensated by multiple stabilizing interactions of polar residues, as observed for L133D. In contrast, the change to K102 destabilizes the salt bridge of D10-R148 by −1.2 kcal/mol while at the same time enhancing the E11-R145 salt bridge by 2.2 kcal/mol. Individual contributions of partner amino acids in the two salt bridges are as follows: −1.6, −2.2, +4.4, and +0.4 kcal/mol for D10, E11, R145, and R148, respectively.

From Fig. 5, favorable changes in the enthalpy of unfolding are notable for Q105 and R145, while unfavorable, negative changes in the relative stability are observed for E11, E22, M106, R125, E128, and R137. To understand the relative small destabilization at K102 and larger neighboring destabilizing effects at M106, we compared the local structure in the WT and variant proteins and noted that in the WT protein M106 is partly stabilized by a π-type interaction with W138. In T4(M102K,H31^{p}), K102 takes over the π-type interaction with W138 and associates with Q105, while pushing M106 away from its original position in the proximity of W138. R137 is also slightly displaced from its original position in T4(H31^{p}). These structural rearrangements allow K102 to find reasonable accommodation in the core of T4(M102K,H31^{p}), while at the same time displacing and dislocating its near neighbors. E22, R125, E128, and R137 are located on the surface of T4 lysozyme and form stable surface salt bridges in T4(H31^{p}), E22 to R137 and R125 to E128. These surface salt bridges are destabilized in T4(M102K,H31^{p}) and the component amino acids display two-state behavior (bound/unbound) while undergoing large fluctuations in solvent exposure and slower convergence of $\Delta\Delta H^{U,i}$.

It is interesting to note that regions of the protein sequence where the amino acids have larger absolute changes in relative enthalpic stability also stand out in an analysis of amino-acid-based relative change in RMSDs. Residue-based RMSD values for each protein are shown in Fig. S3. We also calculate the residue-based relative degree of structural change by subtracting the time-averaged RMSD value for each amino acid in T4(H31^{p}) from its counterpart in the other three proteins. The results for whole amino acids, backbone, and side chains are shown in Fig. S4, Fig. S5, and Fig. S6. The resulting structural picture discussed in the Supporting Material supports our earlier conclusion about the effect of adding a positive charge in the already crowded, positively charged environment of the C-terminal domain in that many and larger structural changes are observed in the C-terminal domain of T4(M102K,H31^{p}) than in T4(L133D,H31^{n}).

Hydration-solvent reorganization enthalpy

We first consider the maximum hydration-solvent reorganization contributions for the dipeptides, i.e., the change in enthalpy when all water molecules in the hydration shell surrounding one dipeptide are released to bulk. Solvent reorganization energies upon dipeptide molecule insertion, λ_s, solvations energies, H^{i}_{solv}, and the maximum change in hydration-solvent reorganization enthalpies, $\Delta H^{U,i,max}_{hyd-sr}$, are listed in Table S1.

Armed with the estimates for the maximum, amino-acid-based, hydration-solvent reorganization contributions to protein unfolding, we then calculate $\Delta H^{U,i}_{hyd-sr}$ for each residue using Eqs. 17 and 18 as described in Model and Theory, above. Our calculations show that on average, each side chain contributes 0.76 kcal/mol hydration-solvent reorganization enthalpy toward protein folded-state stability.

Considering the potential contribution to the hydration-solvent reorganization enthalpy for a given type of side chain is $\Delta H^{U,SC,max}_{hyd-sr}$ (*fifth column* in Table S1), apolar side chains contribute 1.70 kcal/mol, polar side chains contribute roughly one-third (0.61 kcal/mol), and acidic amino acids oppose side-chain desolvation by −0.12 kcal/mol. An examination of the whole amino-acid maximum contributions to hydration-solvent reorganization (*last column* in Table S1) indicates that overall apolar amino acids favor folding by 0.13 kcal/mol, while polar and charged amino acids contribute unfavorably by −0.95 and −1.69 kcal/mol, respectively.

For the case of complete amino-acid backbone burial upon protein folding, both approximations for the calculation of $\Delta H^{U,bb,max}_{hyd-sr}$ outlined in Practical Estimate of $\Delta H^{U,i,max}_{hyd-sr}$, above, give a negative contribution of −1.56 kcal/mol. That is, in our model the main chain has an unfavorable hydration/solvent reorganization contribution to protein folding. This negative hydration-solvent reorganization contribution from the main chain is most likely compensated by favorable intramolecular interactions, formation of main chain to main chain, and main chain to polar side-chain hydrogen-bonding interactions that lead to stable secondary structure formation such as α-helices and β-sheets and others that hold these secondary structures together. These possibly compensating interactions are part of the environmental enthalpy component in our model.

Because the structure of the variants is maintained close to the WT structure, the relative hydration enthalpy contributions, $\Delta\Delta H^{U,i}_{hyd-sr}$, to the total relative enthalpy is practically limited to contributions from the mutated sites. $\Delta\Delta H^{U}_{hyd-sr}$ for each protein are shown in Table 2 and range in value from −4.9 kcal/mol (for T4(L133D,H31^{n}) to −1.3 kcal/mol for T4(M102K,H31^{p}). Given the overall range of the calculated change in enthalpy, these hydration contributions are significant and should not be neglected when assessing the relative change in enthalpy upon

mutation. The relative changes in amino-acid-based hydration contributions to the unfolding enthalpy of the variants with respect to WT are shown in Fig. S14.

DISCUSSION

Numerous experimental and computational studies aim to draw general rules for predicting the effects of mutations of polar and apolar residues to charged residues (9). The hope is that comparisons between the structure and relative stabilities of variant and WT proteins will lead to an understanding of how individual amino acids contribute to protein stability. While these efforts have been very successful in giving valuable insights into quantitative and qualitative understanding of the overall effects of specific mutations on protein relative stability, a quantitative measure of specific, amino-acid-based contributions has been very difficult to achieve (7,8).

In this article, we outline what to our knowledge is a new method for predicting the relative change in protein enthalpic stability upon mutation and mapping a residue-by-residue relative enthalpic stability of variants with respect to WT. The method yields good agreement in rank order of the enthalpy of unfolding for variants of T4 lysozyme that bury charged groups. One of the great advantages of our method is that it allows for a prediction of relative enthalpic stabilities for each component amino acid. Good estimates of $\Delta\Delta H^{U,i}$ may help guide experimental positive/negative design of proteins leading to desired enhancements of their physical and biological properties.

As seen with these results, while the deprotonated T4 lysozyme and two mutants have well-maintained overall structure, their enthalpic stability is affected: small changes in structure do not necessarily translate into small changes in stability. The overall, macroscopic effect of mutation that can be measured experimentally is the combined result of many positive and negative contributions to stability from individual amino acids, some of them far away from the mutation site.

The mutation of L133 to D is an example of extreme local (at the mutation site) destabilization by the immersion of a charged group in the protein core; however, much of this destabilization is offset by numerous surrounding interactions that lead to locally enhanced enthalpic stability. The mutation decreases both the overall charge on the protein and the local domain (domain C) charge, rendering a softer protein environment. Based on the results presented here, it appears that proteins may tolerate burying charged groups that favor a more charge-balanced local environment. A more detailed analysis of the effect of local versus global effects of mutation to charged amino acids is underway.

Our computational method for predicting the relative enthalpic stability of variants with respect to WT makes progress in areas identified as needing improvement (7,8). Better estimates of the effect of polar and charged amino-acid desolvation and hydrogen-bond formation in the folded state is achieved by the consideration of an environment-specific change in enthalpy upon unfolding and the design of a better model for the unfolded state is achieved by the normalization of the unfolded state as a protein-specific dipeptide ensemble.

Although the method evaluates well the relative enthalpic stabilities of variants and wild-type proteins, further work is needed to identify the leading causes for the overall underestimate of $\Delta\Delta H^{U}$. Two areas of investigation in our laboratory currently include extended sampling of the folded state ensemble and a more rigorous method to approximate changes in hydration-solvent reorganization energies upon protein unfolding. These new tools will allow validation and evaluation of the effectiveness of the method for determining relative change in enthalpy for systems where structure is not as well maintained upon mutation and examine other possible applications of the method such as protein-ligand binding.

SUPPORTING MATERIAL

Five sections, one table, 14 figures, and references (39,40) are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00975-7.

We thank Rosen Center for Advanced Computing at Purdue University for providing computing resources and Markey Center for Structural Biology for continuous support of work in our group.

This work was supported in part by the National Institutes of Health (grant No. AI039639 to C.B.P.).

REFERENCES

1. Yue, P., Z. Li, and J. Moult. 2005. Loss of protein structure stability as a major causative factor in monogenic disease. *J. Mol. Biol.* 353:459–473.
2. Wang, Z., and J. Moult. 2001. SNPs, protein structure, and disease. *Hum. Mutat.* 17:263–270.
3. Carneiro, M., and D. L. Hartl. 2010. Colloquium papers: Adaptive landscapes and protein evolution. *Proc. Natl. Acad. Sci. USA.* 107 (*Suppl 1*):1747–1751.
4. Vertrees, J., J. O. Wrabl, and V. J. Hilser. 2009. An energetic representation of protein architecture that is independent of primary and secondary structure. *Biophys. J.* 97:1461–1470.
5. Manson, A., S. T. Whitten, …, V. J. Hilser. 2009. Characterizing the role of ensemble modulation in mutation-induced changes in binding affinity. *J. Am. Chem. Soc.* 131:6785–6793.
6. Vertrees, J., J. O. Wrabl, and V. J. Hilser. 2009. Energetic profiling of protein folds. *Methods Enzymol.* 455:299–327.
7. Potapov, V., M. Cohen, and G. Schreiber. 2009. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng. Des. Sel.* 22:553–560.
8. Kellogg, E. H., A. Leaver-Fay, and D. Baker. 2011. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins.* 79:830–838.
9. Baase, W. A., L. Liu, …, B. W. Matthews. 2010. Lessons from the lysozyme of phage T4. *Protein Sci.* 19:631–641.
10. Grdadolnik, J., V. Mohacek-Grosev, …, F. Avbelj. 2011. Populations of the three major backbone conformations in 19 amino acid dipeptides. *Proc. Natl. Acad. Sci. USA.* 108:1794–1798.

11. Levinthal, C. 1968. Are there pathways for protein folding? *J. Chem. Phys.* 65:44–45.
12. Tanford, C. 1970. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* 24:1–95.
13. Shortle, D. 1996. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* 10:27–34.
14. Smith, L. J., K. M. Fiebig, …, C. M. Dobson. 1996. The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold. Des.* 1:R95–R106.
15. Sharp, K. A., A. Nicholls, …, B. Honig. 1991. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science.* 252:106–109.
16. Kitchen, D. B., L. H. Reed, and R. M. Levy. 1992. Molecular dynamics simulation of solvated protein at high pressure. *Biochemistry.* 31:10083–10093.
17. Levy, R. M., and E. Gallicchio. 1998. Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annu. Rev. Phys. Chem.* 49:531–567.
18. Cui, Q., and V. H. Smith. 2003. Solvation structure, thermodynamics, and conformational dependence of alanine dipeptide in aqueous solution analyzed with reference interaction site model theory. *J. Chem. Phys.* 118:279–290.
19. Wan, S., R. H. Stote, and M. Karplus. 2004. Calculation of the aqueous solvation energy and entropy, as well as free energy, of simple polar solutes. *J. Chem. Phys.* 121:9539–9548.
20. Lazaridis, T. 2000. Solvent reorganization energy and entropy in hydrophobic hydration. *J. Phys. Chem. B.* 104:4964–4979.
21. Shi, Y., C. Z. Zhu, …, P. Ren. 2012. Probing the effect of conformational constraint on phosphorylated ligand binding to an SH2 domain using polarizable force field simulations. *J. Phys. Chem. B.* 116:1716–1727.
22. Dadarlat, V. M., and C. B. Post. 2008. Contribution of charged groups to the enthalpic stabilization of the folded states of globular proteins. *J. Phys. Chem. B.* 112:6159–6167.
23. Tanford, C., K. Kawahara, and S. Lapange. 1967. Proteins as random coils. I. Intrinsic viscosities and sedimentation coefficients in concentrated guanidine hydrochloride. *J. Am. Chem. Soc.* 89:729–736.
24. Dill, K. A., and D. Shortle. 1991. Denatured states of proteins. *Annu. Rev. Biochem.* 60:795–825.
25. Wright, P. E., and H. J. Dyson. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293:321–331.
26. Baldwin, R. L. 2002. A new perspective on unfolded proteins. *Adv. Protein Chem.* 62:361–367.
27. Goldenberg, D. P. 2003. Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* 326:1615–1633.
28. Robic, S., M. Guzman-Casado, …, S. Marqusee. 2003. Role of residual structure in the unfolded state of a thermophilic protein. *Proc. Natl. Acad. Sci. USA.* 100:11345–11349.
29. Wang, Y., J. Trewhella, and D. P. Goldenberg. 2008. Small-angle x-ray scattering of reduced ribonuclease A: effects of solution conditions and comparisons with a computational model of unfolded proteins. *J. Mol. Biol.* 377:1576–1592.
30. Kohn, J. E., I. S. Millett, …, K. W. Plaxco. 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 101:12491–12496.
31. Dao-pin, S., D. E. Anderson, …, B. W. Matthews. 1991. Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry.* 30:11521–11529.
32. MacKerell, Jr., A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
33. MacKerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* 126:698–699.
34. Brooks, B., R. Bruccoleri, …, M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* 4:187–217.
35. Darden, T. A., D. M. York, and L. G. Pedersen. 1993. Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
36. Anderson, D. E., W. J. Becktel, and F. W. Dahlquist. 1990. pH-induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry.* 29:2403–2408.
37. Anderson, D. E., J. Lu, …, F. Dahlquist. 1993. The folding, stability and dynamics of T4 lysozyme: a perspective using nuclear magnetic resonance. *In* NMR of Proteins. G. M. Clore and A. M. Gronenborn, editors. CRC Press, Boca Raton, FL. 258–304.
38. Kumar, S., C. J. Tsai, and R. Nussinov. 2002. Maximal stabilities of reversible two-state proteins. *Biochemistry.* 41:5359–5374.
39. Feig, M. 2008. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *J. Chem. Theory Comput.* 4:1555–1564.
40. Xiong, K., E. K. Asciutto, …, S. A. Asher. 2009. Salt dependence of an α-helical peptide folding energy landscapes. *Biochemistry.* 48:10818–10826.