doi:10.1016/j.jmb.2009.08.010 J. Mol. Biol. (2009) 393, 527–542

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics with a Grain of Salt

Ariel Azia and Yaakov Levy*

Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Received 24 April 2009; received in revised form 1 August 2009; accepted 6 August 2009
Available online 13 August 2009

In recent years, a growing number of protein folding studies have focused on the unfolded state, which is now recognized as playing a major role in the folding process. Some of these studies show that interactions occurring in the unfolded state can significantly affect the stability and kinetics of the protein folding reaction. In this study, we modeled the effect of electrostatic interactions, both native and nonnative, on the folding of three protein systems that underwent selective charge neutralization or reversal or complete charge suppression. In the case of the N-terminal L9 protein domain, our results directly attribute the increase in thermodynamic stability to destabilization of the unfolded ensemble, reaffirming the experimental observations. These results provide a deeper structural insight into the ensemble of the unfolded state and predict a new mutation site for increased protein stability. In the second case, charge reversal mutations of RNase Sa affected protein stability, with the destabilizing mutations being less destabilizing at higher salt concentrations, indicating the formation of charge–charge interactions in the unfolded state. In the N-terminal L9 and RNase Sa systems, changes in electrostatic interactions in the unfolded state that cause an increase in free energy had an overall compaction effect that suggests a decrease in entropy. In the third case, in which we compared the β-lactalbumin and hen egg-white lysozyme protein homologues, we successfully eliminated differences between the folding kinetics of the two systems by suppressing electrostatic interactions, supporting previously reported findings. Our coarse-grained molecular dynamics study not only reproduces experimentally reported findings but also provides a detailed molecular understanding of the elusive unfolded-state ensemble and how charge–charge interactions can modulate the biophysical characteristics of folding.

Keywords: protein folding; energy landscape; charge–charge interaction; electrostatics; nonnative interactions

Edited by M. Levitt

Introduction

The funnel-shaped energy landscape[1–3] of natural proteins suggests that folding is a robust and an efficient process in which the protein sequence is selected to minimize conflicting interactions that can contribute to the accumulation of misfolded conformations. It is therefore often stated that protein folding can be captured by acknowledging only the native interactions as those that govern folding mechanisms. This assumption describes folding as occurring on a perfectly funneled landscape, and its relevance was proven by various studies that use native-topology-based models.[4–13]

While various aspects of the complexity of protein folding are correctly addressed by assuming a perfectly funneled landscape, it is clear that protein sequences are frustrated to some extent. Yet, the exact magnitude of frustration in proteins is rather unclear. In general, frustration can have two

*Corresponding author. E-mail address: koby.levy@weizmann.ac.il.

Abbreviations used: CG-MD, coarse-grained molecular dynamics; NTL9, N-terminal L9; BLA, β-lactalbumin; HEWL, hen egg-white lysozyme; PDB, Protein Data Bank; CSU, Contacts of Structural Units; WHAM, weighted histogram analysis method; LJ, Lennard–Jones; DH, Debye–Hückel.

contributions: topological and energetic.[14] The existence of topological frustration[15–17] (e.g., a knot in the structure) and energetic frustration[18] (e.g., residues that participate in nonnative interactions) does not necessarily indicate poor sequence selection; rather, they can assist folding[19,20] and, more importantly, they may support function.[15,18] Local frustration may, for example, assist in conformational changes and in protein dynamics. Mutations, depending on their chemical nature and location in the sequence, can disturb native contact formation or, alternatively, support the formation of nonnative contacts and, therefore, have the potency to modulate protein folding characteristics. In this study, we focus on the effect of nonnative charge–charge interactions on protein folding.

Charge–charge interactions have been targeted in recent years as a means for enhancing protein stability.[21–28] Following this approach, several proteins that were successfully rationally redesigned to optimize surface charge–charge interactions showed enhanced thermostability[29,30] and, in some cases, retained their biological function.[31] In some cases, the enhanced stability conferred by redesigning the surface charge–charge interactions was reasonably well predicted using computational models based solely on the native-state structure. In other cases, the computational models showed a poor capacity to predict the change in thermostability, even qualitatively.[32] One reason for the limited capability to predict the effect of surface charge–charge interactions on protein stability stems from the fact that these models ignore the unfolded state, assuming that charged residues are well solvated in the unfolded state and interactions between them are screened by the solvent. However, charge–charge interactions in the unfolded state can, in principle, be formed, depending on their position, and their formation may be supported by the residual structure of the unfolded state. When charged residues form favorable interactions in the unfolded-state ensemble, a destabilization effect can be observed. Indeed, in cases where the native state fails to estimate the effect of mutating charge residues on stability, it was found that charge–charge interactions influence the denatured state.[32–34] Recently, it was shown that, to reproduce pairwise distance distributions extracted from single-molecule experiments in denaturing conditions, it is essential to incorporate electrostatic interactions into the model of the unfolded state.[35]

Recent studies of the effect of charge–charge interactions on the unfolded state nicely exemplify how the scientific view of the unfolded state of proteins has evolved, with unfolded proteins no longer viewed as negligible entities but rather acknowledged as players in the folding process. The unfolded state of proteins, which was often described as a random coil, is starting to be appreciated as a state with much richer physics.[36–40] Accordingly, unfolded proteins are now considered to be different from random flexible polymers that have no stabilizing interactions. Therefore, they may have non-zero enthalpy and their entropy may be lower than that of a random coil. It is now understood that, although unfolded proteins follow random-coil statistics, this does not preclude the possibility of the retention of residual structure in the unfolded states as a result of native or nonnative interactions. For example, it has been shown that, where rigid segments (i.e., secondary structure regions) in the unfolded ensemble comprise up to 92% of the protein sequence, self-avoiding walk statistics still prevail.[41] Several experimental studies support residual native-like structural elements in the denatured state.[42–44] Theoretical and computational studies have addressed the role of native interactions in the denatured state and proposed a significant bias toward the native structures.[38,45,46] Nonnative interactions (e.g., nonnative salt bridges, cluster of hydrophobic interactions, or PPII conformations) can be formed in the denatured state as well.[33,40,47,48] It was shown that modulating nonnative electrostatic interactions within the unfolded-state ensemble can significantly change protein stability.[29,32,33,37,49–54] Variations in crowding or confinement conditions[55–57] and attachment of oligosaccharides to proteins[58,59] are reported to have a direct effect on the free energy of the unfolded state. These examples illustrate that understanding the properties of the unfolded state may shed light on various biophysical characteristics of proteins.

In most cases, evidence for the influence of charge–charge interactions in the unfolded-state ensemble on protein stability and kinetics is indirect. For example, the formation of nonnative salt bridges in the unfolded state was supported by observed deviations from expected behavior based only on charge–charge interactions in the native state.[26,32,33,50] Obviously, a direct observation of the interplay between charge–charge interactions is essential for quantifying the frustration introduced by charged residues in protein folding and to take into account these effects when designing proteins with higher thermostability.

In this study, we focus on characterizing the effect of native and nonnative charge–charge interactions on protein folding at the molecular level using coarse-grained molecular dynamics (CG-MD) simulations. We chose to study two protein systems that demonstrate the effect mutating charged residues may have on protein folding. For both the N-terminal L9 (NTL9) protein domain (Fig. 1a) and the RNase Sa protein (Fig. 1b), it was shown experimentally that eliminating or reversing certain charged residues that are extensively exposed to the solvent can significantly change the protein's stability and kinetics. Specifically, some of the mutations in NTL9[33,50,53] and in RNase Sa[29,32,54] principally influence the unfolded-state ensemble, while the folded state remains similar to the wild type. Using a novel, coarse-grained model that includes an electrostatic potential specifically calibrated for this purpose, we reproduced the charge neutralization and charge reversal mutations *in silico* and correlated the thermodynamic and kinetic results from our simulations with experimental data. We also performed a structural analysis that suggests an ex-

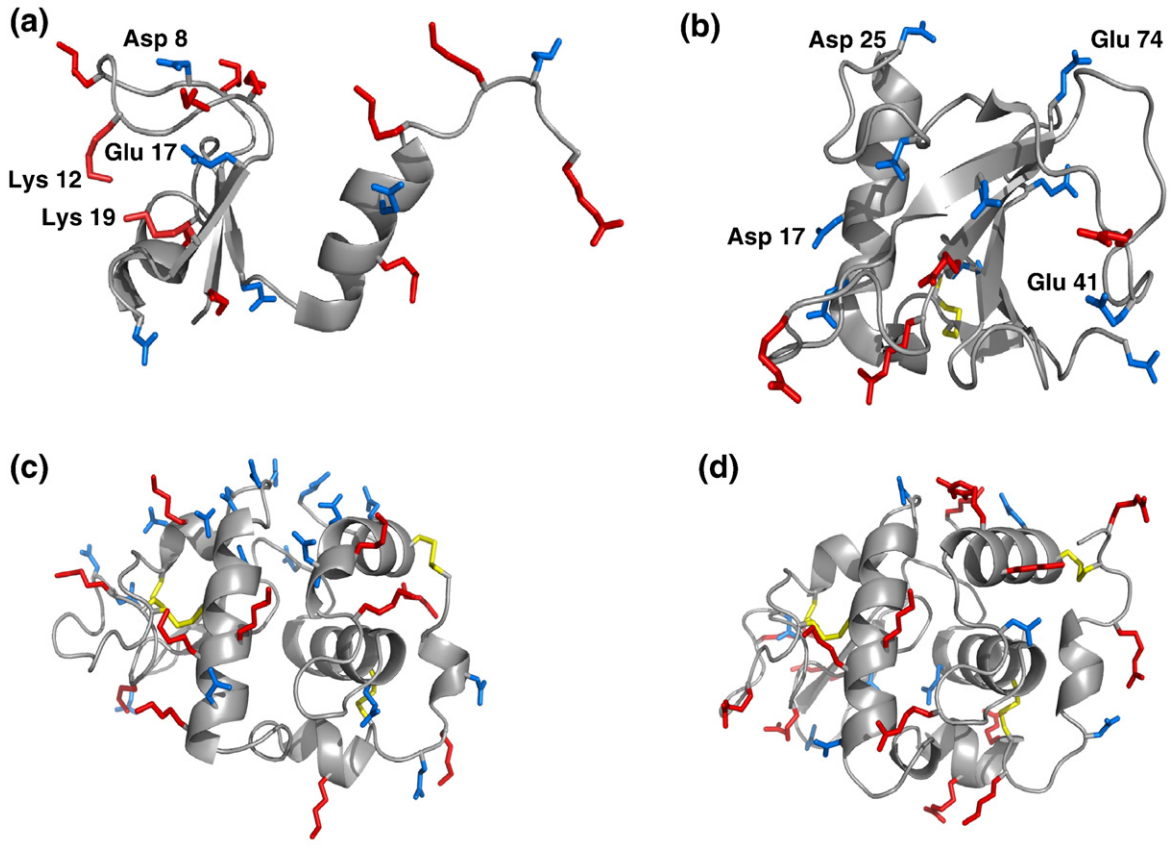

Fig. 1. Cartoon representation of the studied proteins: (a) N-terminal L9 domain, (b) RNase Sa domain, (c) BLA, and (d) HEWL. Positively charged residues are colored red, negatively charged residues are colored blue, and disulfide bonds are colored yellow. For the sake of clarity, only the side chains of the charged residues are shown.

planation for the observed differences due to mutations. The third system used in this study comprised the β-lactalbumin (BLA) and hen egg-white lysozyme (HEWL) proteins (Fig. 1c and d), which are structural homologues but exhibit different folding kinetics. It was recently proposed that the difference in their kinetics is due to each homologue having different charged residues sited in different locations.[60] We simulated both proteins with and without the incorporation of the electrostatic potential to explain the observed experimental difference.

Results and Discussion

Nonnative electrostatic interactions affect folding characteristics

Thermodynamic and kinetic analysis of NTL9 mutants

A recent series of studies by Raleigh *et al.*[28,33,50,53,61–64] focused on NTL9 as a model protein (Fig. 1a) to investigate the effect of electrostatic interactions on protein stability. They performed a set of mutations to remove the electrostatic charge from charged residues by replacing them with similarly shaped uncharged residues. Each mutation causes a relative change in the folding stability with respect to the wild-type domain, with one very pronounced stabilization effect (2 kcal/mol) for the lysine-to-methionine substitution at position 12 (i.e., for K12M L9). These studies present compelling evidence to attribute the change in stability to the denatured-state ensemble. Their results show that, while the folded wild-type and mutated species vary only slightly from each other, the denatured ensembles of some mutations differ significantly in structure from the wild type. Their analysis relies on the fact that exposing a protein to high acidity causes it to denature and its acidic residues to adopt a proton. The Tanford–Wyman equation[65] describes the exact relationship between the thermal stability of a protein and the pH of the solvent for an ideally titrating protein, wherein each charged residue undergoes titration separately. Raleigh *et al.* performed thermal stability analyses and observed that wild-type NTL9 shows a deviation from the calculated ideal titration curve at low pH, which diminishes at higher pH values. They attributed this deviation to salt bridges occurring in the unfolded state. The deviation at low pH is significantly smaller for the K12M L9 mutant, suggesting that fewer salt bridges form in the unfolded-state ensemble of the mutant.

We began by modeling the folding thermodynamics of 11 mutants (K7M, D8N, K10M, K12M, K14M, K15M, E17Q, D23N, E38Q, E48Q, and E54Q) that had been studied experimentally by Raleigh *et al.*[33,50,62–64] Figure 2 compares the experimentally obtained values for the stability change due to mutation with those obtained from our simulations. While results from the simple single-bead model do not correlate with the experimental data (Fig. 2), the two-bead model (in which the charge is placed on the farthest heavy atom of the side chain) shows good agreement with experimental data, particularly in that it shows a high degree of stabilization for the K12M L9 mutant. Examination of ΔH folding and $T\Delta S$ folding of the mutants reveals a striking difference between the K12M mutant and the other mutants in the set. In the folded state, changes in enthalpy and entropy are small and compensate for each other (Fig. 3), yet we may note that K12M L9 shows lower enthalpy and entropy values than the other mutants. In the unfolded state, changes in enthalpy and entropy are greater but still tend to compensate for each other (Fig. 3), resulting in a small ΔG ($G_{\mathrm{mut}}{}^{\mathrm{Unfolded}} - G_{\mathrm{WT}}{}^{\mathrm{Unfolded}}$). The K12M L9 mutant stands out in that it has a similar enthalpy to the wild-type protein but shows a decrease in entropy in the unfolded state. The smaller entropy in the unfolded state of the K12M L9 mutant is the principal contributor to its overall thermodynamic stabilization.

We calculated the folding and unfolding rate of wild-type L9 and of its K12M and E17Q mutants. Unfolding rates differ less significantly than folding rates (Fig. 4), suggesting that the unfolded-state ensemble of the mutants differs from that of the wild type, while the folded state is similar, in accordance with the conclusion of the thermodynamic analysis. The experimental values[50,62–64] of the folding/unfolding rate of K12M L9 mutant qualitatively agrees with the simulations rates (quantitative agreement between simulation and experiments is lacking most likely due to the poor folding cooperativity of our coarse-grained model). The relatively low entropy of K12M L9 in the unfolded state indicates that it is more structured and, thus, folds more easily.

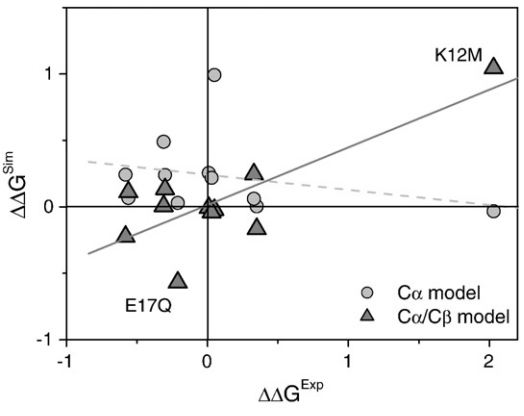

Fig. 2. Correlation between the simulated and experimental values for the stability change due to mutation ($\Delta\Delta G$). The $\Delta\Delta G$ values obtained from the C^{α} and C^{α}/C^{β} models are shown with circles and triangles, respectively. The single-bead model was simulated using a dielectric constant of 50 and a salt concentration 0.1 M, and the calibrated two-bead model was simulated using a dielectric constant of 15 and a salt concentration 0.1 M. The lines were added to guide the eye.

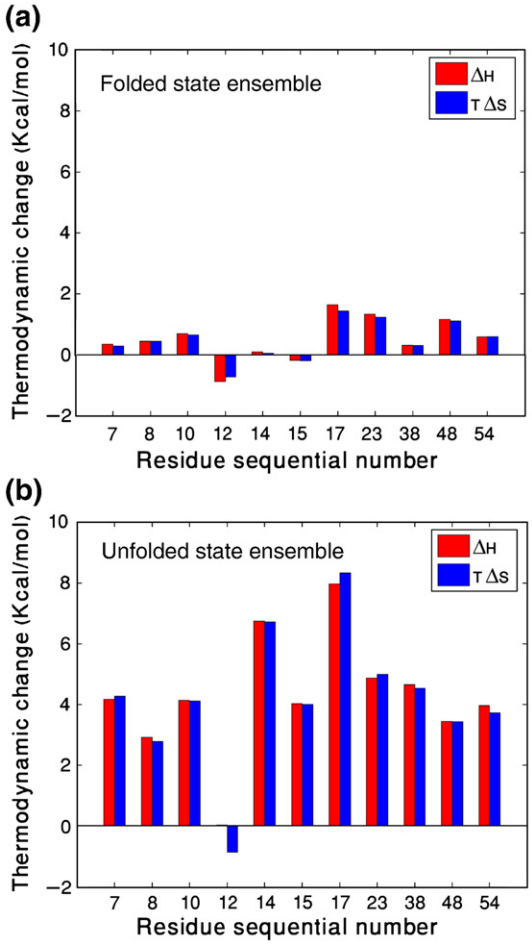

Fig. 3. Thermodynamic analysis of experimentally studied mutants of NTL9, in which the residue index number refers to that which was neutralized through mutation. Thus, residue index 12 relates to the K12M mutant and residue index 17 relates to the E17Q mutant. Entropy ($T\Delta S$) and enthalpy (ΔH) change for the folded state (a) and for the unfolded state (b) with respect to the wild type are shown based on the simulations of the two-bead model.

Structural analysis of the folded and unfolded-state ensembles of NTL9 and its mutants

The detailed trajectories created by our simulations allow us to take a more in-depth look of the effect of mutations on conformational space. We examined the simulation trajectories of wild-type L9 and its K12M (most stabilized) and E17Q (most destabilized) mutants. Trajectories were selected for structural analysis at T_F, where the folded and

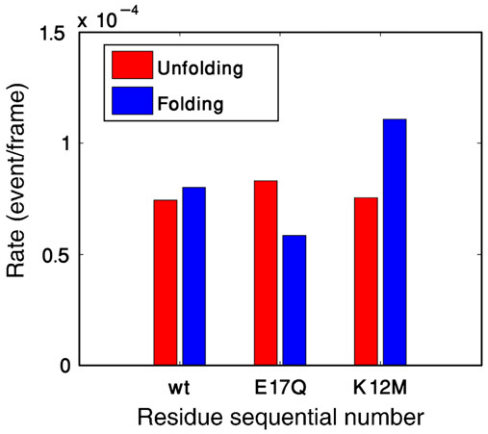

Fig. 4. Kinetic analysis of wild-type and two reported mutants of NTL9, namely, E17Q (in which Glu17 is neutralized) and K12M (in which Lys12 is neutralized), at the T_F of the wild-type NTL9. Folding and unfolding rates are shown in events/frame units.

unfolded states are almost equally populated. We calculated an average distance between the side chains of each residue pair, $\langle R_{ij} \rangle$ (for glycine residues, the backbone bead was used), in both the folded- and unfolded-state ensembles. The differences in the average pairwise distances between each mutant and wild type (termed Δ distance matrices) are shown in Fig. 5. The Δ distance matrices of the folded state for both mutants are similar and close to wild-type values (Fig. 5), while the distances between residues in the unfolded state vary greatly from the wild type (Fig. 5). Looking at the Δ distance matrix of E17Q L9 in the unfolded state (Fig. 5), we see that a region in the loop (residues 1–15) becomes significantly removed from the other residues of the protein. The elimination of the electrostatic charge at position 17 causes the N-terminal part of the domain to disassociate from the rest of it. We also notice distancing of residues 13–17 from residues 1–8, which implies an unwinding of the N-terminal loop structure. In contrast, the Δ distance matrix of unfolded K12M L9 (Fig. 5) shows a trend of overall compaction. Loop region residues are closer to the C-terminal residues, and loop edge residues are about the same distance as in the wild type.

In the K12M L9 mutant, lysine 19 is close to the hinge part of the loop (residues 10–12) in both the folded and unfolded states (Fig. 5). By contrast, lysine 19 is distant from the same region of the loop in the E17Q L9 mutant. Due to this exceptional finding, we decided to study the K19M L9 mutant, which was not reported by Raleigh *et al.* Interestingly, the relative stability of K19M L9 is predicted by our model to be 75% greater than that of K12M L9 (simulation $\Delta\Delta G$ values are 1.82 and 1.04 units, respectively). Analysis of the enthalpy and entropy of both the folded and unfolded states assigns this stabilization effect to a great decrease in the entropy of the unfolded state. The unfolding rate of K19M L9 is similar to that of the other mutants, but its folding rate is triple that of the wild type. The Δ distance matrix of the K19M L9 mutant (Fig. 5) shows a similar pattern to that of K12M L9 (the loop remains closer to the rest of the protein) but with a greater compaction effect.

In order to examine the effect of mutations on a finer level, we calculated histograms of distances between specific residue pairs. Histograms were calculated using the same trajectories as were used in the Δ distance matrix analyses and were similarly separated into folded and unfolded frames. This analysis was performed for wild-type L9 and the mutants K12M, E17Q, and K19M.

The histogram of distances between residues 12 and 17 in the unfolded state is presented in Fig. 6a, while the inset shows distances in the folded state. In the folded state, the average distance between residues 12 and 17 is ~7.5 Å, and no significant difference is observed between the mutants. In the unfolded state, we observe a bimodal histogram with two characteristic distances: one peak around the average folded-state distance and another peak at ~19 Å that corresponds to conformations in which the loop is unstructured. K12M L9 shows a similar distribution to the wild type. In the E17Q mutant, the separation between residues 12 and 17 is increased, while in the K19M mutant, it is reduced. Thus, neutralizing Lys19 removes the unfavorable interaction between residues 12 and 19 (see Fig. 1a) and allows the loop region (and with it, residues 12 and 17) to remain structured in K19M L9. Neutralizing Glu17 increases the repulsion between residues 12 and 19 in E17Q L9, thus causing the loop region to become unstructured. Neutralizing Lys12 in the K12M L9 mutant causes a combined effect, removing, on the one hand, the attraction to Glu17 and, on the other hand, the repulsion from Lys19.

Studying the histogram of distances between residues 12 and 19 (Fig. 6b) shows the overall results to be similar to those for the distances between residues 12 and 17. The folded-state distances for the mutants are similar (Fig. 6b, inset), with the exception of K19M L9, which has a lower typical distance in the folded state. The unfolded-state histogram also shows a bimodal distribution, with wider and shorter peaks. Both the K12M and K19M mutants draw residues 12 and 19 closer together than occurs in the wild type, while E17Q pushes them apart. The first peak of the structured state distance histogram (~4 Å) is not symmetric; such small distances cannot be populated as the repulsion potential becomes significant.

The histogram of distances between residues 8 and 12 (Fig. 6c) hints at the existence of a nonnative salt bridge. Folded-state distances for all mutants are similar (Fig. 6c, inset). The unfolded-state distances histogram shows only one peak at the typical distance of the folded state. The peak is not symmetrical, and a substantial fraction of frames occupy shorter distances, suggesting a nonnative salt bridge between residues 8 and 12, while the loop is

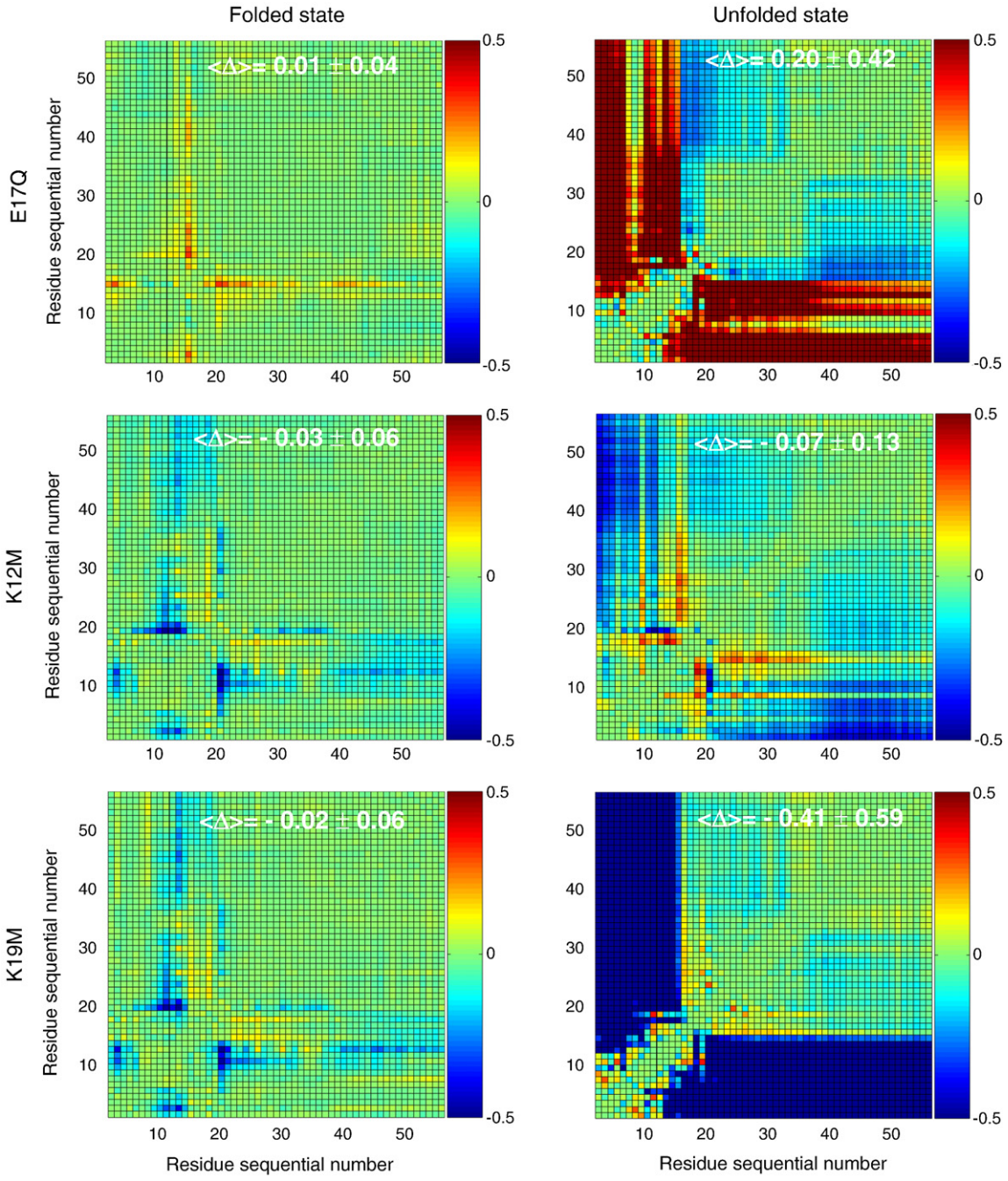

Fig. 5. Δ distance matrices of NTL9 mutants E17Q (top panels), K12M (middle panels), and K19M (bottom panels); distances are in angstroms. The Δ distance matrices are calculated relative to the wild-type NTL9 (i.e., $\langle \Delta_{ij} \rangle = \langle R_{ij}^{\mathrm{mut}} \rangle - \langle R_{ij}^{\mathrm{WT}} \rangle$, where R_{ij} is the distance between residues i and j). The panels on the left correspond to the ensemble of the folded state, and the panels on the right correspond to the ensemble of the unfolded state. For easier comparison, all differences in distances are shown in the range of −0.5 to +0.5, although some distance differences are much larger. The values at the top of each panel indicate the average distance (in angstroms) between the side chains of each residue pair as well as its standard deviation.

unstructured. Since E17Q causes the loop region to be less structured, it frees residues 8 and 12 to form an interaction, resulting in more conformations with short distances between these residues, relative to the wild type. The loop region of mutants K12M L9 and K19M L9 remains structured; hence, both mutants populate fewer short distance conformations than the wild type.

The effect of ion shielding on the stability of RNase Sa mutants

The effect of charge reversal mutations on the thermal stability of the RNase Sa protein (Fig. 1b) was studied by Pace *et al.*[32,66] This study explains an apparent contradiction between expected and observed stability changes due to the charge reversal of

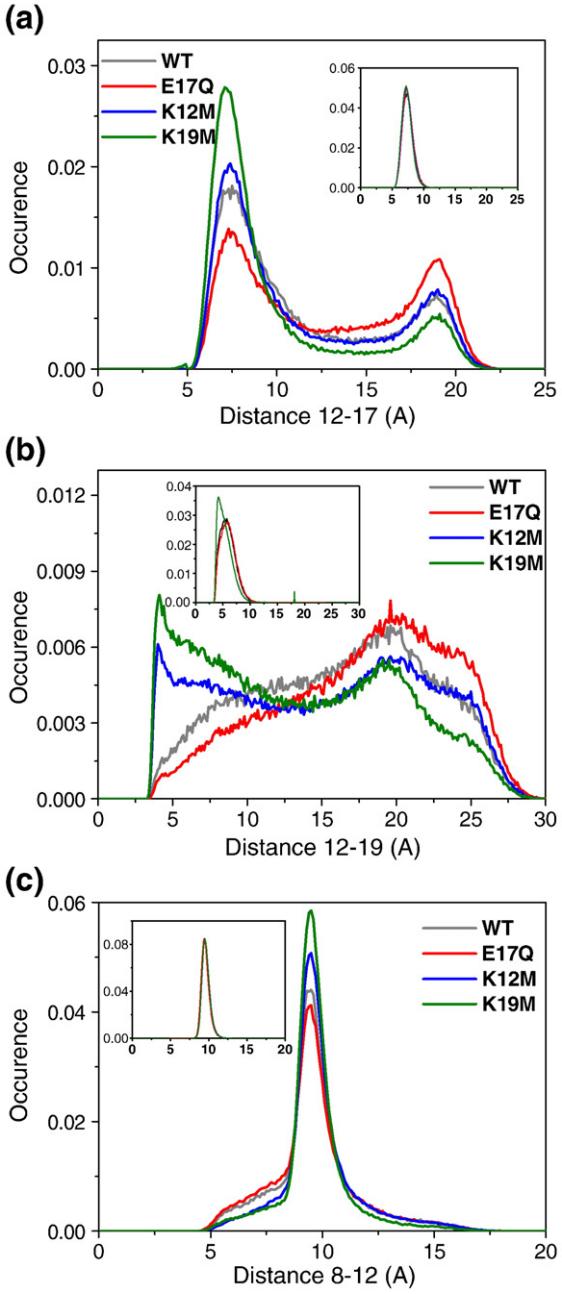

Fig. 6. Distance distributions of selected residue pairs in wild-type NT L9 and the mutants E17Q, K12M, and K19M. The distance distributions are calculated for the ensemble of unfolded conformations for residue pairs 12–17 (a), 12–19 (b), and 8–12 (c). Insets indicate the distance distribution in the folded state.

selected charged residues. These charge reversal mutations can, on the one hand, create intentionally designed salt bridges on the surface of the protein, resulting in higher thermal stability than the wild type due to a decrease in enthalpy of the folded state (the authors originally were motivated by increasing protein stability by improving electrostatic interactions among charged groups on the surface of the folded protein). On the other hand, some of the mutations can create new electrostatic interactions that are favorable in the folded state but are even more favorable in the unfolded state. The new potential salt bridges in the unfolded state may result in more compact conformations (i.e., lower entropy) and may therefore cause an overall stabilization of the protein. The study by Pace *et al.* also maintains that electrostatic interactions in the unfolded state are more exposed to water and are therefore more susceptible to ion screening than are the corresponding interactions in the folded state. From this line of thought, it follows that increasing the salt concentration will increase thermal stability more for destabilizing mutations than for stabilizing mutations.

We studied the folding thermodynamics of the four mutants (D17K Sa, D25K Sa, E41K Sa, and E74K Sa) that were studied experimentally by Pace *et al.*[32] and found a good correlation between the experimental and simulation data on stability changes due to mutation (Fig. 7a), with the exception that we found a destabilization effect for the E41K mutation, while a stabilization effect was reported experimentally. We speculate that the lack of success of our

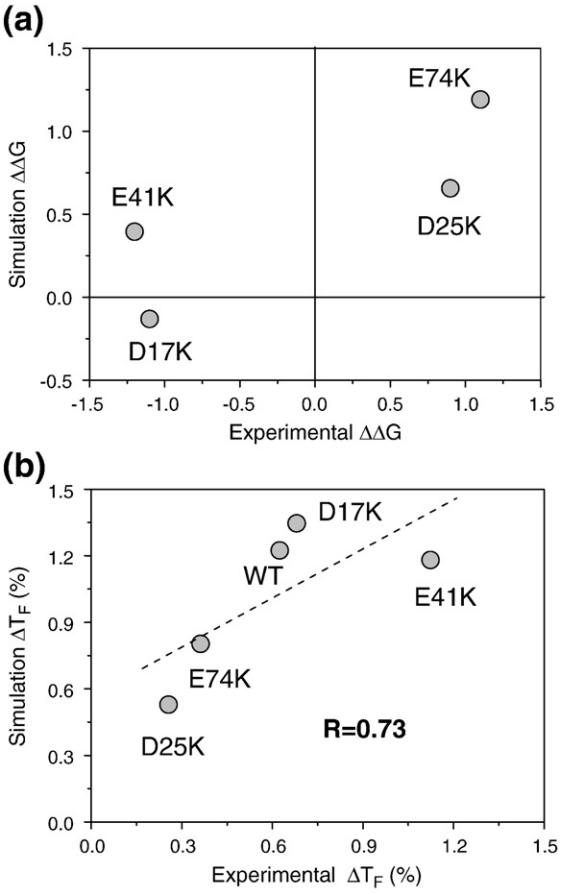

Fig. 7. Thermodynamic analysis of wild-type RNase Sa and reported mutants. (a) Correlation of the effect of mutations on protein stability ($\Delta\Delta G$) between simulation and experimental results. (b) Correlation of dependence of T_F on salt concentration (ΔT_F^{salt}) for wild type and reported mutants. ΔT_F^{salt} is defined as $T_F([salt]=0.1\ M) - T_F([salt]=0\ M)]/T_F([salt]=0\ M)$.

model in predicting the stability of mutant E41K is due to the fact that this point mutation is located at highly flexible loop (high *B*-factor), which limits the validity of the native-topology-based model in estimating the thermodynamic parameters. Analysis of the changes in enthalpy and entropy of the four mutants reveals an effect on both the folded and unfolded states, and a kinetic analysis shows a change in both the folding and unfolding rate constants due to the mutations (data not shown). Significantly, the change in T_F due to a change in salt concentration is highly correlated between the experimental and simulation data for our successfully predicted mutants (Fig. 7b), implying that our model can successfully simulate different ion concentration conditions. As observed by Pace *et al.*,[32] the destabilizing mutations are stabilized to a greater extent than are the stabilizing mutations when the salt concentration is raised, suggesting that the destabilization is due to new electrostatic interactions, which are more susceptible to ion shielding, occurring in the unfolded state.

We have performed a structural analysis for two of the mutants (the stabilized E74K and the destabilized D17K) that shows a negligible change in the folded state for both mutants and a significant change in the unfolded state (Fig. 8). It is interesting to note that the two mutations seem to have an opposite effect on the overall pattern of distances between residues and that the unfolded state of stabilizing E74K shows an overall compaction ($\langle\Delta\rangle = -0.19 \pm 0.68$), while the unfolded state of destabilizing D17K shows an overall expansion ($\langle\Delta\rangle = 0.04 \pm 0.62$).

Modulating the folding kinetics of the BLA and HEWL homologues

Halskau *et al.* have studied the BLA (Fig. 1c) and HEWL (Fig. 1d) homologues and have discovered that these similarly structured proteins have different folding kinetics.[60] These proteins are similar in size, and their contact map suggests that they have similar folding topologies. Both these properties are known to significantly determine folding

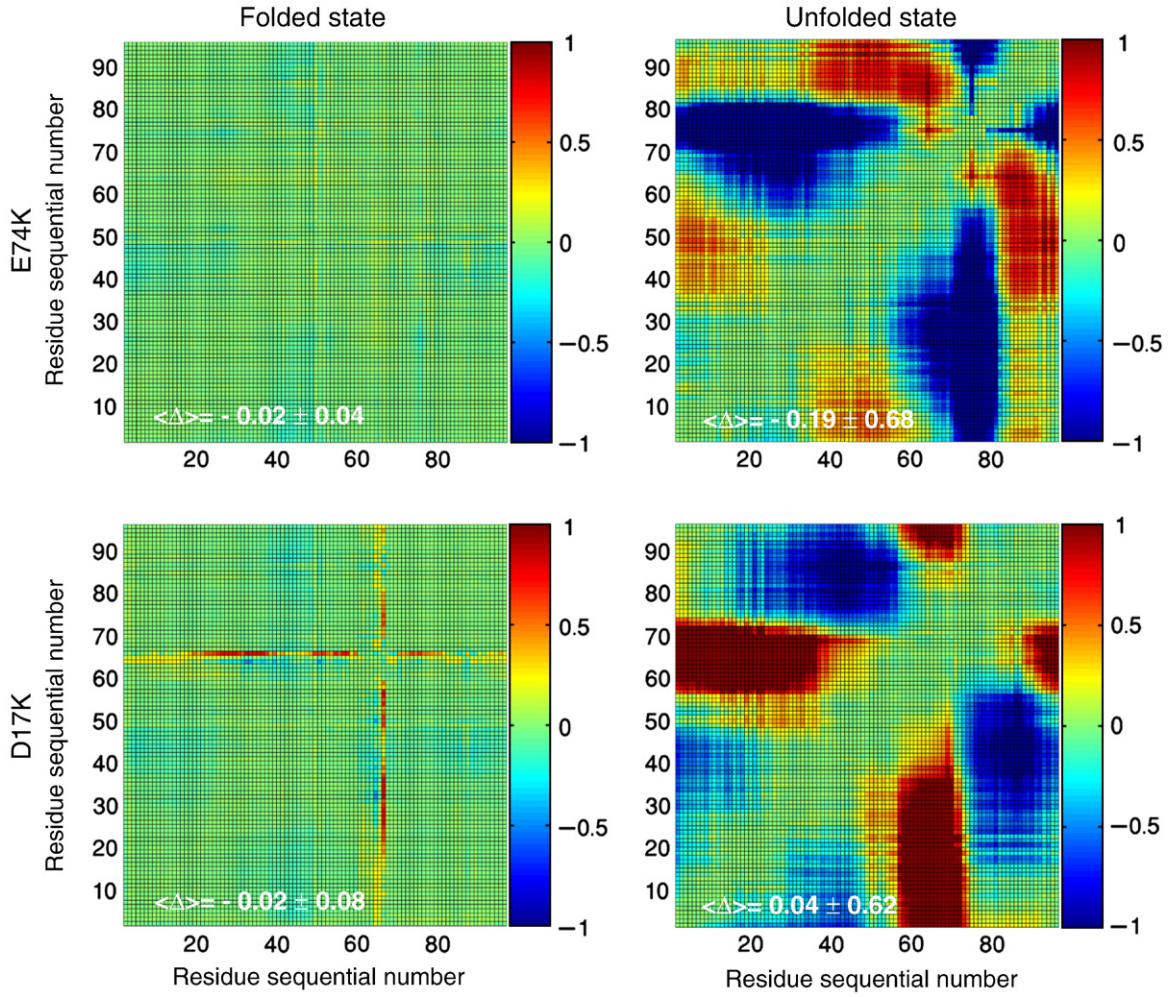

Fig. 8. Δ distance matrices of RNase mutants E74K (top panels) and D17K (bottom panels) relative to wild-type RNase Sa; distances are in angstroms. Panels on the left correspond to the folded-state ensemble, and panels on the right correspond to the unfolded-state ensemble. For easier comparison, all differences in distances are shown in the range of −1.0 to +1.0, although some distance differences are much larger. The value at the bottom of each panel indicates the average distance (in angstroms) between the side chains of each residue pair as well as its standard deviation.

rates;[4,67–69] thus, the *a priori* expectation is that they will fold at a similar rate. However, their surprising results show that the BLA protein folds faster and has a lower T_F than HEWL (1.34 kcal/mol and 318 K for BLA and 8.84 kcal/mol and 339 K for HEWL for the folding barrier heights and T_F, respectively). Their work suggests that the main cause of this effect is electrostatic interactions. Halskau *et al.* performed several analyses to corroborate their claim.[60] They measured the chemical shifts of acidic residues as a function of pH and associated them with the acceptance of a proton by the carboxylic side chain, and they indirectly measured pK_a values for those residues. They proceeded to compare experimental pK_a values of both BLA and HEWL to theoretically predicted pK_a values using the Tanford–Kirkwood approximation;[70] this analysis showed that the pK_a values of HEWL residues are highly correlated with the expected pK_a values, while the pK_a values of BLA residues are not. In addition, they applied a statistical mechanics model that generated 10^7 partially folded conformations. Using this conformational space, they evaluated the dependence of electrostatic free energy on the fraction of native contacts for both proteins, showing that stabilizing electrostatic interactions occur in the unfolded state of BLA, while no such interactions appear in the unfolded state of HEWL.

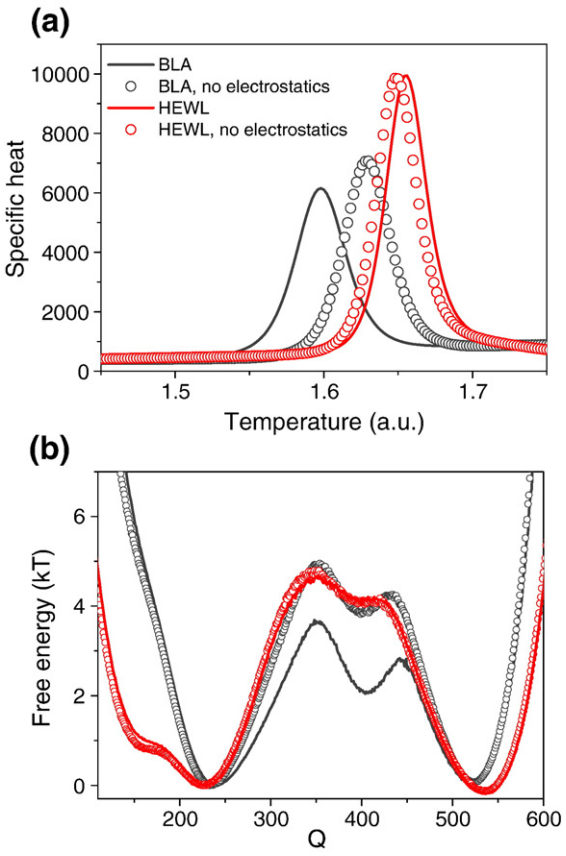

Fig. 9. Thermodynamic analysis of the BLA and HEWL proteins. (a) The specific heat curves; (b) the potential of mean force profiles. Open circles correspond to simulations with the electrostatic potential neutralized.

Using our new model, we were able to reproduce the experimental specific heat curve (Fig. 9). The specific heat curve of HEWL peaks at a higher temperature and is sharper than that of BLA, indicating a more cooperative folding process. Neutralizing electrostatic interactions caused the T_F of both proteins to become similar, indicating that they acquired a similar stability. Analyzing the potential of mean force plot (Fig. 9b) shows the effect of electrostatics on the folding barrier. Remarkably, the transition state of a BLA protein in which electrostatic interactions have been neutralized is similar to that of HEWL. The presence of an intermediate is consistent with the BLA protein-specific heat peak being smaller and wider and with the experimental results. We note that the stability and folding kinetics of BLA and HEWL are similar but not identical when the electrostatic interactions are excluded (i.e., their folding is studied using the simple native-topology-based model) since their topology is the same but their network of tertiary contacts includes some differences.

Figure 10 shows the Δ distance matrices for BLA and HEWL in their unfolded- and folded-state ensembles. A conformation was considered to be in the transition state if the number of native contacts for that conformation was between 290 and 450 (see Fig. 9b). Neutralizing electrostatic interactions has opposite effects on each protein model. The lack of an electrostatic potential causes several residue pairs of the BLA system to draw nearer to each other, while similarly positioned residue pairs draw apart in the HEWL system. For both systems, significant changes in the folded-state and unfolded-state ensembles are observed, with stronger effects witnessed in the HEWL system. The compaction effect observed in the BLA system when all electrostatic interactions were turned off suggests that the long-range electrostatic interactions network of the BLA protein is relatively weak and its interactions are in conflict with the native topology.[60] Thus, the removal of these interactions has a net stabilizing effect on the BLA system (Fig. 9a). Conversely, the long-range electrostatic interaction network of HEWL is in concert with the native structure.[60] The removal of these interactions is equivalent to removing some of its native contacts and has a net destabilizing effect on the HEWL system (turning off the electrostatic interactions resulted in a more expanded folded state) (Fig. 9a). In both systems, the effect seems to be dependent on changes in enthalpy. Changes in entropy, which contribute oppositely to stability, are probably less pronounced, due to the four disulfide bonds in the fold imposing strong constraints on the systems that greatly shrink the conformational space.

Conclusions

Three protein systems were investigated using CG-MD simulations to understand the effect of charge–charge interactions on the biophysics of the proteins and particularly on the unfolded state.

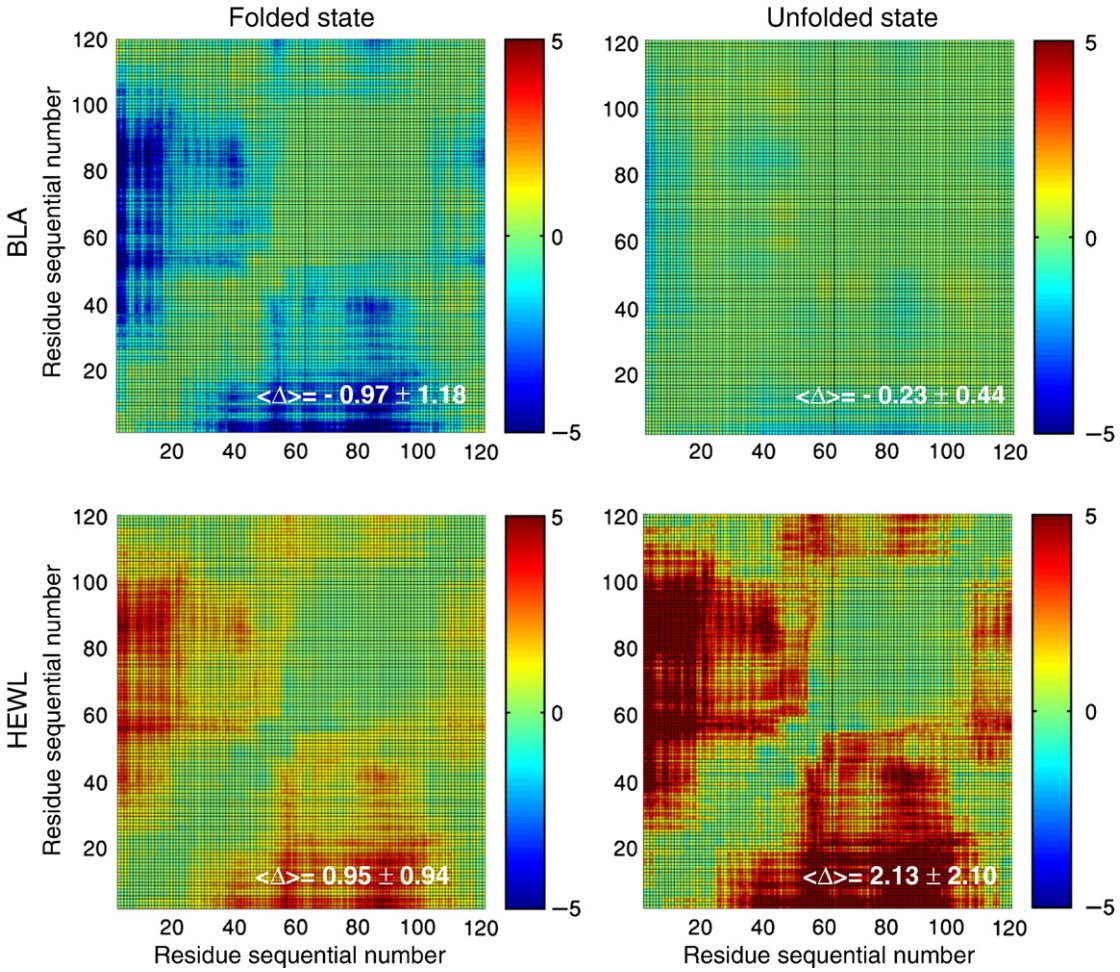

Fig. 10. Δ distance matrices for the BLA (top panels) and HEWL (bottom panels) proteins ($\langle \Delta_{ij} \rangle = \langle R_{ij}^{\text{No-electrostatic}} \rangle - \langle R_{ij}^{\text{WT}} \rangle$); distances are in angstroms. Panels on the left correspond to the folded-state ensemble, and panels on the right correspond to the unfolded-state ensemble.

Results from the L9 system suggest that the removal of the charge from lysine 12 (K12M mutant) causes a decrease in enthalpy due to the elimination of the repulsive electrostatic interaction. The decrease in enthalpy is present to only a negligible extent in the unfolded state. In addition, we can note that neutralizing the charge guides lysine 19 into closer contact with lysine 12 and the residues surrounding it, located in the middle of the loop region of NTL9. The entire region encompassing residues 12 and 19 remains more structured in the unfolded conformations of K12M L9 than it does in the wild type. This stabilization of residual structure reduces the entropy of the unfolded ensemble of the K12M mutant in a manner exceeding the enthalpy contribution alone, resulting in the destabilization of the unfolded ensemble and the thermal stabilization of the NTL9 protein domain. This finding is consistent with the results of Raleigh *et al.*, who found the K12M mutation to cause a significant stabilization effect ($\Delta\Delta G \sim 2$ kcal/mol) and attributed it to destabilization of the unfolded state.[33,53] Their analysis suggests that the destabilization of the unfolded state is due to the non-formation of a nonnative salt bridge between residues 8 and 12 that is formed (at least transiently) in wild-type NTL9. We find evidence for such interactions in our simulations examining the distance profile between residues 8 and 12. In the wild-type species, the greater loop region unfolds more frequently, allowing transient salt bridges to form. Neutralizing the charge of lysine 12 locks the region in its folded configuration and fewer transient salt bridges can form. The E17Q mutation intensifies the unfavorable interaction between residues 12 and 19, causing the loop region to unfold and allowing transient salt bridges to form more frequently.

Electrostatic interactions occurring in the unfolded state are more susceptible to ion shielding. Both native and nonnative electrostatic interactions are likely to be more exposed to water and salt ions when the protein is unfolded. Thus, the presence of water increases the likelihood that ions will coordinate around the side chain and mask its charge. In the folded state, native electrostatic interactions are forced into existence by the overall fold of the protein and are therefore less dependent on ion concentration. Our findings directly corroborate this

understanding. Using a model of the RNase Sa wild type and several charge reversal mutants, we measured the change in thermal stability at different ion concentrations. Similarly to the experimental results reported by Pace *et al.*,[32] changing the salt concentration from 0 to 0.1 M increased the thermal stability to a greater extent for the destabilizing mutations of RNase Sa than for its stabilizing mutations. This is consistent with thermal destabilization being caused by overall favorable electrostatic interactions (e.g., formation of transient salt bridges) that principally occur in the unfolded state and that a high salt concentration eliminates. Stabilized mutants gain new favorable electrostatic interactions in their folded state (in accordance with the applied design strategy of improving salt bridges on the protein surface) and, as expected, gain less additional thermal stabilization when the salt concentration is increased.

The BLA and HEWL homologues are very similar in size and topology and are expected to have similar folding rates. The observed difference in their folding kinetics is attributed to the different distribution of charged residues along their sequence and structure, demonstrating that electrostatic interactions can significantly modulate protein folding kinetics, which is mostly governed by the native topology. We sharpened this understanding, which was recently illustrated in the experimental work of Halskau *et al.*,[60] by neutralizing the electrostatic potential in our simulation. Eliminating the electrostatic charge caused BLA and HEWL to have similar thermal stabilities and similar folding barriers.

The Δ distance matrix gives a structural insight into the effect of mutation on the state of the protein. In all three systems, compaction effects, which can be correlated with lower entropy, caused an increase in the free energy for that state. On the other hand, expansion effects, which are correlated with higher entropy, caused a decrease in free energy.

We set out to try to attribute a stabilizing effect to unfolded-state interactions, but we stumbled upon a more complex effect that includes both folded and unfolded species. Our finding strengthens the need to consider the unfolded ensemble not as a random coil, sampling all conformations equally, but as having a residual structure. This result has implications for understanding protein folding as well as protein design. When planning a mutation that will influence the folded state favorably (charge neutralization in the case of L9 and charge reversal in the case of RNase Sa), one may also be altering the tendency of the local structure to remain intact when the protein undergoes unfolding. Both folded and unfolded species may form the new favorable interaction; thus, total stabilization may depend more on the relative loss of entropy in the unfolded state than on the gain of enthalpy in the folded state. In such a case, the exact outcome of the mutation is hard to predict, since the change in entropy may not be easily gauged. We speculate that placing a mutation designed to add favorable interactions in the folded state in marginally stable areas of the protein will improve the prediction of stabilization effects. In choosing such a location, our aim is that the segment around the site chosen for mutation will break up during unfolding, but even if the residual native structure is not present after the protein unfolds, there are still alternative encounter possibilities for the mutated residue in the unfolded ensemble.

Methods

Simulation model

To study the effect of electrostatic interactions on the biophysical characteristics of proteins, we developed a coarse-grained model in which each amino acid is represented by two beads. One bead, representing the backbone, was located on the C^{α} atom, and one bead, representing the side chain, was located on the farthest heavy atom of the side chain. We ran CG-MD simulations using the Langevin equation to sample the conformational space of our protein models in both the folded and unfolded states. The force field applied in our simulations uses a native-topology-based potential. The potential in this model rewards conformations that resemble the native fold. The potential of a particular conformation [$V(\Gamma, \Gamma_0)$, where Γ denotes a particular conformation and Γ_0 denotes the native conformation] along the CG-MD simulation trajectory consists of the following terms:

$$
\begin{aligned}
V(\Gamma,\Gamma_0) = & \sum_{\text{bonds}} K_{\text{bonds}}\left(b_{ij} - b^0{}_{ij}\right)^2 + \sum_{\text{angles}} K_{\text{angles}}\left(\theta_{ijk} - \theta^0{}_{ijk}\right)^2 \\
& + \sum_{\text{dihedrals}} K_{\text{dihedrals}}[1 - \cos(\varphi_{ijkl} - \varphi^0{}_{ijkl}) \\
& - \cos(3\varphi_{ijkl} - \varphi^0{}_{ijkl})] \\
& + \sum_{i\neq j} K_{\text{contacts}}\left[5\left(\frac{A_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{A_{ij}}{r_{ij}}\right)^{10}\right] \\
& + \sum_{i\neq j} K_{\text{repulsion}}\left(\frac{0.7C_{ij}}{r_{ij}}\right)^{12} + V_{\text{electrostatics}} + V_{\text{chirality}}
\end{aligned}
$$

where $K_{\text{bonds}} = 100$ kcal mol^{-1} Å^{-2}, $K_{\text{angles}} = 20$ kcal mol^{-1}, and $K_{\text{dihedrals}}$, K_{contacts}, and $K_{\text{repulsion}}$ are each valued at 1 kcal mol^{-1}. b_{ij} is the distance (in angstroms) between bonded beads i–j, and b^0_{ij} is the optimal distance (in angstroms) between bonded beads i–j. φ_{ijk} is the angle (in radians) between sequentially bonded beads i–j–k, and θ^0_{ijk} is the optimal angle between subsequently bonded beads i–j–k. φ^0_{ijkl} is the dihedral angle (in radians) between subsequently bonded backbone beads i–j–k–l, and φ^0_{ijkl} is the optimal dihedral angle between subsequently bonded backbone beads i–j–k–l. A_{ij} is the optimal distance (in angstroms) between beads i and j that are in contact with each other, and r_{ij} is the distance (in angstroms) between beads i and j in a given conformation along the trajectory. Optimal values were calculated from the atomic coordinates of the relevant Protein Data Bank (PDB) structure file. C_{ij} is the sum of radii for any two beads not forming a native contact; the repulsion radius of the backbone bead was 1.9 Å, and the radius (in angstroms) of the side-chain bead was set as half the distance between the backbone and side-chain beads for that residue taken from the PDB. Repulsion terms were rescaled by a factor of 0.7 to prevent strong steric clashes that hinder folding.[71] The repulsion potential between immediately adjacent side-chain beads was removed, and the chirality term prevented them from

overlapping with other beads. To test the overlap of repulsion spheres, we considered an imaginary ellipsoid enveloping both backbone and side chain, and violation of its space by other beads was recorded for a sample trajectory. No notable violation of any ellipsoid was found.

The electrostatic potential was represented by the Debye–Hückel (DH) model, which is a linearization of the Poisson–Boltzmann equation.[72] All charged residues (i.e., lysine and arginine are positively charged; glutamic acid and aspartic acid are negatively charged) were given a full charge according to their electrostatic charge at neutral pH. The charge was placed on the side-chain bead. The explicit potential term was:

$$V_{\text{electrostatics}} = \sum_{i,j} K_{\text{electrostatics}} B(\kappa) \frac{q_i q_j \exp^{-\kappa r}}{\varepsilon_r r_{ij}}$$

where $K_{\text{electrostatics}} = 332$ kcal Å mol^{-1} e^{-2}, q_i is the sign of the charge for charged beads i, ε_r is the dielectric constant, κ is the screening factor (in Å^{-1}), $B(\kappa)$ is the salt-dependent coefficient, and r_{ij} is the distance (in angstroms) between charged beads i and j. For the calculation of the screening factor κ and the salt-dependent coefficient $B(\kappa)$, solvent density was taken as 1 kg l^{-1} and the ion radius was taken as 1.4 Å (i.e., the average radius of the Na^+ and Cl^- ions). More details of the DH model can be found in Ref. 73.

In order to maintain the correct chirality of the side-chain beads, we applied a chirality potential.[71] The chirality potential was based on the triple product of three vectors calculated from coordinate differences between the following bead pairs: (1) current backbone bead–previous backbone bead, (2) current backbone bead–next backbone bead, and 3) current backbone bead–current side-chain bead. This potential maintains the relative orientation of the side-chain bead with respect to the plane on which the backbone beads are located.

The explicit potential term was:

$$V_{\text{chirality}} = \sum_{\text{chirality}} K_{\text{chirality}} \left(c_{ijk} - c^0_{ijk} \right)^2$$

$$c = r_{C_\alpha j \to C_\beta j} \cdot \left(r_{C_\alpha j \to C_\alpha i} \times r_{C_\alpha j \to C_\alpha k} \right)$$

$$c_0 = r_{0C_\alpha j \to C_\beta j} \cdot \left(r_{0C_\alpha j \to C_\alpha i} \times r_{0C_\alpha j \to C_\alpha k} \right)$$

where $K_{\text{chirality}} = 20$ kcal mol^{-1} and c_{ijk} is the value of the triple product for sequential residues i,j,k (c^0_{ijk} is the optimal value of the triple product).

Definition of native contacts

Contacts in the native state were determined using Contacts of Structural Units (CSU) software.[74] CSU acknowledges a contact between two atoms based on the intersection of their putative surfaces and their hydrophobicity. Atoms were grouped as either backbone or side-chain atoms, and contacts were acknowledged using the full atom contact list provided by CSU and according to the respective group of the atom. Optimal distances for contacts were calculated between the positions of respective beads. All contact potential coefficients were initially set to 1 kcal/mol, but the contact coefficient was corrected for bead pairs with an attractive electrostatic potential. Backbone–backbone, side-chain–backbone and side-chain–side-chain bead contacts were considered as contact candidates if their residue indices differed by 4 or more, 3 or more, or 2 or more, respectively. Thus, all bead pairs not participating in a bonded term were considered contact candidates, which is an implicit criterion for contact consideration in the single-bead model.

Simulation analyses

The thermodynamic properties of the protein systems were calculated using the weighted histogram analysis method (WHAM).[75] We used the WHAM analysis to compare between the enthalpy and entropy of the folded and unfolded ensemble or the overall stability of wild-type and mutant proteins. The potential of mean force was calculated as a function of the number of native contacts (Q). Conformations were divided into either folded or unfolded. A conformation was acknowledged as folded if its number of native contacts was higher than the number of contacts at the highest point of the transition barrier and as unfolded if otherwise. If an intermediate exists, the unfolded state was defined as having contacts smaller than those that define the first transition state. The folded state accordingly is defined as having more native contacts than those that define the second transition state. Folding and unfolding rates were calculated as the inverse of the average passage times for simulation trajectories at the folding temperature (T_F). The folding temperature is defined as the simulation temperature at which the ratio of folded to unfolded conformations was the closest to 1, which also correlated with the difference in free-energy values between the folded and unfolded states being 0. Unfolding passage time was defined as the number of steps it takes the protein model to move from the folded state to the unfolded state, and the folding passage time was defined vice versa. This method was used in previous studies[76,77] and was shown to correlate with the height of the respective transition barrier as determined by thermodynamic analysis and with experimental results. Entropy (S) values were calculated by subtracting the enthalpy (H) value from the free-energy (G) values. The free energy of the folded or unfolded states was calculated by

$$G_X = -k_B T \log \left[\sum_{Q_a}^{Q_b} \sum_i n(Q, E_i) \exp(-E_i / k_B T) \right]$$

$$H_X = \sum_{Q_a}^{Q_b} \sum_i E_i n(Q, E_i) \exp(-E_i / k_B T)$$

$$TS_X = H_X - G_X$$

where X stands for the folded or unfolded ensembles that are defined by the range of Q $\{Q_a, Q_b\}$. $n(Q,E_i)$ is the density of states for conformations with Q native contacts and an energy, E_i, which is obtained from the WHAM analysis.

Model calibration

The parameters of the designed model required calibration relative to the equivalent parameters of the native interactions so that mutating charged residues would modulate biophysical characteristics in the most reliable and accurate way. We used the change in the Gibbs free energy ($\Delta\Delta G = \Delta G_{\text{mutant}} - \Delta G_{\text{wild type}}$, which represents the change in protein stability due to mutation) as a fitness score for our model by correlating the $\Delta\Delta G$ values yielded by our model with those obtained experimentally. We chose L9 due to the large number of $\Delta\Delta G$

values available for comparison. Our results are based on the calibrated model, whose simulation data are in good agreement with the experimental data.

We first attempted to simply introduce the electrostatic potential into a single-bead model, placing a charge on the C^{α} bead of the charged residues. The details of the single-bead model force field can be found elsewhere.[4] When we used the C^{α} representation, we observed that supplementing the native-based model with electrostatic interactions stabilized the wild-type L9 domain, while the folding thermodynamics of the wild-type SH3 domain was unaffected. Simulations were performed using an extensive set of dielectric constants ranging between 15 and 150 and ion concentrations ranging from 0 to 0.15 M. Changing the dielectric constant or the salt concentration did not significantly affect the folding stability of the L9 domain (data not shown). Comparison of average distances between lysine 12 and other charged residues (in the ensemble of unfolded conformations) showed no change between the wild-type L9 and K12M L9, contrary to the result of a pH-dependent analysis reported by Raleigh *et al.*[33,53] Finally, the simulated $\Delta\Delta G$ bore no correlation to experimental $\Delta\Delta G$ values (Fig. 2). These observations illustrate that the C^{α} model is inappropriate for the study of charged residue effects on folding thermodynamics.

During calibration of the two-bead models, two possible positions for the side-chain bead were considered: on the C^{β} atom and on the farthest heavy atom taken from the PDB coordinates. Analysis of C^{β} model trajectories yielded a poor correlation with experimental $\Delta\Delta G$ values, similar to that obtained from the single-bead model results (data not shown). Placing the side-chain bead on the farthest heavy atom yielded a better correlation, probably because the charge was placed at a more precise position.

Using the native-topology-based potential in a simulation that involves a non-bonded potential, other than a simple Lennard–Jones (LJ) repulsion potential, presents a problem of scaling. A native contact involving oppositely charged beads (i.e., a salt bridge) will have both LJ and electrostatic attractive potentials, which will effectively strengthen their coefficient and shorten their optimal contact distance. For contacts involving same-sign charges, the reverse effect will occur. We have therefore tested four schemes to address this issue of calibrating the electrostatic strength. (1) In the first scheme, we applied both the LJ and DH potentials independently. The energetic contribution of the DH potential for two opposite charges located at a distance of 5 Å is 0.83 kcal/mol (for a dielectric constant of 80). In our model, each LJ contributes 1 kcal/mol. We note that this relative strength of DH compared to LJ interactions is similar to that in atomistic force fields such as CHARMM and AMBER. (2) In the second scheme, electrostatic interactions were applied for nonnative charged–charged interactions and were excluded for cases in which native contacts are formed between charged residues. (3) The third scheme compensated for electrostatic energy formed between charged beads that also form native contacts. The electrostatic energy was calculated at the optimal distance length of the native contact, and the LJ coefficient of that contact was changed in such a way that the sum of LJ and DH of that contact will be 1 kcal/mol. As a result, the potential between the pair of charged beads resembled an LJ potential at near contact distances and an electrostatic potential at far distances. (4) The fourth scheme was similar to the third scheme but applied only to native contacts between residues of opposite charges. For a native contact that is formed between residues of opposite charges, the LJ coefficient has to be reduced so that the total contribution of LJ and DH is 1 kcal/mol. The fourth scheme yielded the best correlation when comparing $\Delta\Delta G$ of the 11 mutants of NT L9, yet all the four models indicate that the mutant K12M is unique and exhibits stabilization effect.

Protein systems

Three protein systems were investigated:

1. The N-terminal domain of the ribosomal protein L9 (NTL9) is a regulatory RNA-binding module that binds to 23rRNA (PDB code: 1CQU[78]). NTL9 is composed of 56 residues, of which 12 are positively charged and 6 are negatively charged (Fig. 1a). NTL9 simulations were performed using a dielectric constant of 15 and an ion concentration of 0.1 M. The dielectric constant and ion concentration were chosen after extensive calibration (described in Model calibration) and showed the best correlation with experimental data. Similar results, although less pronounced, were obtained using a dielectric constant of 75 and an ion concentration of 0.1 M. Simulations were performed for the wild-type protein, 11 experimentally reported mutants, and 1 previously unreported mutant. The 11 experimentally reported charge removal mutations were as follows: Lys to Met at position 7, Asp to Asn at position 8, Lys to Met at position 10, Lys to Met at position 12, Lys to Met at position 14, Lys to Met at position 15, Glu to Gln at position 17, Asp to Asn at position 23, Glu to Gln at position 38, Glu to Gln at position 48, and Glu to Gln at position 54. The one previously unreported charge removal mutation was Lys to Met at position 19. Simulations were performed at five temperatures, ranging between 1.37 and 1.43 arbitrary units, which sufficiently sampled both the folded and unfolded states of the modeled protein. The NT L9 has 243 native contacts and the folded and unfolded states are defined by $Q \geq 143$ and $Q < 143$, respectively.
2. Ribonuclease Sa (RNase Sa) is a 96-residue enzyme secreted by *Streptomyces aureofaciens* (PDB code: 1T2H). RNase Sa has 5 positively charged and 12 negatively charged residues (Fig. 1b). RNase Sa simulations were performed using a dielectric constant of 75 and a salt concentration of 0 and 0.1 M, similar to the experimental conditions with which the results were compared. Simulations were performed for the wild type and for four mutants that had undergone charge reversal mutations: Asp to Lys at position 17, Asp to Lys at position 25, Glu to Lys at position 41, and Glu to Lys at position 74. Simulations were performed at 11 temperatures, ranging between 1.12 and 1.32 arbitrary units, which sufficiently sampled the folded, unfolded, and intermediate states of the modeled protein. RNase Sa has one disulfide bond, which was modeled by adding a bond potential between the side-chain beads of the cysteine residues participating in the disulfide bond. The RNase Sa has 439 native contacts and the folded and unfolded states are defined by $Q \geq 270$ and $Q < 270$, respectively.
3. BLA and HEWL are two homologous proteins of similar size and structure (Fig. 1c and d). The proteins have a similar number of residue–residue contacts (BLA has 651 and HEWL has 690) and their structural contact maps show a common folding

topology. Both proteins have four disulfide bonds, which overlap when structurally aligning the native structures of the proteins (PDB codes: 1F6R and 1DPX for BLA and HEWL, respectively). BLA has 13 positively charged and 18 negatively charged residues, while HEWL has 17 positively charged and 9 negatively charged residues. BLA and HEWL simulations were performed using a dielectric constant of 75 and an ion concentration of 0 M, which match the experimental conditions, and also without applying the electrostatic potential. BLA and HEWL simulations were performed at eight temperatures, ranging between 1.61 and 1.68, which sufficiently sampled the folded, unfolded, and intermediate states of the modeled proteins. The folded and unfolded states of BLA and HEWL are defined by $Q \geq 430$ and $Q < 350$, respectively.

Acknowledgements

This work was supported in part by the Kimmleman Center for Macromolecular Assemblies, the Israel Science Foundation and the Center for Complexity Science (Y.L). Y.L. is the incumbent of the Lilian and George Lyttle Career Development Chair. We thank Tzachi Hagai, Dalit Shental and Agnes Toth for insightful discussions.

References

1. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence–structure relationship. *Proc. Natl Acad. Sci. USA*, **89**, 8721–8725.
2. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600.
3. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19.
4. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953.
5. Simler, B. R., Levy, Y., Onuchic, J. N. & Matthews, C. R. (2006). The folding energy landscape of the dimerization domain of *Escherichia coli* Trp repressor: a joint experimental and theoretical investigation. *J. Mol. Biol.* **363**, 262–278.
6. Gu, Z., Rao, M. K., Forsyth, W. R., Finke, J. M. & Matthews, C. R. (2007). Structural analysis of kinetic folding intermediates for a TIM barrel protein, indole-3-glycerol phosphate synthase, by hydrogen exchange mass spectrometry and Go model simulation. *J. Mol. Biol.* **374**, 528–546.
7. Levy, Y., Caflisch, A., Onuchic, J. N. & Wolynes, P. G. (2004). The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. *J. Mol. Biol.* **340**, 67–79.
8. Levy, Y., Cho, S. S., Onuchic, J. N. & Wolynes, P. G. (2005). A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.* **346**, 1121–1145.
9. Karanicolas, J. & Brooks, C. L. (2003). Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J. Mol. Biol.* **334**, 309–325.
10. Ding, F., Guo, W., Dokholyan, N. V., Shakhnovich, E. I. & Shea, J. E. (2005). Reconstruction of the src-SH3 protein domain transition state ensemble using multiscale molecular dynamics simulations. *J. Mol. Biol.* **350**, 1035–1050.
11. Turjanski, A. G., Gutkind, J. S., Best, R. B. & Hummer, G. (2008). Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput. Biol.* **4**, e1000060.
12. Faraldo-Gomez, J. D. & Roux, B. (2007). On the importance of a funneled energy landscape for the assembly and regulation of multidomain Src tyrosine kinases. *Proc. Natl Acad. Sci. USA*, **104**, 13643–13648.
13. Zhang, Z. & Chan, H. S. (2009). Native topology of the designed protein Top7 is not conducive to cooperative folding. *Biophys. J.* **96**, L25–L27.
14. Shea, J. E., Onuchic, J. N. & Brooks, C. L. (1999). Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc. Natl Acad. Sci. USA*, **96**, 12512–12517.
15. Gosavi, S., Chavez, L. L., Jennings, P. A. & Onuchic, J. N. (2006). Topological frustration and the folding of interleukin-1 beta. *J. Mol. Biol.* **357**, 986–996.
16. Hills, R. D., Jr & Brooks, C. L., 3rd (2008). Subdomain competition, cooperativity, and topological frustration in the folding of CheY. *J. Mol. Biol.* **382**, 485–495.
17. Sulkowska, J. I., Sulkowski, P. & Onuchic, J. (2009). Dodging the crisis of folding proteins with knots. *Proc. Natl Acad. Sci. USA*, **106**, 3119–3124.
18. Ferreiro, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. (2007). Localizing frustration in native proteins and protein assemblies. *Proc. Natl Acad. Sci. USA*, **104**, 19819–19824.
19. Plotkin, S. S. (2001). Speeding protein folding beyond the G(o) model: how a little frustration sometimes helps. *Proteins*, **45**, 337–345.
20. Zarrine-Afsar, A., Wallin, S., Neculai, A. M., Neudecker, P., Howell, P. L., Davidson, A. R. & Chan, H. S. (2008). Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc. Natl Acad. Sci. USA*, **105**, 9999–10004.
21. Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I. & Sanchez-Ruiz, J. M. (1999). Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge–charge interactions to protein stability. *Biochemistry*, **38**, 8138–8149.
22. Pace, C. N. (2000). Single surface stabilizer. *Nat. Struct. Biol.* **7**, 345–346.
23. Sanchez-Ruiz, J. M. & Makhatadze, G. I. (2001). To charge or not to charge? *Trends Biotechnol.* **19**, 132–135.
24. Kumar, S. & Nussinov, R. (2002). Close-range electrostatic interactions in proteins. *ChemBioChem*, **3**, 604–617.
25. Xiao, L. & Honig, B. (1999). Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* **289**, 1435–1444.
26. Gribenko, A. V. & Makhatadze, G. I. (2007). Role of the charge–charge interactions in defining stability and halophilicity of the CspB proteins. *J. Mol. Biol.* **366**, 842–856.
27. Marshall, S. A., Morgan, C. S. & Mayo, S. L. (2002). Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol.* **316**, 189–199.
28. Luisi, D. L., Snow, C. D., Lin, J. J., Hendsch, Z. S., Tidor, B. & Raleigh, D. P. (2003). Surface salt bridges,

double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein l9. *Biochemistry*, **42**, 7050–7060.
29. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L. *et al.* (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **8**, 1843–1849.
30. Strickler, S. S., Gribenko, A. V., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T. *et al.* (2006). Protein stability and surface electrostatics: a charged relationship. *Biochemistry*, **45**, 2761–2766.
31. Gribenko, A. V., Patel, M. M., Liu, J., McCallum, S. A., Wang, C. & Makhatadze, G. I. (2009). Rational stabilization of enzymes by computational redesign of surface charge–charge interactions. *Proc. Natl Acad. Sci. USA*, **106**, 2601–2606.
32. Pace, C. N., Alston, R. W. & Shaw, K. L. (2000). Charge–charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci.* **9**, 1395–1398.
33. Cho, J. H., Sato, S., Horng, J. C., Anil, B. & Raleigh, D. P. (2008). Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability. *Arch. Biochem. Biophys.* **469**, 20–28.
34. Zhou, H. X. (2002). Residual electrostatic effects in the unfolded state of the N-terminal domain of L9 can be attributed to nonspecific nonlocal charge–charge interactions. *Biochemistry*, **41**, 6533–6538.
35. Weinkam, P., Pletneva, E. V., Gray, H. B., Winkler, J. R. & Wolynes, P. G. (2009). Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles. *Proc. Natl Acad. Sci. USA*, **106**, 1796–1801.
36. Dill, K. A. & Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825.
37. Elcock, A. H. (1999). Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability. *J. Mol. Biol.* **294**, 1051–1062.
38. Zagrovic, B., Snow, C. D., Khaliq, S., Shirts, M. R. & Pande, V. S. (2002). Native-like mean structure in the unfolded ensemble of small proteins. *J. Mol. Biol.* **323**, 153–164.
39. Zhou, H. X. (2002). A Gaussian-chain model for treating residual charge–charge interactions in the unfolded state of proteins. *Proc. Natl Acad. Sci. USA*, **99**, 3569–3574.
40. Mor, A., Haran, G. & Levy, Y. (2008). Characterization of the unfolded state of repeat proteins. *HFSP J.* **2**, 405–415.
41. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl Acad. Sci. USA*, **101**, 12497–12502.
42. Yi, Q., Scalley-Kim, M. L., Alm, E. J. & Baker, D. (2000). NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* **299**, 1341–1351.
43. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature*, **437**, 1053–1056.
44. Shortle, D. & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science*, **293**, 487–489.
45. Pappu, R. V., Srinivasan, R. & Rose, G. D. (2000). The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proc. Natl Acad. Sci. USA*, **97**, 12565–12570.
46. Ding, F., Jha, R. K. & Dokholyan, N. V. (2005). Scaling behavior and structure of denatured proteins. *Structure*, **13**, 1047–1054.
47. Cortajarena, A. L., Lois, G., Sherman, E., O'Hern, C. S., Regan, L. & Haran, G. (2008). Nonrandom coil behavior as a consequence of extensive PPII structure in the denatured state. *J. Mol. Biol.* **382**, 203–212.
48. Shan, B., Eliezer, D. & Raleigh, D. (2009). The unfolded state of the C-terminal domain of the ribosomal protein L9 contains both native and non-native structure. *Biochemistry*, **48**, 4707–4719.
49. Shortle, D., Chan, H. S. & Dill, K. A. (1992). Modeling the effects of mutations on the denatured states of proteins. *Protein Sci.* **1**, 201–215.
50. Cho, J. H. & Raleigh, D. P. (2006). Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure. *J. Mol. Biol.* **359**, 1437–1446.
51. Kundrotas, P. J. & Karshikoff, A. (2002). Modeling of denatured state for calculation of the electrostatic contribution to protein stability. *Protein Sci.* **11**, 1681–1686.
52. Stigter, D., Alonso, D. O. & Dill, K. A. (1991). Protein stability: electrostatics and compact denatured states. *Proc. Natl Acad. Sci. USA*, **88**, 4176–4180.
53. Anil, B., Li, Y., Cho, J. H. & Raleigh, D. P. (2006). The unfolded state of NTL9 is compact in the absence of denaturant. *Biochemistry*, **45**, 10110–10116.
54. Alston, R. W., Lasagna, M., Grimsley, G. R., Scholtz, J. M., Reinhart, G. D. & Pace, C. N. (2008). Tryptophan fluorescence reveals the presence of long-range interactions in the denatured state of ribonuclease Sa. *Biophys. J.* **94**, 2288–2296.
55. Takagi, F., Koga, N. & Takada, S. (2003). How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. *Proc. Natl Acad. Sci. USA*, **100**, 11367–11372.
56. Baumketner, A., Jewett, A. & Shea, J. E. (2003). Effects of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape. *J. Mol. Biol.* **332**, 701–713.
57. Cheung, M. S., Klimov, D. & Thirumalai, D. (2005). Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl Acad. Sci. USA*, **102**, 4753–4758.
58. Shental-Bechor, D. & Levy, Y. (2008). Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc. Natl Acad. Sci. USA*, **105**, 8256–8261.
59. Hoiberg-Nielsen, R. R., Arleth, L. & Westh, P. (2009). The effect of glycosylation on interparticle interactions and dimensions of native and denatured phytase. *Biophys. J.* **96**, 153–161.
60. Halskau, O., Jr., Perez-Jimenez, R., Ibarra-Molero, B., Underhaug, J., Munoz, V., Martinez, A. & Sanchez-Ruiz, J. M. (2008). Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics. *Proc. Natl Acad. Sci. USA*, **105**, 8625–8630.
61. Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. (1999). pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry*, **38**, 4896–4903.
62. Cho, J. H., Sato, S. & Raleigh, D. P. (2004). Thermodynamics and kinetics of non-native interactions in

protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *J. Mol. Biol.* **338**, 827–837.
63. Cho, J. H. & Raleigh, D. P. (2005). Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. *J. Mol. Biol.* **353**, 174–185.
64. Anil, B., Craig-Schapiro, R. & Raleigh, D. P. (2006). Design of a hyperstable protein by rational consideration of unfolded state interactions. *J. Am. Chem. Soc.* **128**, 3144–3145.
65. Tanford, C. (1968). Protein denaturation. *Adv. Protein Chem.* **23**, 121–282.
66. Trefethen, J. M., Pace, C. N., Scholtz, J. M. & Brems, D. N. (2005). Charge–charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa. *Protein Sci.* **14**, 1934–1938.
67. Ivankov, D. N., Garbuzynskiy, S. O., Alm, E., Plaxco, K. W., Baker, D. & Finkelstein, A. V. (2003). Contact order revisited: influence of protein size on the folding rate. *Protein Sci.* **12**, 2057–2062.
68. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
69. Onuchic, J. N. & Wolynes, P. G. (2004). Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75.
70. Tanford, C. & Kirkwood, J. (1957). Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* **79**, 5333–5339.
71. Cheung, M. S., Finke, J. M., Callahan, B. & Onuchic, J. N. (2003). Exploring the interplay between topology and secondary structural formation in the protein folding problem. *J. Phys. Chem. B*, **107**, 11193–11200.
72. Schlick, T. (2002). Molecular modeling and simulation—An interdisciplinary guide, pp. 298–301.
73. Givaty, O. & Levy, Y. (2009). Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.* **385**, 1087.
74. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E. & Edelman, M. (1999). Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15**, 327–332.
75. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* **13**, 1011–1021.
76. Hagai, T. & Levy, Y. (2008). Folding of elongated proteins: conventional or anomalous? *J. Am. Chem. Soc.* **130**, 14253–14262.
77. Koga, N. & Takada, S. (2001). Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* **313**, 171–180.
78. Luisi, D. L., Kuhlman, B., Sideras, K., Evans, P. A. & Raleigh, D. P. (1999). Effects of varying the local propensity to form secondary structure on the stability and folding kinetics of a rapid folding mixed alpha/beta protein: characterization of a truncation mutant of the N-terminal domain of the ribosomal protein L9. *J. Mol. Biol.* **289**, 167–174.