JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Predictions from an Ising-like Statistical Mechanical Model on the Dynamic and Thermodynamic Effects of Protein Surface Electrostatics

Athi N. Naganathan*

Department of Biotechnology, Indian Institute of Technology Madras, Chennai-600036, India

Supporting Information

ABSTRACT: Charged residues on the surface of a protein are known hot-spots for post-translational modification, protein/ligand-binding, and tuning conformational stabilities. Recent experimental evidence points to the fact that surface electrostatics can also modulate thermodynamic barriers and hence folding mechanisms. To probe for this behavior across different proteins, we develop a novel version of the Wako−Saitô−Muñoz−Eaton (WSME) model in which we include an electrostatic potential term in the energy function while simplifying the treatment of solvation free energy. Both of the energy terms are obtained by quantitatively fitting the model to differential scanning calorimetry (DSC) experiments that carry critical information on the protein partition function. We characterize four sets of structural/functional homologues (HEWL/BLA, CspB, engrailed, α-spectrin) either by fitting the experimental data of a single domain in the homologous set and predicting the conformational behavior of the rest with the same set of parameters or by performing semiblind predictions. The model with the added electrostatic term is able to successfully reproduce the order of thermodynamic stabilities and relaxation rates of most of the homologues. In parallel, we predict diverse conformational features including a wide range of thermodynamic barriers (∼9−40 kJ/mol), broad native ensembles in helical proteins, structured unfolded states and intermediates, rugged folding landscapes, and further provide an independent protein-specific estimate of the folding speed limit at 298 K (1/(7−300 μs)). Our results are evidence that protein surface electrostatics can be tailored to not only engineer stabilities but also folding mechanisms and the ruggedness of the underlying landscape.

INTRODUCTION

Charged amino acids are unique among the 20-letter code fine-tuned by the mechanism of Natural Selection. They are asymmetrically distributed with a large relative occurrence on the surface of proteins that makes them a principal target for post-translational modifications like phosphorylation, methylation, acylation, etc., thus playing a major role in modulating the functions of proteins and enzymes.[1] Their strategic location on the protein surface enables them to recruit binding partners (protein or ligand) through long-range electrostatic interactions that in turn influence a variety of cellular pathways in response to external stimuli. The longer length-scales of charge−charge interactions have been further exploited by Nature to tune the conformational stabilities of proteins to match with the functional requirements.[2] Recognizing this, a large amount of work has gone into stabilizing proteins by optimizing electrostatic interactions either in the native[3−5] or in the unfolded states.[6]

Recent experimental work, however, reveals that protein surface electrostatics can also modulate folding mechanisms.[7] Specifically, the structural homologues hen egg-white lysozyme (HEWL) and the apo-form of bovine lactalbumin (BLA) have almost identical tertiary structure (RMSD ∼1.5 Å) and size (121 and 128 residues, respectively) but perform very different functions in the cell. Their peak DSC temperatures differ by more than 30 K and show a marked difference in the broadness despite the structural similarities. A closer analysis revealed that this difference can be attributed to the distinct distribution of charged residues in the two proteins.[7] The authors estimate the difference in thermodynamic barriers to be a remarkable 33 kJ/mol and hint that functional requirements are the source of these contrasting folding behaviors. Coarse-grained simulations with electrostatic terms provide qualitative support to the differences in thermodynamic behavior.[8]

The relevant question is then how common is this kind of folding behavior? In other words, how much say does electrostatics have in determining thermodynamic barriers? This is a nontrivial problem as it is challenging to quantitatively reproduce equilibrium experimental data (e.g., the heat capacity profile) from all-atom simulations mainly due to force-field and sampling limitations. Coarse-grained molecular dynamics simulations are increasingly performing better in reproducing equilibrium unfolding and rates particularly upon the introduction of Tanford's molecular transfer model.[9,10] The Wako−Saitô−Muñoz−Eaton (WSME) model provides a practical and attractive alternative in this regard as it relies on an ensemble description of the folding process albeit with assumptions that local contacts form first and that long-range contacts require all of the intervening residues to be folded. Despite the limiting assumptions, it has been successfully used to quantitatively characterize the folding of peptides[11] and proteins.[12−14] The exact-solution approach to the WSME model[15−18] that enumerates the contribution to the unfolding energetics from 2^N structured microstates (where N is the protein length) presents an even more physical description of

Received: May 1, 2012
Published: October 5, 2012

dx.doi.org/10.1021/ct300676w | J. Chem. Theory Comput. 2012, 8, 4646−4656

the folding process particularly with the addition of solvation energetics.[19]

Experimentally, DSC is the method of choice to study global protein unfolding thermodynamics. It has been long known that such measurements of the temperature dependence of enthalpy fluctuations in proteins carry information on the partition function.[20−22] However, the means to extract the underlying distribution of states by quantitatively analyzing experimental DSC profiles while at the same accounting for the ensemble nature of the folding process was lacking due to the absence of simple statistical models. This scenario changed recently starting with the use of the Single Sequence Approximation (SSA) approach to the WSME model (i.e., a quasi-continuous description of the folding ensemble constructed by accounting for only single stretch of folded-like residues) to identify downhill folding in the helical protein BBL[13] and the development of the Variable Barrier (VB) model that bases its description on the Landau theory of phase transitions.[23] The thermodynamic barriers predicted by analyzing the heat capacity profiles using the VB model scale well the experimental relaxation rates[24] (r = 0.95), highlighting the power of the approach. Since then, both the WSME in its various forms and the VB models have been successfully employed to determine the folding thermodynamic barriers for several fast-folding proteins[23,25−28] that are not only in agreement with each other but also with independent approaches.[27,29]

In the current work, we take advantage of the availability of the high-quality DSC data on HEWL/BLA to introduce electrostatic potential term into the exact-solution version of the WSME model and simplify the previous treatment of solvation energetics. HEWL/BLA is an exciting system as size, topological effects, and any influence of experimental conditions are factored out, providing clear insight into the subtle sequence-specific contributions to their conformational behavior. Moreover, because electrostatic interactions are the only major difference between HEWL and BLA, analyzing them affords a systematic procedure to parametrize electrostatics in the WSME model. We then use it study three sets of homologues: structural homologues that differ in the source organisms and functions (engrailed family[30]), structural and functional homologues from organisms that grow at different temperatures (CspB family[31]), and structural and functional homologues from within a tandem-repeat protein (α-spectrin family[32]), which span a size range of 50−128 residues and cover all of the three structural classes (all-α, all-β, and α/β).

THEORETICAL BASIS

The Wako−Saitô−Muñoz−Eaton (WSME) model has been discussed extensively in the original works.[11,12,15,16] Simply, the model bases its description on the native structure of the protein, and hence the energetics are purely Gō-like.[33] Each residue is assumed to sample two sets of conformations, folded and unfolded, which are allotted the binary variables 1 and 0, respectively. Because the unfolded conformations significantly outnumber the former, fixing each residue in a native conformation is associated with a large entropic cost that is accounted for by the parameter ΔS_{conf}. Two amino acids are assumed to interact only if all of the intervening residues are also native-like, that is, a string of 1's connecting the two interacting residues, resulting in 2^N microstates for an N-residue protein. It is possible to calculate the exact solution to this simple model, that is, enumerate all of the species and calculate their contribution to the partition function, either through transfer-matrix formalisms[16,17] or by accumulating the partition function from every structured species.[18]

In earlier exact-solution descriptions,[16−18] the effective free energy of a microstate with a string of native-like residues located between residues m and n is determined by the simplest form of the Gibbs relation that assigns a mean interaction energy to contacting residues (ξ_{int}) within that stretch and a conformational entropic cost (ΔS_{conf}) for fixing each residue in native conformation:

$$\Delta G_{m,n} = \sum_{m,n} \xi_{int} - T \sum_{m}^{n} \Delta S_{conf} \tag{1}$$

where the first summation includes all of the residue−residue interactions within (m, n). In a recent work, a more physical treatment of energetics that included solvation was developed (called WSME-S where S stands for solvation).[19] Here, enthalpy and entropy are obtained by integrating the heat capacities, which are themselves calculated as temperature-dependent empirical polynomial functions of the molecular weight, polar, and apolar surface areas of the microstates derived from the work of Freire.[34] The surface areas are approximated from the number of contacts in the microstate relative to that in the fully folded native state. However, WSME-S does not differentiate between charged and uncharged residues that is a requisite for a more complete description of protein systems.

In this regard, we modify the WSME model in two aspects. First, we include an electrostatic potential term in the energy-function based on the Debye−Hückel model.[35] Second, we employ a simpler approach to the solvation free energy by employing the Gibbs relation as discussed below. The first part of eq 1 now becomes the stabilization free energy of a particular microstate (m,n) represented as a sum of van der Waals interactions (E_{VdW}), electrostatic potential (E_{elec}), and solvation free energy (ΔG_{solv}):

$$\Delta G^{stab}_{m,n} = E_{VdW} + E_{elec} + \Delta G_{solv} \tag{2}$$

The interacting partners that contribute to the van der Waals energy are identified by setting a distance cutoff (r_{cut}) to the pairwise heavy-atom partners (i, j) calculated from the PDB file.

$$E_{VdW} = \sum_{m,n} \xi_{i,j}\rho \tag{3}$$

where ρ = 1 if $r_{ij} \le r_{cut}$ and ρ = 0 otherwise.

The Debye−Hückel (DH) theory has been successfully employed to model electrostatic effects into coarse-grained treatments to study the folding and binding of both proteins[8,36,37] and nucleic acids.[38−41] This prompted us to employ a similar description for the electrostatic potential term:

$$E_{elec} = \sum_{m,n} K_{Coulomb} \frac{q_i q_j}{\varepsilon_{eff} r_{ij}} \exp(-r_{ij}\kappa) \tag{4}$$

wherein $K_{Coulomb}$ is the Coulomb constant (1389 kJ·Å/mol), q_i is the charge on the atom i, r_{ij} is the distance between charge centers i and j, and ε_{eff} is the effective dielectric constant (see below). $1/\kappa$ is the Debye screening length that depends on ε_{eff}, solvent ionic-strength (I), and temperature (T) as

$$\kappa^2 = \frac{8\pi e^2 I}{\varepsilon_{eff} k_B T} \tag{5}$$

dx.doi.org/10.1021/ct300676w | J. Chem. Theory Comput. 2012, 8, 4646−4656

where k_B is Boltzmann's constant, and e is the elementary charge.

The solvation free energy is approximated as being proportional to the number of formed contacts ($x_{cont}^{m,n}$) in that microstate with the proportionality constant being ΔC_p^{cont}, which is the temperature-independent heat capacity change upon fixing a native contact. This is similar to the previous treatment,[19] but accessible surface areas and hence Freire's empirical treatment of heat capacity are not included as we pose the relations directly in the form of free energy. Although the treatment of solvation in this approximation is simplifying, it liberates the model from the dependence on Freire's parameters that are empirical. Moreover, it has been shown before that the number of residues in a protein, the change in heat capacity and accessible surface areas upon unfolding, and the number of contacts are mutually correlated with each other.[42] Therefore:

$$\Delta G_{solv} = x_{cont}^{m,n} \Delta C_p^{cont} [(T - T_{ref}) - T \ln(T/T_{ref})] \quad (6)$$

where T_{ref} is the reference temperature, which is fixed to 385 K.[43] For the sake of simplicity, a uniform conformational entropy cost (ΔS_{conf}) is assigned to residues independent of sequence or native secondary structure.

The partition function (Z) is calculated employing the transfer-matrix formalism of Wako and Saitô.[15,16] The excess heat capacity is derived from the partition function as:

$$C_p^{ex} = 2RT\left(\frac{d \ln Z}{dT}\right) + RT^2\left(\frac{d^2 \ln Z}{dT^2}\right) \quad (7)$$

from which DSC curves can be reproduced employing:

$$C_p^{fit} = C_p^{ex} + C_p^{Unf} \quad (8)$$

and

$$C_p^{Unf} = [a + (b/1000)(T - 273)] \cdot M_r \quad (9)$$

where a and b are the intercept and slope of the linear unfolded (Unf) baseline, respectively, and M_r is the molecular mass of the protein in g/mol.

Model Parameterization. I employ experimental data and particularly DSC thermograms to parametrize the current version of WSME. The model has six parameters: the interaction energy per native contact ($\xi_{i,j}$; henceforth referred to as ξ), the entropic cost (ΔS_{conf}), the heat capacity change upon fixing a native contact (ΔC_p^{conf}), a and b that determine the slope and intercept of the unfolded baseline, and the effective dielectric constant (ε_{eff}). Apart from these, r_{cut} and the definition of nearest neighbors are variables that need to be defined to identify native contacts. Atoms of charged residues are given the following charges following the pH 7 protonation states unless specified otherwise: atoms NE, NH1, NH2 of arginine were assigned a charge of 0.33 each, NZ of lysine a charge of 1, and OD1, OD2 of aspartate and OE1, OE2 of glutamate a charge of −0.5. All-to-all electrostatic interactions are considered that eliminate assumptions implicit in using cut-offs. Ionic strength values are taken from the concerned literature. We choose r_{cut} to be 6 Å and consider only those nearest neighbors defined by $j > i + 1$. This choice of contact description reduces the relative contribution of local interactions to the total energetics, thus making the protein systems as two-state-like as possible. In other words, the barriers will be slightly overestimated as compared to previous predictions using the VB or free-energy surface-based approaches. However, because we are dealing only with homologues in this study, any effect of the choice of contact description is relative, thus rendering the results robust. All parameters except ε_{eff} can be directly obtained by fitting the model to experimental DSC thermograms.

The choice of ε_{eff} is however a nontrivial issue,[44] and several values exist in the literature ranging from ∼4 (protein interior) to 78.5 (water). For models such as the WSME where the effect of solvent is implicitly included in the energetics, the value of ε_{eff} should fall within the range specified above. Furthermore, because we are interested only in surface electrostatics, that is, charged residues exposed to the solvent, it is instructive to calculate the distribution of relative accessible surface areas (ASA) of charged residues in proteins. Accordingly, we constructed a nonredundant database of 2515 single-domain proteins spanning the size-range of 30−150 residues without any bound ligands from the protein data bank (PDB). Statistical analysis indicates that charged residues (K, R, D, E) are predominantly solvent exposed (>83% using the typical relative ASA cutoff of 30%) with a distinct Gaussian-like distribution when compared to that of highly hydrophobic residues (V, L, I, F, W) (Figure S1). Moreover, the charged atom is typically more exposed to the solvent than the mean % ASA value for the charged residues binned in Figure S1. This suggests that it is possible to identify a generic ε_{eff} that can be used as a zeroth-order approximation to treat surface electrostatics in simple models.

A simple procedure to extract this value employing the WSME model is to analyze structural homologues that differ only in the surface charge distribution. In such a scenario, all energetic terms would be expected to be very similar between them, and the predicted stability can be matched with the experimental ones by simply tuning the magnitude of ε_{eff}. Taking advantage of the HEWL/BLA protein set, we first fit the thermogram of HEWL by fixing ε_{eff} to different values between 4 and 80 and floating the other five parameters (ξ, ΔS_{conf}, ΔC_p^{cont}, a, b) with the contact description discussed above (as the experiments were carried out at pH 4.5 for HEWL and pH 5.6 for apo-BLA, we also assigned a charge of 0.5 to the atoms ND1 and NE2 of histidine). We then simply replace the contact map of HEWL with that of apo-BLA for each of the fit parameter sets (with no additional fitting, the contact-map replacement procedure) and check if the predicted peak DSC temperature of apo-BLA matches with that of the experiment (Figure 1). This systematic exercise resulted in a value of 29 for ε_{eff}.

The magnitude of 29 for ε_{eff} is also sufficient to qualitatively reproduce the unique thermodynamic behavior of the one-state downhill folding protein BBL that refolds continuously upon increasing ionic strength starting from low pH conditions[45] (Figure S2). This indicates that the attractive and repulsive forces arising primarily from charged residues are well accounted for in this approach. To eliminate possible solution-specific contribution to the energetics of folding (ion-binding, entropic effects of solvent ordering in concentrated ionic solutions, etc.), we focus on validating the model comparing homologues. For the CspB- and engrailed-family homologues, we compare the experimental rates at the chemical midpoint to that calculated from the free-energy surface at T_0 (temperature at which folded and unfolded minima have similar free energy). This procedure is expected to work well as the magnitude of folding free-energy barrier

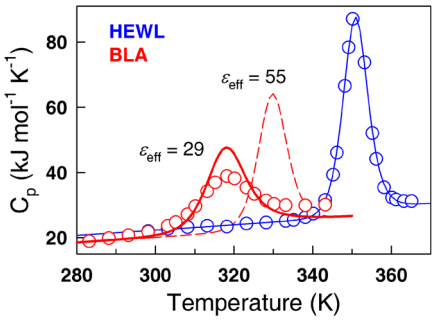

Figure 1. Estimating ε_{eff} employing the contact-map replacement procedure. Experimental heat capacity profiles of HEWL (blue ○) and apo-BLA (red ○) together with the fits from the WSME model (with electrostatics) following the same color code for different values of the effective dielectric constant.

heights changes little at temperatures below the midpoint of folding due to known entropy−enthalpy compensations.[46]

RESULTS

Thermodynamic Features of the HEWL/BLA Family. Figure 2A plots the predicted heat capacity curve for apo-BLA (red curve) from the model with electrostatics. The agreement is quite good but is still sharper than the experimental thermogram.[7] It is important to emphasize that the parameters are identical for both HEWL and apo-BLA except for the contact map (and hence the distribution of surface electrostatics), which is obtained from the PDB file. It is also clear that both experimental and predicted profiles are significantly broader than those of HEWL. Yet is this broadness a result of the temperature dependence of enthalpy (as they unfold at different temperature ranges) or due to changes in thermodynamic barriers? A simple approach to answer this question is to destabilize the HEWL thermogram by changing ξ and matching the predicted peak DSC temperature with that of the apo-BLA. Such a procedure results in a much sharper thermogram for HEWL (black in Figure 2A) that differs significantly from that of apo-BLA. The difference in broadness is therefore an unambiguous consequence of changes in thermodynamic barriers.

The barriers at around the respective characteristic temperatures are indeed quite different for both proteins ranging from ∼36 kJ/mol for HEWL to just ∼13 kJ/mol for apo-BLA (Figure 2B). An independent estimate of the thermodynamic barriers that employed the variable barrier model[23] came up with values of ∼37 and ∼3 kJ/mol, for HEWL and apo-BLA, respectively.[7] The estimated barrier for apo-BLA in the current study is higher but is also consistent with the sharper thermogram we predict. The WSME model without the electrostatic potential term is unable to reproduce the differences in stability or the broadness and hence the barriers (Figure S3). These results provide further compelling evidence that surface electrostatics is responsible for such large changes in thermodynamic barriers. The differences in barriers should correspond to >10 000 times faster folding for apo-BLA as compared to HEWL. However, experiments reveal only a minor difference in the midpoint relaxation rates with apo-BLA folding slightly faster (2.5×10^{-3} for apo-BLA vs 3.2×10^{-4} s^{-1} for HEWL).[47]

To investigate the possible origins of slower folding in apo-BLA, we calculate the simplest form of the free-energy landscape as represented by the single-sequence approximation (SSA, sequences with a single stretch of folded-like residues) with the same set of parameters from the exact-solution fit. The probability of SSA species, that is, Z_{SSA}/Z, is small (∼10%) but provides a simple avenue to look for intermediates and their structure. We find that the landscape of HEWL is almost

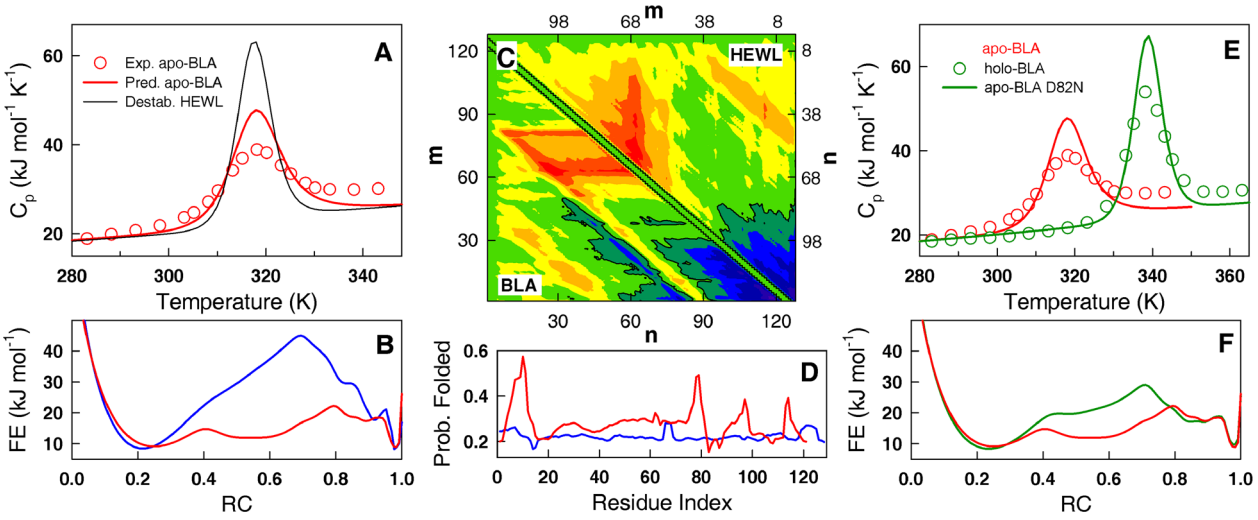

Figure 2. Thermodynamic features of the HEWL/BLA family. (A) Experimental (red ○) and predicted (red curve) heat capacity profiles of apo-BLA with ε_{eff} of 29. The hypothetical thermogram of destabilized HEWL is shown as a black curve. (B) Predicted one-dimensional free-energy surfaces for HEWL (blue) and apo-BLA (red) at T_0 with the fraction of native-residues as a reaction coordinate (RC). (C) A contour plot of the free-energy landscape as calculated by the single-sequence approximation for HEWL (upper right triangle; right and top axes) and apo-BLA (lower left triangle; left and bottom axes). Each species is represented by the coordinate (m,n) where m is the starting residue and n is the number of structured residues. For example, the native-state of apo-BLA is represented as (1,121). Note that the axes scales and identities are different for the proteins. Spectral color-coding is employed ranging from low (violet; high probability) to high free energy (red; low probability). (D) Residual unfolded state structure in apo-BLA as compared to that from HEWL at temperatures equidistant from the respective peak DSC temperatures (333 and 365 K). (E) The experimental heat capacity curve of holo-BLA (green ○) and the predicted profile from a D82N apo-BLA mutant (green curve). (F) The predicted free-energy surfaces at T_0 following the color code used in panel E.

dx.doi.org/10.1021/ct300676w | J. Chem. Theory Comput. 2012, 8, 4646−4656

perfectly "funneled" toward the native state (upper right triangle in Figure 2C and Figure S4), although the one-dimensional projection (Figure 2B) hints at a ∼2% probability of a partially structured state at a reaction coordinate value corresponding to 92% of structured residues. Apo-BLA, on the other hand, reveals a contrasting landscape characterized by several valleys and ridges with even the native state being quite heterogeneous (lower left triangle in Figure 2C and Figure S4). Apart from the native well, the β-rich region encompassing residues 40−80 is the most structured with residues 1−80 of apo-BLA forming an intermediate-like state (i.e., residues 81−121 are partially unstructured). Previous NMR studies point to the C-terminus region (K108−E121) to be displaying the largest flexibility and with the C-helix (T86−V99) showing only minimal chemical shift index information,[48] suggesting significant conformational heterogeneity in this stretch. In other words, the C-terminal region including the C-helix is expected to be the least stable completely consistent with our predictions. The presence of several partially structured states for apo-BLA is also in agreement with folding kinetic studies that display multiple phases.[47,49] Our results thus indicate that the relatively slower folding of apo-BLA despite the smaller thermodynamic barrier possibly arises from a more rugged landscape characterized by several valleys that act as kinetic traps during folding along with the high degree of electrostatic frustration in the native structure,[7] thus slowing the effective diffusion coefficient (D_{eff}, or the pre-exponential factor in the rate equation).

No attempt was made to overfit the apo-BLA thermogram with the model as its native ensemble is quite heterogeneous as shown before[7] and as seen from the SSA free-energy landscape. The unfolded state of apo-BLA is more structured than HEWL as is apparent from both the free-energy profiles and the residue probabilities at high temperatures that show islands of structure (Figure 2D). It is also known experimentally that apo-BLA has a significantly structured unfolded state,[50] further validating our results. The ability of charge distributions to modulate folding barrier and stability raises an interesting question: can apo-BLA be mutated such that it exhibits properties similar to those of the calcium-bound holo-form? In the holo-BLA structure, an array of aspartic residues coordinate with the calcium ion, thus effectively reducing the net repulsion between aspartic residues and as a result increasing both the conformational stability (DSC peak at ∼339 K) and the thermodynamic barrier (5.6 kJ/mol). To mimic the presence of calcium ion, we computationally mutate the aspartic acid 82 to asparagine (D82N) in apo-BLA, thus reducing the net negative charge. This simple mutation is sufficient to reproduce the stability of holo-BLA quite accurately (Figure 2E). The barrier is also increased to ∼19.9 kJ/mol (Figure 2F) in line with independent predictions from the VB model although to a larger extent that is again consistent with the predicted sharper thermogram.

The agreement of the model predictions with experiments at a quantitative level apart from the consistency with several independent analyses (calorimetry, NMR, kinetics) despite not explicitly including disulfide bond information is evidence that the effective dielectric constant of 29 is a reasonable estimate to deal with electrostatics in this model. However, the correspondence with just one set of data might be a coincidence. We therefore study more structural homologues to not only validate the model further but also to look for similar characteristic role for electrostatics in protein folding.

Thermodynamic and Dynamic Behavior of the CspB Family. The distribution of charged residues on the protein surface among homologues from mesophilic, thermophilic, and hyperthermophilic organisms is one of the factors known to influence their conformational stability. Specifically, homologues from mesophilic organisms are least stable, while those from hyperthermophiles have maximal stability. This stabilizing effect has been directly linked to the increase in the number and network of salt bridges on the protein surface.[2,51]

To check if the model captures this behavior, we analyze the homologues from the all-β cold shock protein CspB family as equilibrium and kinetic data are available including the structural information.[31] They are CspB from *Bacillus subtilis* (mesophile; Bs-CspB; 68 residues; 7 salt-bridges), *Bacillus caldolyticus* (thermophile; Bc-CspB; 67; 15), and *Thermotoga maritima* (hyper-thermophile; Tm-CspB; 67; 17). The mean pairwise RMSD between them is just ∼2 Å, indicating that the three-dimensional structures are very similar. We first fit the model to the available DSC data on Bs-CspB[4] by fixing the ε_{eff} to 29 while floating the other five parameters. The fit is of very good quality and is comparable to that from a two-state model (Figure 3A). As before, we then employ the same fitted parameter set and simply replace the contact map of Bs-CspB with that from Bc-CspB and Tm-CspB and thus predict the heat capacity profiles of the thermophilic proteins at constant ionic strength of 0.1 M. The predicted thermograms are

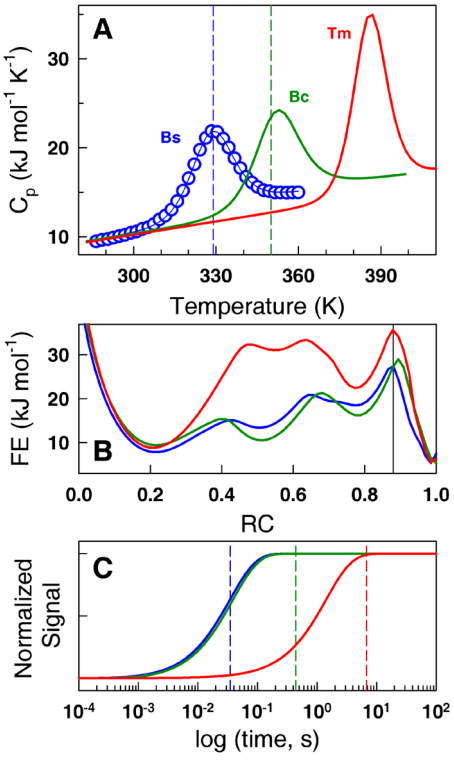

Figure 3. Thermodynamic and dynamic behavior of the CspB family. (A) Experimental heat capacity profile of Bs-CspB (○) together with its fit and the corresponding predictions for Bc-CspB and Tm-CspB. The dashed vertical lines are the experimental melting temperatures. (B) One-dimensional free-energy surfaces of the homologues at T_0. (C) Simulated relaxation profiles using a signal cutoff on the FES (vertical line in panel B) upon a jump from an unfolded state centered at low RC values mimicking kinetic chemical denaturation experiments. The dashed vertical lines signal the experimental relaxation times.

dx.doi.org/10.1021/ct300676w | J. Chem. Theory Comput. 2012, 8, 4646−4656

characterized by higher melting temperatures with the T_m of Bc-CspB in very good agreement with experiments. Tm-CspB displays an even higher melting temperature of 387 K that is as expected higher than the optimal growth temperature of *Thermotoga maritima* (~353 K). However, Tm-CspB also displays a more compact packing of side-chains that translates to more native contacts and higher stability (Figure S5A) possibly due to differences in the structure calculation protocol (NMR for Tm-CspB while X-ray for Bs- and Bc-CspB), thus precluding an unbiased comparison with the experimental T_m of ~355 K at 0.02 M ionic strength.[52] Irrespective of this, we show that the difference in the distribution of surface electrostatics is sufficient to increase the stability of thermophilic proteins in agreement with previous analysis,[51] further validating the model and the estimated ε_{eff}.

Another interesting feature is the kinetic behavior of the CspB family homologues. The relaxation rates at the chemical denaturation midpoint differ by more than 2 orders of magnitude with Bs-CspB folding fastest and Tm-CspB folding slowest.[31] Is this behavior an effect of the difference in the surface charge distribution? The free-energy profiles at around the T_0 are characterized by several bumps with Tm-CspB displaying the highest barrier qualitatively consistent with the trend in relaxation rates and the observed sharper thermogram (Figure 3B). These main features are also evident in predictions from the basic WSME model without electrostatics, although the maximal barriers are almost identical for all three proteins (Figure S5B). A recent Gō-model simulation identified potential intermediates in the CspB family[53] consistent with our predictions of a complex multistate-like free-energy surface. The predicted maximal barrier height for Bs-CspB at T_0 from WSME model with electrostatics is ~19.5 kJ/mol, resulting in an effective diffusion coefficient of ~1/(26 μs) consistent with recent low temperature experimental measures on mesophilic downhill proteins[54,55] and predictions from free-energy surface models.[24,29] A one-dimensional diffusive calculation on the FES[56] is able to reproduce the exponential decays observed experimentally but is unable to account for the difference in midpoint relaxation rates between the homologues (Figure 3C and Figure S6).

What is the source of this discrepancy? One reason for this could simply be that the model and/or its approximations are not adequate. On the other hand, a closer look at the experimental folding data on CspB homologues reveals that Bs- and Bc-CspB display similar equilibrium *m*-values of 7.4 and 7.5 kJ/(mol·M), respectively, while that of Tm-CspB is higher (7.9 kJ/(mol·M)).[31] This suggests that Tm-CspB has a sharper unfolding transition and hence a larger thermodynamic barrier[29] than its homologues that should show similar barriers. In agreement with this expectation, the predicted barriers for Bs-CspB and Bc-CspB at T_0 are identical (~19.5 and 19.6 kJ/mol) with that of Tm-CspB being higher (~26.8 kJ/mol). This correlation suggests that the difference in rates possibly arises from a more frustrated landscape[57] in thermophilic proteins. From the predicted barrier and experimental relaxation rate at the chemical denaturation midpoint, we calculate D_{eff} to be ~1/(324 μs) for Bc-CspB and ~1/(267 μs) for Tm-CspB that are much slower than for Bs-CspB and other independent experimental estimates at ~298 K.[54,55]

Modulation of Stability and Kinetics in the Engrailed Family by Surface Electrostatics. In the characterization of the HEWL/BLA and CspB families, we first fit the model to available experimental data and then use it to predict the equilibrium and kinetic behavior of the homologues. In this section and the next, we go further and do semiblind predictions (as we use the structural information and also adjust ξ to match the predicted T_m of one of the homologues to experiments) on both equilibrium and kinetics of the all-helical domains from the engrailed[30] and α-spectrin[32] families.

We could not use the published DSC data of engrailed homeodomain (En-HD[58]) as it is not reported in absolute units and possibly carries a low-temperature artifact.[24] We employ the same parameters from the fit to the Bs-CspB thermogram but slightly modify the interaction energy to match the experimental T_m of 325 K for En-HD. We obtain the following order of equilibrium stability, hRAP1 < En-HD < hTRF1 < c-Myb (Figure 4A), with the stability order of the latter three

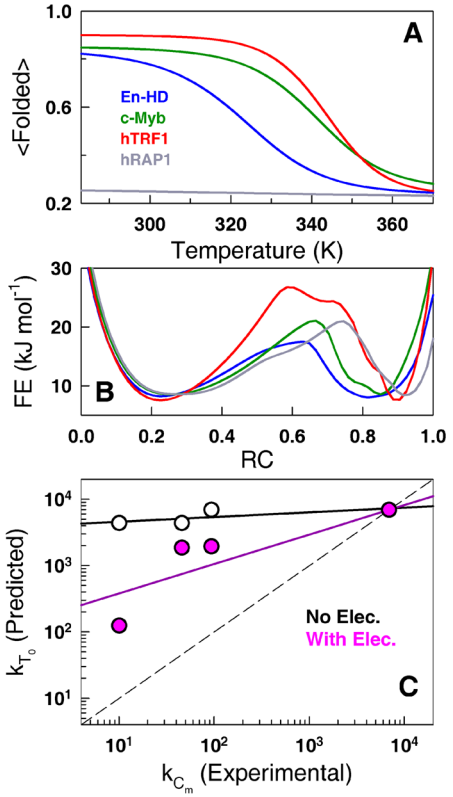

Figure 4. Modulation of stability and kinetics in the engrailed family by surface electrostatics. (A,B) Predicted unfolding curves and free-energy surfaces at T_0. (C) Correlation between experimental rates at chemical midpoint (C_m) and predicted rates at T_0 from the WSME model with (pink ●) and without (○) electrostatics. The dashed line corresponds to the expected 1:1 correlation.

homologues agreeing with experiments.[30] For hRAP1, the model predicts a mean folded probability of just ~0.25 at 298 K, indicating that the assumed stabilization energy is not sufficient to fold the protein (see Discussion).

In terms of kinetics, the engrailed family displays a remarkable 3 order difference in their midpoint relaxation rates.[30] In agreement with this, the predicted folding free-energy barriers from the model follow the exact same trend: En-HD (~9.3 kJ/mol) < c-Myb (12.5) < hRAP1 (12.4) < hTRF1 (19.2) (Figure 4B). For hRAP1, the barrier is calculated at 335 K after modifying ξ to obtain a T_m of 333 K (the mean T_m of single-domain proteins[43]). The barriers for c-Myb and hRAP1 are almost identical, agreeing with the very similar

dx.doi.org/10.1021/ct300676w | J. Chem. Theory Comput. 2012, 8, 4646−4656

experimental midpoint relaxation rates of 46 and 94 s^{-1}, respectively, thus providing a nice internal control to the model predictions. Again, the trend in barriers is reflected in the *m*-values: the homologue with the smallest barrier (En-HD) displays the lowest *m*-value and vice versa as shown before from a one-dimensional FES analysis.[29] Assuming a D_{eff} of 1/(7 μs), we obtain a good correlation between experimental and predicted rates and am able to account for an ∼2 order difference in the midpoint relaxation rates between homologues arising purely from added electrostatic terms into the model (Figure 4C).

The detail that stands out from a closer look at the unfolding probabilities (Figure 4A) and FES (Figure 4B) is that the "native state" of En-HD is just ∼80% folded at room temperature and the native basin is broad extending from 0.75 to 0.9 in nativeness units. In other words, the native state of En-HD is quite frustrated due to unfavorable electrostatic interactions and hence not completely folded even at 298 K. Experimental evidence for this comes from the progressively increasing stabilization of En-HD and its mutants with salt concentration up to a remarkable 3 M NaCl.[59] The folded minima in the FES of the other engrailed homologues are also shifted, centered ∼90% of structured residues. This is however not seen in small β-rich proteins (CspB) or larger domains (HEWL/BLA), which are >98% folded (Figures 2B, 3B). This is first evidence that there is a significant structural heterogeneity in the native states of small helical proteins even if the main folding barrier is large. The basic WSME model fails to predict the differences in stability or barriers (Figures S7 and 4C).

Evidence for Landscape Roughness in the α-Spectrin Family. The homologues R15, R16, and R17 are also helical domains but are ∼110 residues in length as compared to the ∼50-residue engrailed family. They have very similar stabilities in chemical denaturation experiments with large differences in their minimum relaxation rates. Specifically, the minimum relaxation rate of R15 is ∼80 times faster than R16 and ∼600 times faster than R17. This translates into larger relative thermodynamic barriers of ∼11 and ∼16 kJ/mol for R16 and R17, respectively. Kinetic experiments in the presence of viscogens however suggest that most of the difference in relaxation rates possibly arises from internal friction effects[32] (i.e., landscape roughness). This study did not explicitly account for any possible changes in thermodynamic barriers. To investigate further, we use the same set of parameters from the fit to the Bs-CspB heat capacity profile and adjust ξ such that the inflection point on the R15 unfolding curve is 333 K.

The model predicts R15 and R17 to have similar stabilities upon temperature denaturation ($\Delta T_m \approx 12$ K), while a much larger difference is predicted for R16 ($\Delta T_m \approx 30$ K) (Figure 5A) unlike what is seen in chemical denaturation experiments. The thermodynamic barriers, on the other hand, show a clear trend increasing from ∼34.2 kJ/mol for R15 to ∼38.5 and ∼40.3 kJ/mol for R16 and R17, respectively, in qualitative agreement with the relative order of experimental rates (Figure 5B). The basic WSME also predicts similar magnitude for the barriers but lacks a trend: 32.5, 40.8, and 30.5 kJ/mol for R15, R16, and R17, respectively (Figure 5B). The relative difference in predicted thermodynamic barriers from the model with electrostatics is however not large enough to account for the differences in relaxation rates, strengthening the previous interpretation that landscape roughness plays a big role in slowing the folding kinetics of R16 and R17. Still, a part of the rate difference (a factor of ∼6 and ∼12 between R15 and R16/R17) can simply be explained by larger thermodynamic barriers than invoking frustration.

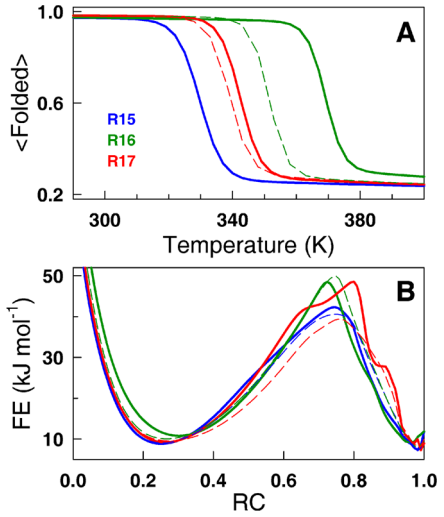

Figure 5. Evidence for landscape roughness in the α-spectrin family. (A,B) Predicted unfolding probabilities and free-energy surfaces at T_0 from the WSME model with (continuous curves) and without (dashed curves) electrostatics.

DISCUSSION

Model and Parameters. The energy function of the novel WSME model now resembles a mini-force-field with the addition of the electrostatics term and simplified treatment of solvation free energy. The model is characterized by just four thermodynamic parameters (ξ, ΔS_{conf}, ΔC_p^{cont}, ε_{eff}), which we objectively parametrize by analyzing equilibrium unfolding experiments on small single-domain proteins. Importantly, we use differential scanning calorimetry data as they report on the global protein unfolding thermodynamics and are related to the protein partition function.[20−22] we find that the availability of both kinetic relaxation rates and equilibrium heat capacity profiles serve as a stringent two-way control to constrain statistical models. The contact-map replacement procedure we develop is a simple but efficient technique as it helps to narrow the parameter space and serves as a reality check to the effectiveness of the model when characterizing homologues. In other words, upon the availability of DSC data for all homologues within the same set, one could in principle fit each of them individually and compare how the parameter sets are compatible with each other by replacing contact maps and predicting both equilibrium and kinetics.

In terms of parameters in the current analysis, the mean interaction energy ($\langle\xi\rangle$) from the fits (−52.1 ± 6.6 J/mol) is of the same order as the van der Waals interaction energy for a C−C interaction calculated from atomic-level force-field parameters:[60] −46.1 J/mol at 6 Å. The mean entropic cost of −11.4 J/(mol·K) per residue (from HEWL and CspB-Bs fits alone) is smaller than that estimated from calorimetry analysis of two-state proteins[43] (−16.5 J/(mol·K)); the lower value is consistent with the fact that the WSME model overcounts the number of possible configurations. The mean heat capacity per residue is just ∼22.2 J/(mol·K), much smaller than the 58 J/(mol·K) expected for larger two-state systems. As with the entropic cost, this might be a result of the microscopic nature of the WSME model, while the reference parameters are obtained

from a macroscopic two-state analysis. Despite the minor differences, the magnitude of thermodynamic parameters is within the physically reasonable range expected for small single-domain proteins.

Thermodynamics and Dynamics. The availability of high-quality DSC data on the HEWL/BLA proteins enabled us to obtain a physically reasonable estimate for ε_{eff}: in the act of reproducing the conformational stability of apo-BLA, we are able to capture the large change in thermodynamic barriers and conformational features in agreement with a variety of independent evidence that includes DSC experiments, NMR data, and predictions from the variable barrier model. We further validate the magnitude of ε_{eff} by reproducing the thermodynamic stabilities of the CspB family members, and in the case of engrailed family (for which we do not have a suitable reference point), reproduce the order of stabilities for three of the four homologues. It is important to note that the basic WSME model without electrostatics fails to reproduce the thermodynamic characteristics of all three families. Therefore, although a simple contact description is sufficient to quantitatively characterize experimental data (as in basic WSME model[13,19,27,61]), a more accurate deception of energetics is afforded by the model with the added electrostatic term.

We find that slower folding of apo-BLA as compared to HEWL despite the smaller thermodynamic barrier is due to the presence of several native-like partially structured states together with its electrostatically frustrated native state that impedes efficient folding by lowering the D_{eff} in agreement with previous experimental studies. In the CspB family, the model predicts that most of the differences in the midpoint relaxation rates at 298 K arise from differences in D_{eff} than from thermodynamic barriers. It is important to note that the experiments on CspB family have been performed using an ionic denaturant (guanidinium hydrochloride) that should screen all electrostatic interactions at high concentrations. Therefore, the source of increased frustration (or slower dynamics) in the thermophilic CspBs that have a larger content of charged residues and salt bridges could be the result of either an increase in the frequency non-native electrostatic interactions or a decrease in the frequency of favorable native-like electrostatic interactions during the folding process. This result is not surprising as thermophilic proteins are evolutionarily optimized for folding and function at temperatures typically greater than 325 K and not at 298 K. A direct consequence is that the folding speed limit might be slower for thermophilic proteins as compared to their mesophilic counterparts. This prediction is in agreement with the recent comparison of the ∼40 residue α-helical downhill folding homologous proteins from thermophilic (PDD, from *Bacillus stearothermophilus*) and mesophilic organisms (BBL, from *Escherichia coli*). Both proteins fold downhill under native-like conditions at 298 K, and hence the measured relaxation rates report on the folding diffusion coefficient, $\sim 1/(25\ \mu s)$ for BBL[54] and $\sim 1/(143\ \mu s)$ for PDD,[28] that interestingly differ by a factor of ∼5.7. Using this direct measure as a reference and accounting for the larger size of CspB translates to a $D_{eff} \approx 1/(228\ \mu s)$ for the 67-residue protein. This value is close to the predicted effective diffusion coefficients of $1/(324\ \mu s)$ for Bc-CspB and $1/(267\ \mu s)$ for Tm-CspB. More experiments comparing homologues across different organisms are needed to ascertain if this is indeed a general phenomenon and if the dynamic term matches at the respective growth temperatures. These predictions further open an avenue to quantitate the degrees of frustration arising from native- and non-native electrostatic interactions through not only comparing homologues but also by modifying the ionic nature of chaotropes (i.e., urea vs guanidinium).

In the engrailed family, there is a distinct trend in the model predictions consistent with experimental relaxation rates. Moreover, for two proteins within the engrailed family that have similar relaxation rates (c-Myb and hRAP1), we predict similar barriers highlighting the robustness of the calculation. The agreement is however not quantitative. For the spectrin family, there is experimental evidence pointing to a slower folding diffusion coefficient for R16 and R17 relative to R15 arising primarily from non-native interactions,[32] in tune with our predictions that the effect of increase in thermodynamic barriers is minimal. It is not as clear-cut for the engrailed homeodomain homologues as there are no available experimental data supporting either an increased thermodynamic barrier or slower diffusion coefficients for the three homologues relative to En-HD. Our results hint that it might be a combination of both as we reproduce a 2-order difference in relaxation rates by merely but explicitly including the differences in the distribution of charged residues between the homologues in the model.

An unexpected result of this study is that relative thermodynamic barriers are easier to predict than relative changes in stability (Figures 4 and 5). This observation is consistent with the fact that simple models[12,18,62] and descriptors[63,64] are able to reasonably predict the changes in folding rate constants while performing worse when it comes to conformational stability. Although protein size is an important factor determining protein stabilities,[65] it is clear from the current work that finer structural details such as the distribution of surface electrostatics may also play a dominant role driven by strong selection pressure to remain folded under high temperatures for functional reasons. In this aspect, the inability to reasonably reproduce the similar stabilities of spectrin domains and the stability order of hRAP1 (in the engrailed family) might just be a result of comparing chemical with thermal denaturation as the unfolding mechanisms are completely different.[66] Moreover, spectrin domains are rod-shaped as compared to the globular HEWL/BLA proteins used for parametrizing the effective dielectric constant thus possibly requiring an equivalent reference to estimate ε_{eff}. This is evident in the fact that the WSME model without electrostatics predicts similar stabilities for the three spectrin domains: range of $T_m \approx$ 40 and ∼20 K with and without electrostatics (Figure 5A). An alternative viewpoint is that the model is missing contributions from other energy terms. A good candidate is the hydrogen bond that plays a role in stabilizing not only the secondary structural elements but also turns and long loops (hRAP1 has longer serine-rich loops as compared to other members in the engrailed family) and whose effects are quite complex.[67] Another factor could be the intrinsic conformational preferences of residues to be in different secondary structural elements.[68] This can be introduced into the model in the form of residue identity and structure-dependent entropic costs. Work is currently underway to introduce these terms into the model.

Critique of the Model. The increased predictive ability of the current version of the model is the result of its makeup that directly incorporates two fundamental aspects of protein folding processes: (1) the large degrees of freedom intrinsic

to the polymeric protein chain are accounted for by an ensemble of 2^N microstates, and (2) the diversity of weak noncovalent energies are represented in an uncomplicated fashion by individual terms for short-range (van der Waals), long-range interactions (electrostatics), and hydrophobic effect (solvation free energy). In addition, the model can be solved exactly in just a matter of few seconds; that is, the contribution to the partition function from all of the 2^N microstates for a range of temperatures can be generated very quickly as compared to the months of simulation time required in all-atom representations of proteins. However, the ability to solve the model analytically and its inherent simplicity come at an expense as we discuss below.

First, the WSME model and its various flavors assume a particular mechanism of folding: that local interactions form earlier than nonlocal interactions with multiple nucleation sites distributed throughout the length of the protein, and for any two nonlocal residues to interact all of the intervening residues should also be folded (i.e., a string of 1's should connect any pair of interacting residues). Therefore, the model should be used with extreme care to analyze the folding data of proteins that have been experimentally shown to form nonlocal interactions separated by unfolded stretches. It is of interest to note that the successful COREX algorithm[69] is constructed on equivalent mechanistic principles (i.e., local interactions form first), while empirical analysis of mutational data,[70] Gō-model,[71] and large-scale MD simulations[72] also points to a similar mechanism of folding at work. These works underline the critical role of local interactions in folding and point to a possible reason as to why the WSME model is one of the most successful models of folding.

Second, in the spirit of mean-field Gō-like treatments, the model assumes that only native contacts (i.e., contacts identified from the PBD structure) contribute to the folding energetics. Although this is a drastic assumption, it is well within the framework of the energy-landscape theory,[57] the now well-accepted reference frame for interpreting experiments. Moreover, recent long molecular simulations on fast folding proteins suggest that the barrier top has native-like topology with only a few non-native interactions and that even unfolded states possess local structure similar to that found in the native states.[72] These observations hint that a simple Gō-like treatment should be sufficient, at a first approximation, to parametrize the energetics of protein folding processes. However, the situation is more complex when considering non-native electrostatic interactions, as they are inherently long range in nature. Both experiments[73] and simulations[8,37] point to the fact that electrostatic interactions are only weakly formed at the barrier top, suggesting that there is an ample scope for non-native effects arising from charged residues during the folding of a protein. In such cases, the WSME model with electrostatics can still be employed as any difference between experiments and model predictions (especially in kinetics) can be directly attributed to these effects. This is particularly relevant when comparing homologues as overfitting is avoided and complex solution effects are eliminated (e.g., the presence of salt or chemical denaturants). In other words, the drawback can be turned into an advantage to quantify the effect of non-native interactions (as we show in the case of CspB and α-spectrin families). As a note of caution pertaining to the current work, we would like to add that the ε_{eff} of 29 is parametrized to treat only protein surface electrostatics using the popular Debye–Hückel theory and not buried charges with significantly perturbed pK_a that are inherently more challenging to model.

Last, any errors or differences in the structural calculation protocol can influence the results as the model is intricately linked to the native PDB structure (illustrated using the Tm-CspB in the current study). One possible solution to this problem is to compare mutants of the same protein to look for trends or to parametrize the model for any additional energy terms. Moreover, parametrizing using the native structure is expected to work well for two-state-like proteins as the PDB structure is a truer representation of the native state than downhill folders that are typically characterized by a heterogeneous native ensemble.[74]

To summarize, the WSME model should be seen as a simple and quick, yet physical, predictive and analysis tool with details that can be extracted at the residue level. The predictions can then be tested either using atomic-level simulations or more directly by experiments.

CONCLUSIONS

By explicitly including an electrostatic potential term in the Ising-like WSME model and by characterizing purely the equilibrium experimental data on homologous proteins, we identify several different features of folding landscapes that are influenced by the distribution of charged residues on the protein surface. These include the effect of electrostatics on stability, barrier heights, width of the native-state ensemble, and the ruggedness of the folding landscape in two-state-like proteins. The order of stability and folding relaxation rates agree more with experiments only upon the introduction of the electrostatic term into the model. The predicted free-energy surfaces are quite complex, particularly for HEWL/BLA and CspB family, despite displaying a single-peaked DSC thermogram and a sigmoidal equilibrium unfolding curve, highlighting the intricate nature of the folding process even in two-state-like proteins. However, a definitive answer to the precise thermodynamic and dynamic contributions to the relaxation rates in the CspB, engrailed, α-spectrin, and other families relies on constant feedback from experiments and particularly by performing calorimetry measurements. Characterizing such data in combination with relaxation rates employing the model would provide further tests to its predictive ability while simultaneously bringing the worlds of experiments, theory, and simulations closer.

ASSOCIATED CONTENT

Supporting Information

Results from the structural database analysis, predictions from the WSME model with electrostatics on the conformational behavior of BBL, predictions from the basic WSME model with the simplified treatment of solvation for HEWL/BLA, engrailed, and CspB family including the unfolding curves/DSC and the free-energy surfaces at T_0, the single-sequence free-energy landscape of HEWL and apo-BLA, single-exponential fits to the predicted kinetic decays of CspB, and a table with the model parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: athi@iitm.ac.in.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Jose M. Sanchez-Ruiz and Beatriz Ibarra-Molero for sharing the experimental heat capacity profiles of HEWL and BLA, Pierpaolo Bruscolini for valuable comments and critical reading of the manuscript, and S. Karthigayini for help in generating the nonredundant protein structural database.

■ REFERENCES

(1) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. *Angew. Chem., Int. Ed.* **2005**, *44*, 7342−7372.
(2) Szilagyi, A.; Zavodszky, P. *Structure* **2000**, *8*, 493−504.
(3) van den Burg, B.; Eijsink, V. G. H. *Curr. Opin. Biotechnol.* **2002**, *13*, 333−337.
(4) Makhatadze, G. I.; Loladze, V. V.; Gribenko, A. V.; Lopez, M. M. *J. Mol. Biol.* **2004**, *336*, 929−942.
(5) Strickler, S. S.; Gribenko, A. V.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.; Loladze, V. V.; Makhatadze, G. I. *Biochemistry* **2006**, *45*, 2761−2766.
(6) Bi, Y.; Cho, J. H.; Kim, E. Y.; Shan, B.; Schindelin, H.; Raleigh, D. P. *Biochemistry* **2007**, *46*, 7497−7505.
(7) Halskau, O.; Perez-Jimenez, R.; Ibarra-Molero, B.; Underhaug, J.; Muñoz, V.; Martinez, A.; Sanchez-Ruiz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8625−8630.
(8) Azia, A.; Levy, Y. *J. Mol. Biol.* **2009**, *393*, 527−542.
(9) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. *J. Am. Chem. Soc.* **2012**, *134*, 979−987.
(10) Liu, Z. X.; Reddy, G.; O'Brien, E. P.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 7787−7792.
(11) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199.
(12) Muñoz, V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11311−11316.
(13) Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Muñoz, V. *Science* **2002**, *298*, 2191−2195.
(14) Kubelka, J.; Henry, E. R.; Cellmer, T.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18655−18662.
(15) Wako, H.; Saito, N. *J. Phys. Soc. Jpn.* **1978**, *44*, 1931−1938.
(16) Wako, H.; Saito, N. *J. Phys. Soc. Jpn.* **1978**, *44*, 1939−1945.
(17) Bruscolini, P.; Pelizzola, A. *Phys. Rev. Lett.* **2002**, *88*, 258101.
(18) Henry, E. R.; Eaton, W. A. *Chem. Phys.* **2004**, *307*, 163−185.
(19) Bruscolini, P.; Naganathan, A. N. *J. Am. Chem. Soc.* **2011**, *133*, 5372−5379.
(20) Cooper, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2740−2741.
(21) Freire, E.; Biltonen, R. L. *Biopolymers* **1978**, *17*, 463−479.
(22) Chan, H. S. *Proteins* **2000**, *40*, 543−571.
(23) Muñoz, V.; Sanchez-Ruiz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17646−17651.
(24) Naganathan, A. N.; Sanchez-Ruiz, J. M.; Muñoz, V. *J. Am. Chem. Soc.* **2005**, *127*, 17970−17971.
(25) Naganathan, A. N.; Perez-Jimenez, R.; Sanchez-Ruiz, J. M.; Muñoz, V. *Biochemistry* **2005**, *44*, 7435−7449.
(26) Fung, A.; Li, P.; Godoy-Ruiz, R.; Sanchez-Ruiz, J. M.; Muñoz, V. *J. Am. Chem. Soc.* **2008**, *130*, 7489−7495.
(27) Godoy-Ruiz, R.; Henry, E. R.; Kubelka, J.; Hofrichter, J.; Munoz, V.; Sanchez-Ruiz, J. M.; Eaton, W. A. *J. Phys. Chem. B* **2008**, *112*, 5938−5949.
(28) Naganathan, A. N.; Li, P.; Perez-Jimenez, R.; Sanchez-Ruiz, J. M.; Muñoz, V. *J. Am. Chem. Soc.* **2010**, *132*, 11183−11190.
(29) Naganathan, A. N.; Doshi, U.; Muñoz, V. *J. Am. Chem. Soc.* **2007**, *129*, 5673−5682.
(30) Gianni, S.; Guydosh, N. R.; Khan, F.; Caldas, T. D.; Mayor, U.; White, G. W. N.; DeMarco, M. L.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13286−13291.
(31) Perl, D.; Welker, C.; Schindler, T.; Schroder, K.; Marahiel, M. A.; Jaenicke, R.; Schmid, F. X. *Nat. Struct. Biol.* **1998**, *5*, 229−235.
(32) Wensley, B. G.; Batey, S.; Bone, F. A. C.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J. *Nature* **2010**, *463*, 685−689.
(33) Taketomi, H.; Ueda, Y.; Go, N. *Int. J. Pept. Protein Res.* **1975**, *7*, 445−459.
(34) Gomez, J.; Hilser, V. J.; Xie, D.; Freire, E. *Proteins* **1995**, *22*, 404−412.
(35) Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*; Springer: New York, 2000.
(36) Kim, Y. C.; Hummer, G. *J. Mol. Biol.* **2008**, *375*, 1416−1433.
(37) Zarrine-Afsar, A.; Zhang, Z. Q.; Schweiker, K. L.; Makhatadze, G. I.; Davidson, A. R.; Chan, H. S. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 858−870.
(38) Beard, D. A.; Schlick, T. *Biopolymers* **2001**, *58*, 106−115.
(39) Hyeon, C.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6789−6794.
(40) Levy, Y.; Onuchic, J. N.; Wolynes, P. G. *J. Am. Chem. Soc.* **2007**, *129*, 738−739.
(41) Savelyev, A.; Papoian, G. A. *J. Am. Chem. Soc.* **2007**, *129*, 6060−6061.
(42) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138−2148.
(43) Robertson, A. D.; Murphy, K. P. *Chem. Rev.* **1997**, *97*, 1251−1267.
(44) Vicatos, S.; Roca, M.; Warshel, A. *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 670−684.
(45) Desai, T. M.; Cerminara, M.; Sadqi, M.; Muñoz, V. *J. Biol. Chem.* **2010**, *285*, 34549−34556.
(46) Akmal, A.; Muñoz, V. *Proteins* **2004**, *57*, 142−152.
(47) Ikeguchi, M.; Kuwajima, K.; Mitani, M.; Sugai, S. *Biochemistry* **1986**, *25*, 6965−6972.
(48) Halskau, O.; Underhaug, J.; Froystein, N. A.; Martinez, A. *J. Mol. Biol.* **2005**, *349*, 1072−1086.
(49) Schlepckow, K.; Wirmer, J.; Bachmann, A.; Kiefhaber, T.; Schwalbe, H. *J. Mol. Biol.* **2008**, *378*, 686−698.
(50) Griko, Y. V.; Freire, E.; Privalov, P. L. *Biochemistry* **1994**, *33*, 1889−1899.
(51) Kumar, S.; Tsai, C. J.; Nussinov, R. *Protein Eng.* **2000**, *13*, 179−191.
(52) Wassenberg, D.; Welker, C.; Jaenicke, R. *J. Mol. Biol.* **1999**, *289*, 187−193.
(53) Huang, L.; Shakhnovich, E. I. *Protein Sci.* **2012**, *21*, 677−685.
(54) Li, P.; Oliva, F. Y.; Naganathan, A. N.; Muñoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 103−108.
(55) DeCamp, S. J.; Naganathan, A. N.; Waldauer, S. A.; Bakajin, O.; Lapidus, L. J. *Biophys. J.* **2009**, *97*, 1772−1777.
(56) Lapidus, L. J.; Steinbach, P. J.; Eaton, W. A.; Szabo, A.; Hofrichter, J. *J. Phys. Chem. B* **2002**, *106*, 11628−11640.
(57) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167−195.
(58) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518−13522.
(59) Religa, T. L.; Markson, J. S.; Mayor, U.; Freund, S. M. V.; Fersht, A. R. *Nature* **2005**, *437*, 1053−1056.
(60) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(61) Naganathan, A. N.; Perez-Jimenez, R.; Muñoz, V.; Sanchez-Ruiz, J. M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17064−17076.
(62) De Sancho, D.; Muñoz, V. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17030−17043.
(63) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *227*, 985−994.
(64) Gromiha, M. M.; Selvaraj, S. *J. Mol. Biol.* **2001**, *310*, 27−32.
(65) de Sancho, D.; Doshi, U.; Muñoz, V. *J. Am. Chem. Soc.* **2009**, *131*, 2074−2075.
(66) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. *J. Am. Chem. Soc.* **2007**, *129*, 7346−7353.
(67) Enciso, M.; Rey, A. *Biophys. J.* **2011**, *101*, 1474−1482.

(68) Muñoz, V.; Serrano, L. *Proteins* **1994**, *20*, 301−311.
(69) Hilser, V. J.; Freire, E. *J. Mol. Biol.* **1996**, *262*, 756−772.
(70) Naganathan, A. N.; Muñoz, V. *Proc. Natl. Acad. Sci. U.S.A* **2010**, *107*, 8611−8616.
(71) Naganathan, A. N.; Orozco, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 15166−15174.
(72) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.
(73) Luisi, D. L.; Raleigh, D. P. *J. Mol. Biol.* **2000**, *299*, 1091−1100.
(74) Naganathan, A. N.; Orozco, M. *J. Am. Chem. Soc.* **2011**, *133*, 12154−12161.