Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles

Patrick Weinkam[a], Ekaterina V. Pletneva[b,1], Harry B. Gray[b,2], Jay R. Winkler[b,2], and Peter G. Wolynes[a,2]

[a]Center for Theoretical Biological Physics and Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093; and [b]Beckman Institute, California Institute of Technology, Pasadena, CA 91125

Contributed by Harry B. Gray, December 23, 2008 (sent for review December 3, 2008)

The denatured state of proteins is heterogeneous and susceptible to general hydrophobic and electrostatic forces, but to what extent does the funneled nature of protein energy landscapes play a role in the unfolded ensemble? We simulate the denatured ensemble of cytochrome *c* using a series of models. The models pinpoint the efficacy of incorporating energetic funnels toward the native state in contrast with models having no native structure-seeking tendency. These models also contain varying strengths of electrostatic effects and hydrophobic collapse. The simulations based on these models are compared with experimental distributions for the distances between a fluorescent donor and the heme acceptor that were extracted from time-resolved fluorescence energy transfer experiments on cytochrome *c*. Comparing simulations to detailed experimental data on several labeling sites allows us to quantify the dominant forces in denatured protein ensembles.

denatured state | funnel | hydrophobic collapse | structure-based

The guiding forces of the folding funnel dominate the folding mechanism but are also of key importance in the denatured state. The denatured state is, however, strongly fluctuating and therefore also highly susceptible to weak nonnative electrostatic and hydrophobic interactions. Measurements of the radius of gyration (R_g) suggest that proteins in unfolded states are nearly random coils (1) perhaps modulated by the effects of electrostatic interactions (2–8) or general hydrophobicity (9, 10). Yet a number of techniques, including circular dichroism and NMR, indicate that there can be local helical structure (11–14) and sometimes remnant tertiary structure in chemically denatured ensembles (15). These discrepant views of the denatured ensemble arise from biases inherent to techniques that measure different average properties. Indeed, both experimental (16–23) and theoretical (2, 9, 24) work supports the existence of a multiplicity of populations of unfolded configurations, both compact and extended, in denatured ensembles. Analysis of time-resolved fluorescence energy transfer (TRFET) measurements can reveal composition and structural features of heterogeneous protein ensembles. These experiments have provided distance distributions ($P(r)$) from dansyl fluorophores to the heme in unfolded *iso*-1-ferricytochrome *c* (21–23). In this article we show that models based on funneled (structure-based) energy landscapes (25–30) but additionally incorporating generic hydrophobic collapse and electrostatic interactions lead to heme-fluorophore distance distributions that, in the main, compare quite well with the experimental results. The models correctly predict the effects of salt and pH on the denatured ensemble of cytochrome *c* (22, 31) and shed new light on the interplay between native structure seeking funneled interactions and electrostatics.

Results

Experimental Data. We have examined TRFET from a dansyl fluorophore (Dns) to the Fe(III) heme in six variants of *Saccharomyces cerevisae iso*-1-cytochrome *c* under GuHCl- and acid-denaturing conditions (21–23). Labeling sites at residues 4, 39, 50, 66, 85, and 99 sample various regions in the protein, permitting comprehensive analysis of the denatured states. The observed nonexponential kinetics reflect distributions of donor (D)–acceptor (A) distances characteristic of the heterogeneous ensembles. The probability distributions of logarithmically spaced rate constants, $P(k)$, were extracted directly from experimental data I(t) (Eq. **1**) and then converted to the probability distributions of a range of distances $P(r)$ (Eq. **2**). The R_0 value of 39 Å for the dansyl–heme pair limited the D-A distances that can be measured. At longer distances (> 1.5 R_0), the energy transfer efficiency is extremely low and nearly identical rate constants are observed despite the different Dns-heme separation distances. Therefore, single bars at 59 Å in the experimental $P(r)$ distributions represent the structures with $r \geq 59$ Å.

$$I(t) = \sum_{k=k_0}^{\infty} P(k)e^{-kt} \quad [1]$$

$$k = k_0\left(1 + \left(\frac{R_0}{r}\right)^6\right) \quad [2]$$

Our TRFET data have revealed considerable diversity in both denatured ensembles: there are several distinct populations of different structures, some extended and others compact. A significant population of structures with relatively short D-A distances even under strongly denaturing conditions is evidence for favorable interactions within unfolded polypeptides. Comparison of experimental $P(r)$ distributions for the GuHCl- and the acid-denatured protein suggests that the latter polypeptide ensemble is more compact.

In our simulations we have tracked distances between the heme iron and the C^{α} atom of each of the six labeled residues. To account for the length of the Dns linker in the experimental data, 6 Å were added to these distances. As was done for TRFET data, all of the structures with $r \geq 59$ Å were placed into a single bin. The fluorescence decay curves were calculated from theoretical distributions $P(r)$ according to Eq. **1**.

The GuHCl-Denatured Ensemble. We examined several models that include varying tendencies for nonnative hydrophobic collapse and various strengths of electrostatic forces in addition to forces that are energetically funneled toward the native state (29–30). The potentials were designed such that increasing the magnitude of the electrostatic contribution causes the protein to expand, thereby decreasing the effects of hydrophobic collapse. Electrostatic effects are included using a Coulomb potential with a Debye screening term to account for salt effects (32–33). The electrostatic contribution is scaled by adjusting the parameter n_{ε} (see *Materials and Methods* for more information). Simulations using the simple

Author contributions: P.W. and P.G.W. designed research; P.W. performed research; P.W., E.V.P., H.B.G., J.R.W., and P.G.W. contributed new reagents/analytic tools; P.W., E.V.P., H.B.G., J.R.W., and P.G.W. analyzed data; and P.W., E.V.P., H.B.G., J.R.W., and P.G.W. wrote the paper.

The authors declare no conflict of interest.

[1]Present address: Department of Chemistry, Dartmouth College, 6128 Burke Laboratory, Hanover, NH 03755.

[2]To whom correspondence may be addressed. E-mail: hbgray@caltech.edu, winklerj@caltech.edu or pwolynes@ucsd.edu.

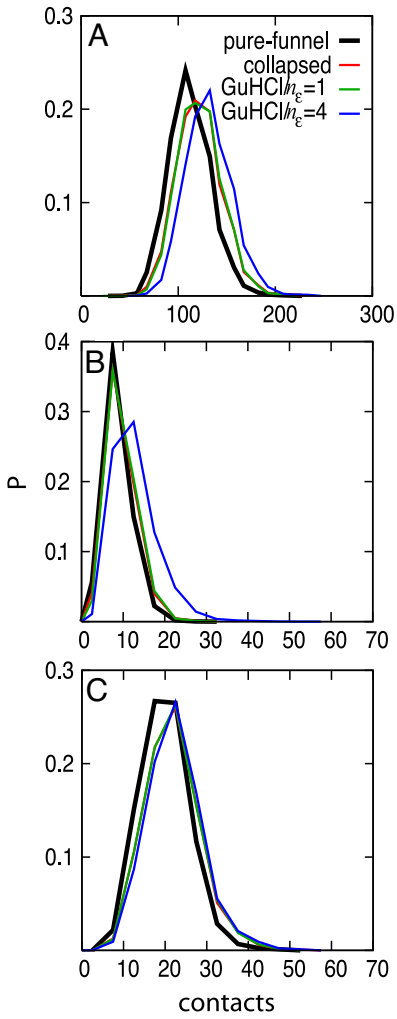

Fig. 1. Probability distributions for (*A*) the number of total contacts, (*B*) the number of charge-charge contacts, and (*C*) the number of hydrophobic-hydrophobic contacts. Contacts are defined between those C^{α}s that are >2 residues apart and that approach closer than 10 Å.

pure-funnel (structure-based) model yield fewer contacts than do models based on funneled landscapes that also include generic collapse (Fig. 1*A*). A further increase in the number of contacts is seen upon increasing the magnitude of the electrostatic term by changing n_ε from 1 to 4. The increase in the total number of contacts largely reflects an increased number of charge-charge contacts (Fig. 1*B*). There is no significant change in the number of hydrophobic-hydrophobic contacts due to varying n_ε (Fig. 1*C*). Despite the corresponding increase in the number of contacts with increased electrostatic contribution, all of the simulations predict a similar radius of gyration (27 ± 3 Å). This value agrees with the value extracted from small-angle X-ray scattering experiments (19).

Detailed distance distributions extracted from TRFET data clearly demonstrate the heterogeneity of unfolded ensembles (21–23). Although the pure-funnel model correctly predicts the presence of more extended structures, it fails to predict many of the partially collapsed states. The effects of nonnative hydrophobic collapse and electrostatics are needed to generate the heterogeneous unfolded ensemble observed experimentally. Fig. 2 shows $P(r)$ distributions calculated from a simulation that qualitatively agrees with the experimental data. The fidelity of the simulation to observation is more obvious when the fluorescence decay data are used directly for the comparison than when the inferred distance distributions are compared (Fig. 2).

We introduce an order parameter, Q_{charge}, which is a distance similarity measure restricted only to native charge–charge contacts less than 10 Å. The probability distribution of Q_{charge} suggests that the electrostatic contributions help steer the denatured ensemble to the native state (Fig. 3). In contrast, generic hydrophobic collapse alone does not increase the similarity of the unfolded ensemble to the native state. However, the electrostatic forces primarily help order the unfolded ensemble for short-range interactions ($3 \leq i-j < 5$) but play little role for medium- ($5 \leq i-j < 12$) and long-range ($i-j \geq 12$) interactions. It appears that local segments of the protein chain are partially guided by electrostatic forces toward a native-like conformation.

Histidine Misligation in the GuHCl-Denatured Ensemble. In GuHCl solutions at pH 7, there are three histidine residues in *iso*-1-ferricytochrome *c* (His-26, His-33, and His-39) that can bind the heme, replacing the native Met-80 ligand (34). The addition of imidazole inhibits this misligation. We find that most of the heme-residue distance distributions are quite similar for unfolded ensembles with and without misligation (Fig. 2). However, the experimental $P(r)$ distributions between the heme and residue 39 and residue 50 are sensitive to heme ligation and can be compared with simulation. We carried out independent simulations of the three distinct histidine-misligated proteins by introducing a harmonic potential between each of the histidines and the heme. Fig. 4 shows the ratio of the experimental $P(r)$ distributions including misligation effects to the distributions computed without misligation and the $P(r)$ distributions computed for each of the simulated histidine-misligated structures. Experimentally, the misligated unfolded ensemble has an increased probability for short distances between the heme and residue 39 compared with the nonmisligated unfolded ensemble (Fig. 4*A*). These distances (19 and 24 Å) are most consistent with the maximally probable distance predicted for misligation by His-33 but are less consistent with misligation by His-26. The heme to residue 50 distributions primarily support the existence of misligation by His-39 but do not rule out misligation by His-33. These results are consistent with folding studies of the structurally similar horse heart cytochrome *c* (a protein that lacks His-39) that have been interpreted in terms of His-33-misligation (35, 36). Moreover, studies of *iso*-1-ferricytochrome *c* mutants with single histidine residues suggest that His-26, His-33, and His-39 have comparable heme binding affinities at pH 7 (37, 38). It is likely, then, that the increased population of shorter heme-Dns distances in the presence of misligation (pH 7) can be attributed primarily to binding of His-33 and His-39 to the ferriheme.

Acid-Induced Denatured Ensemble. The ferricytochrome *c* energy landscape changes significantly under acidic conditions as the net positive charge on the protein increases from +8 to +23. The TRFET data of the acid-induced unfolded protein (Fig. 5) show that the polypeptide ensemble is composed of a more diverse range of structures under these conditions than found in the GuHCl-induced unfolded state (22). We simulate the salt effects in the acid-induced unfolded ensemble by varying the Debye screening factor. Under high salt conditions (low Debye screening factor) the simulations yield a highly collapsed protein, or molten globule, with $P(r)$ distributions whose peak is close to the native separation. Similar to experiments, simulations show that there is a small population of extended structures in the molten globule ensemble. At low salt concentrations, charge repulsion increases the variety of structures found, as both collapsed and extended conformations become populated. The heme to residue distributions of the simulated acid-induced unfolded ensembles are qualitatively similar to those inferred from experiment (Fig. 5) and the agreement is also apparent in the directly measured fluorescence decay curves (Fig. 5).

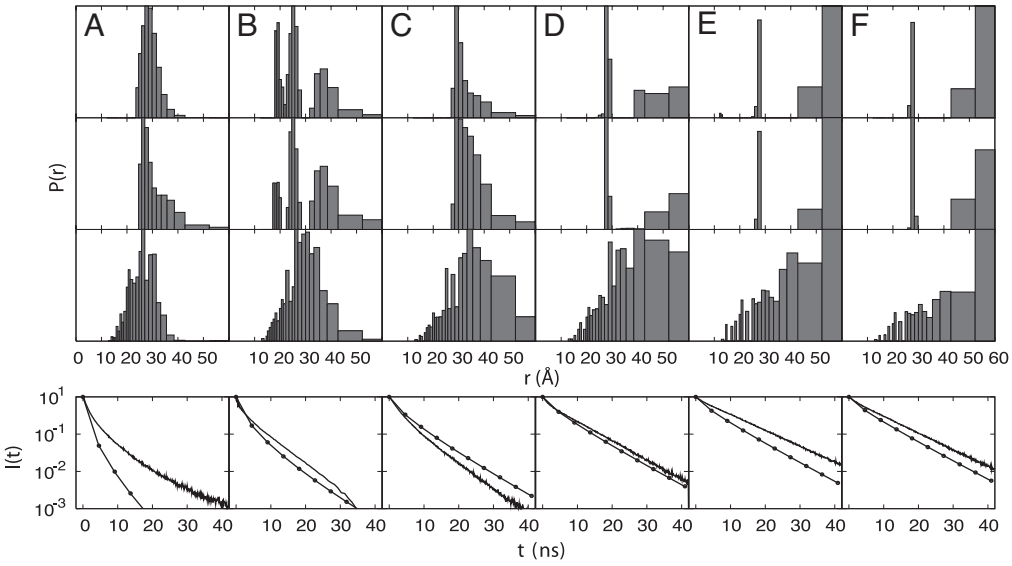

Fig. 2. Heme–residue distance probability distributions and the corresponding fluorescence decay curves for the GuHCl-unfolded ensemble involving residues (*A*) 4, (*B*) 39, (*C*) 50, (*D*) 66, (*E*) 85, and (*F*) 99. From top to bottom are: the distributions extracted from TRFET measurements under conditions favoring heme misligation in 3.0 M GuHCl at pH 7 (23), the distributions extracted from TRFET measurements without misligation in 3.0 M GuHCl and 0.15 M imidazole at pH 7 (23), the distributions calculated from simulations based on the GuHCl model with $n_\varepsilon = 4$, and the fluorescence decay curves. The experimental decay curves obtained at 3.0 M GuHCl without misligation are shown in solid lines and the decay curves calculated using the GuHCl model with $n_\varepsilon = 4$ are shown in lines with dots. See *Materials and Methods* for more information.

Discussion

Nonspecific electrostatic effects, hydrophobic collapse, and funneled interactions on the energy landscape synergistically improve the agreement of the $P(r)$ distributions calculated from the simulation with those extracted from TRFET experiments. The effect of adding structure-based potentials to the *tabula rasa* model is especially clear when comparing some specific distance distributions, such as that for the heme to residue 39 pair. The shorter distance ranges (<30 Å) are not populated to the extent observed in TRFET experiments if the funneling interactions of the folding landscape are not included (Fig. 6). A quantitative assessment of the agreement for all six $P(r)$ distributions is shown using an information theoretic surprisal measure. The surprisal averaged over each heme-residue pair, but restricted to pairs having distances <30 Å, makes clear that funneled energetic terms are necessary to generate appropriately collapsed structures. The surprisal also shows that models based on funneled energy landscapes agree more closely with experiment than do models based on nonfunneled, random polymer chains (Fig. 6 *A* and *B*). Furthermore, the models that lack a collapse potential score better for residues in the C-terminal end, particularly the heme to residue 85 distribution. At pH 7, the C-terminal sequence has a high density of polar and charged residues that would not drive hydrophobic collapse. Indeed, increasing the electrostatic contribution, by increasing n_ε, causes the C-terminal end to fray. In contrast, the surprisal measure for the distance of heme to residue 50, which is in a more hydrophobic region of the sequence, decreases for any model in which there is an increase in hydrophobic collapse. Inhomogeneously distributed regions of polar/hydrophobic and positively/negatively charged residues are responsible for the observed heterogeneity of the GuHCl-unfolded ensemble.

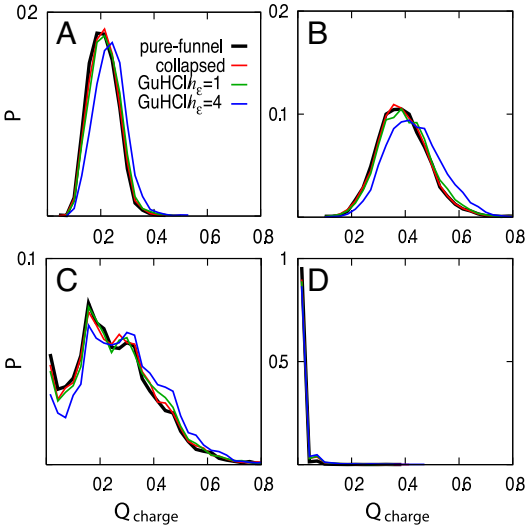

Fig. 3. Probability distributions of Q_{charge} for (*A*) all native contacts, (*B*) short-range native contacts ($3 \leq i-j < 5$), (*C*) medium-range native contacts ($5 \leq i-j < 12$), and (*D*) long-range native contacts ($i-j \geq 12$). Q_{charge} is a distance similarity measure of charge-charge contacts that are less than 10 Å in the crystal structure analogous to the folding reaction coordinate Q.

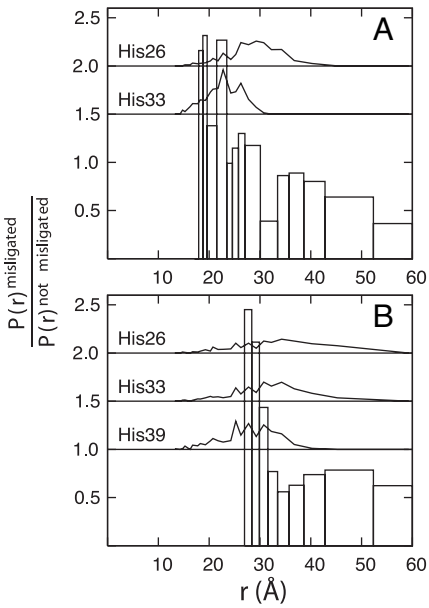

Fig. 4. Ratios of the $P(r)$ distributions extracted from TRFET measurements with misligation to those without misligation are shown by using open bars for (*A*) the heme to residue 39 pair and (*B*) the heme to residue 50 pair. The heme-residue $P(r)$ distributions from simulations in which a histidine is permanently misligated to the heme are shown within the plot using black lines. No $P(r)$ distributions for His-39-misligated structures are shown for the heme to residue 39 pair because cysteine mutation and Dns labeling prevent misligation.

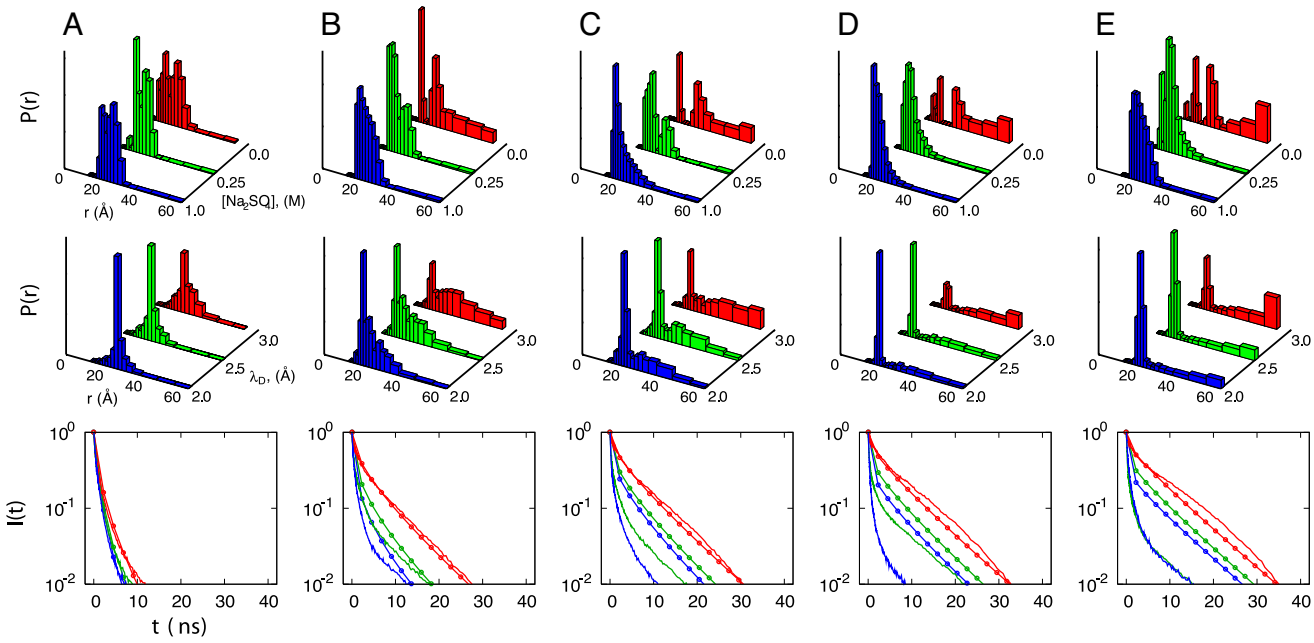

Fig. 5. Heme–residue distance probability distributions and the corresponding fluorescence decay curves for the acid-induced unfolded ensembles for labeling sites at: (*A*) 4, (*B*) 39, (*C*) 50, (*D*) 66, and (*E*) 99. From top to bottom are: the distributions extracted from TRFET measurements at pH 2 and different Na_2SO_4 concentrations (22), the distributions calculated from simulations based on the acid/salt model with $n_\varepsilon = 4$ using different Debye screening factors (Å), and the corresponding fluorescence decay curves. The colors of the decay curves correspond to the colors used in the above *P*(*r*) distributions with experimental data shown in solid lines and simulation results shown in lines with dots. See *Materials and Methods* for more information.

The acid-induced unfolded ensemble exhibits a broader range of structures than the GuHCl-denatured ensemble, suggesting differences in its energy landscape. The driving force for the formation of acid-induced unfolded structures is the chemical frustration caused by increasing the net charge on the protein as a result of protonation. Extended structures arise from like charge repulsion,

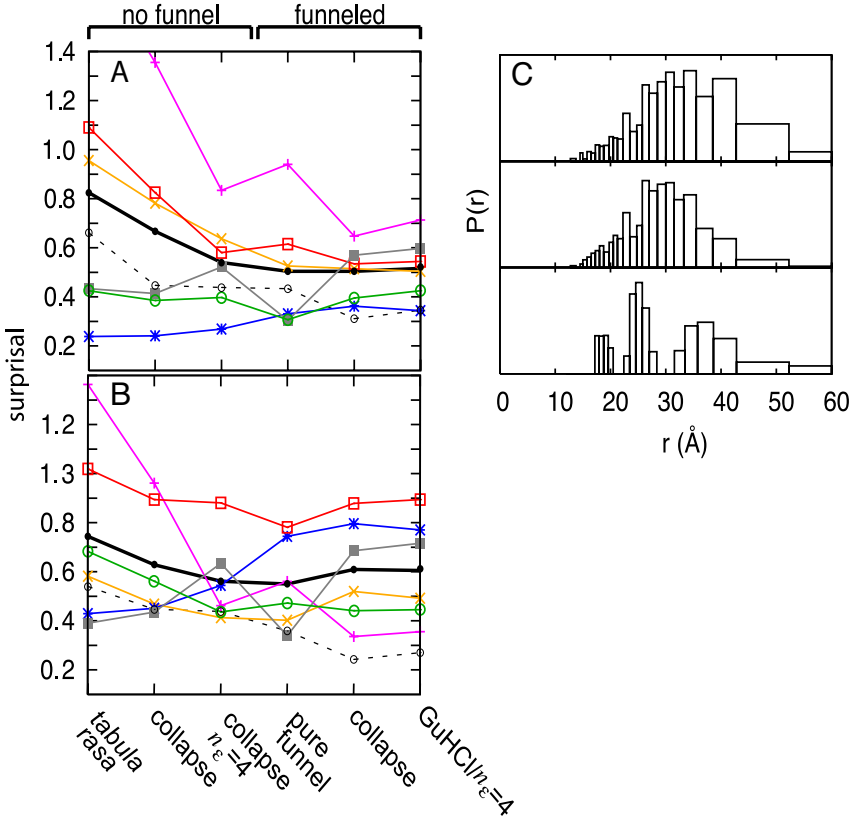

Fig. 6. A surprisal measure is used to compare the results of several models to the distributions determined from TRFET measurements (*A*) with misligation and (*B*) without misligation. Surprisal is a measure of the similarity between two distributions, in which small values imply agreement. The surprisal is plotted as lines for each pair distribution involving residues: 4 (*), 39 (×), 50 (+), 66 (□), 85 (■), and 99 (○). The average surprisals of all heme-residue pairs are plotted as black lines. The average surprisals for all pairs, at distances <30 Å, are plotted as dashed lines. See *Materials and Methods* for further description of the models and of the surprisal calculation. (*C*) Heme to residue 39 *P*(*r*) distributions calculated from simulations based on the nonfunneled, collapse/$n_\varepsilon = 4$ model (*Top*), the funneled GuHCl/$n_\varepsilon = 4$ model (*Middle*) and the distribution extracted from TRFET measurements without misligation (*Bottom*).

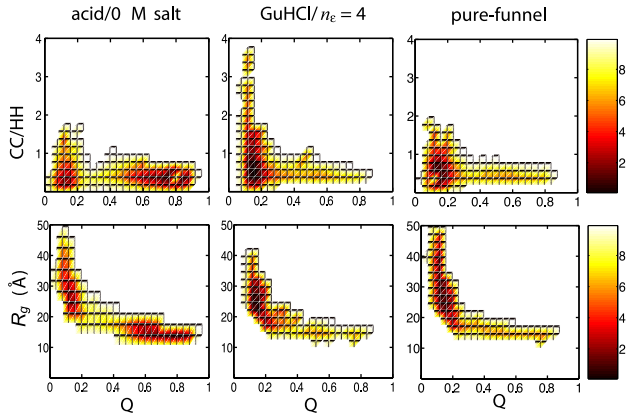

Fig. 7. Free energy surfaces in units of k_BT calculated from simulations based on the acid/0 M salt, GuHCl with $n_\varepsilon = 4$, and pure-funnel models. The free energy surfaces are plotted as a function of the folding reaction coordinate Q (related to the fraction of native contacts)and either the ratio of the number of charge-charge contacts (CC) to hydrophobic-hydrophobic contacts (HH) (*Upper*) or the radius of gyration (*Lower*).

which causes frustration and opposes the funneled nature of the energy landscape. The differences in unfolded ensembles can be visualized by using free energy surfaces as a function of the folding reaction coordinate Q and other structural measures as seen in Fig. 7. Q is a distance similarity measure that is related to the fraction of native contacts and is close to 1 if the topology of the configuration is similar to that of the native state. The reaction coordinate Q has been used in several studies (25, 27), including one on cytochrome *c* (29) in which a model with a perfectly funneled energy landscape was shown to predict the same sequential ordering of folding units as inferred by hydrogen exchange (39). The unfolded structures ($Q \sim 0.1$) generated using each model are extended and dispersed, as evidenced by their radii of gyration (Fig. 7), and they contract to a near native size at the folding transition state ($Q \sim 0.3$). Consistent with small-angle X-ray scattering data, the simulated GuHCl-denatured ensemble is composed of extended structures ($R_g > 20$ Å), whereas the simulated acid-unfolded ensemble contains both collapsed and extended structures (R_g between 15 and 40 Å) (19, 20). The more compact structures in the acid-unfolded ensemble can have a high fraction of native contacts with Q ranging between 0.5 and 0.9.

The ratio of charge–charge to hydrophobic–hydrophobic contacts shows that the residual structure in the unfolded ensemble generated from the pure-funnel model is mainly due to hydrophobic contacts (Fig. 7). By including electrostatic contributions, the unfolded ensemble reorganizes, owing to an increased number of attractive charge-charge interactions (Fig. 1). Electrostatics affect only local segments of the protein chain because R_g does not change significantly. These electrostatic interactions steer the protein chain as it collapses and they may even drive the preliminary stages of folding as evidenced by Q_{charge} (Fig. 3). However, because of repulsive electrostatic interactions, the ratio of charge-charge to hydrophobic-hydrophobic contacts is lower in the acid-unfolded ensemble than in the GuHCl-denatured ensemble.

We have demonstrated that there are specific patterns of structures in chemically denatured ensembles of cytochrome *c*. The heterogeneous distribution of hydrophobic and polar/charged residues results in nonrandom coils with varying amounts of collapse or extension in different regions of the sequence. Denatured ensembles have pockets of local, native-like structure, often guided by electrostatics, but these ensembles are globally unstructured above the folding temperature or at GuHCl concentrations greater than the unfolding midpoint. The destabilizing effects of GuHCl cause the protein to populate higher energy regions of the folding funnel; however, stabilizing nonnative interactions due to misligation can cause rather deep traps (Fig. 8 *A* and *B*). Below the folding midpoint or temperature, acidic conditions at low salt concentration can be used to induce unfolding. The structures in the acid-unfolded ensemble are diverse because the funneled energy landscape responsible for folding is opposed by chemically frustrating electrostatic effects from protonation that cause unfolding (Fig. 8). The success of these largely structure-based models in predicting the properties of chemically denatured ensembles suggests that features of a funneled energy landscape are still important even in the unfolded state. The interplay of nonnative hydrophobic and electrostatic effects with the funneled energetics of the landscape determines the properties of unfolded ensembles in agreement with those determined by TRFET experiments.

Materials and Methods

To distinguish and quantify the various interaction effects, we simulate several models of the unfolded ensembles of cytochrome *c*. We start with a perfectly funneled energy landscape Hamiltonian based on the native topology alone. We call this the pure-funnel model and the potential is written as $H_{pure-funnel} = V_{back} + V_{contact} + V_{heme}$. The backbone potential, V_{back}, does not contain funneled energetic terms, rather it merely acts to maintain backbone geometry (bond angles and distances) and accounts for excluded volume. The contact potential, $V_{contact}$, includes attractive Gaussian well potentials derived from the native X-ray structure summed over all C^αs and C^βs for the protein–protein contacts. In addition, there are explicit interactions between the C^αs and the heme pseudoatoms that model the heme-protein contacts. The native structure stabilizing term, $V_{contact}$, is not a simple pair potential but contains 40% nonadditivity (in the native-state conformation). The nonadditivity accounts for cooperative effects that arise from preaveraging over side chain and solvent interactions. The size of the nonadditivity we employ has been shown to account quite well for protein folding barriers and was used in our earlier study of the local unfolding events near the native ensemble (29). The backbone and interaction potentials have been described in detail (29–31). The potential describing the heme, V_{heme}, includes 8 harmonic constraints and a planarity potential that maintains the square-planar geometry of the five heme pseudoatoms and two harmonic constraints to mimic the covalent bonds. We also study another model that contains, in addition to funneled

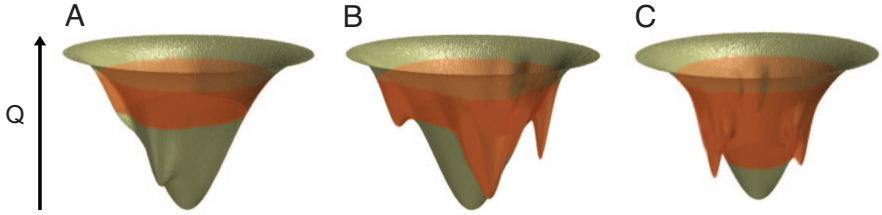

Fig. 8. Representations of the folding energy landscapes for cytochrome *c*. The height corresponds roughly to the structural similarity to the native state, whereas the width corresponds to entropy. In all cases, the folding energy landscape is highly funneled, even in the denatured region. Nevertheless, the region of the funnel that is populated, shown in orange, varies significantly with solvent conditions. (*A*) The GuHCl-denatured ensemble at pH 7 with imidazole to inhibit misligation. GuHCl acts to destabilize native state contacts causing the ensemble to populate higher regions of the funnel. Electrostatic forces are highly screened and do not cause significant traps, allowing the funnel to be fairly smooth. (*B*) The GuHCl-denatured ensemble at pH 7 with several large traps due to misligation by histidine residues. These misligated intermediates can be structurally similar to the native state as indicated by the depth of their minima in the funnel. (*C*) The acid-induced unfolded ensemble with modest traps caused by chemical frustration. These frustrated traps are due to nonnative electrostatic interactions that result from protonation of residues at low pH.

contact terms, a weak potential that leads to a generic hydrophobic collapse of the protein: $V_{Rg} = 0.01(Rg - Rg^N)^2$ where Rg^N is the radius of gyration of the native state. The Hamiltonian of the hydrophobic collapse model is: $H_{\text{collapse}} = H_{\text{pure}-\text{funnel}} + V_{Rg}$. A long-range electrostatic potential, with charges placed on the C^β atoms, was included to describe denaturation induced by GuHCl and acid which changes the charge-charge interactions:

$$V_{elec} = \Sigma_{|i-j|>2}\gamma V_{Rg} \times \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \times e^{(-r_{ij}/\lambda_D)}.$$

Ion screening effects are accounted for using a Debye screening term $(e^{(-r_{ij}/\lambda_D)})$. The charge on a residue is +1, −1, or 0, depending on the identity of that residue, r_{ij} is the distance between C^βs of residues i and j, and γ is the strength of a native contact. Because the funneled energy terms already implicitly contain native electrostatic effects, we chose a dielectric constant so that the long-range and nonnative electrostatics have a more dominant effect in a manner similar to that described (31, 32). The dielectric constant ($\varepsilon = 30$) is chosen to be smaller than that for water but is still significantly larger than the value of 3.5 calculated for ferricytochrome *c* (33). The long-range electrostatic potential is scaled by the amount of collapse to avoid overcompensating for electrostatics implicitly included in the contact terms that are funneled to the native structure. This model allows for the interplay of electrostatic effects and hydrophobic collapse on the energy landscape of the unfolded ensemble and also allows for a nearly unfrustrated energy landscape near the native ensemble. The Hamiltonian used for the GuHCl unfolded ensemble simulations is given by: $H_{\text{GuHCl}} = H_{\text{pure}-\text{funnel}} + V_{Rg} + n_\varepsilon \times V_{\text{elec}}$. The Hamiltonian used for the acid/salt unfolded ensemble simulations is: $H_{\text{acid/salt}} = H_{\text{pure}-\text{funnel}} + V_{Rg} + n_\varepsilon \times (V_{\text{elec}} + V_{\text{prot}})$. The values of λ_D theoretically correspond to the experimental ionic strengths and are set to 1.7 Å for 3.0 M GuHCl and either 2.0, 2.5, or 3.0 Å for 1.0, 0.25, and 0 M Na_2SO_4, respectively. n_ε scales the contribution of electrostatic effects on the ensemble. The protonation potential, V_{prot}, accounts for the energetic frustration reflected in the destabilization of favorable native contacts due to protonation and is given by:

$$V_{prot} = -\sum_{i \text{ or } j \text{ prot}} 0.02\gamma\varepsilon_{ij}(r_{ij})$$

(31). $\varepsilon_{ij}(r_{ij})$ is a unit-depth Gaussian well contact potential between residues i and j. The factor 0.02 was determined by first calculating the electrostatic energy change upon protonation obtained by placing charges on the C^βs in the native conformation. The total destabilization energy distributed among all of the contacts involving protonated residues is then ≈0.02. We also simulated random polymer models that are not based on funneled energy landscapes. The nonfunneled *tabula rasa* model contains excluded volume and backbone connectivity only: $H^{n.f.}_{tabularasa} = V_{back} + V_{heme}$. Models with generic collapse (and no funneling) along with electrostatics are used: $H^{n.f.}_{collapse} = V_{back} + V_{heme} + V_{Rg}$, $H^{n.f.}_{collapse/elec} = V_{back} + V_{heme} + V_{Rg} + n_\varepsilon \times V_{elec}$.

We chose the temperature for the simulations of the unfolded ensembles to be greater than T_f such that the free energy difference between the folded and unfolded basins is 5 k_BT and the temperature for the simulations of the acid/salt ensembles to be below T_f such that the free energy difference between the high salt molten globule basin and its unfolded basin is −5 k_BT.

The folding order parameter Q measures a structure's similarity to the native crystal structure and is given by:

$$Q = \frac{2}{(N-1)(N-2)} \sum_{i<j-1} exp\left[\frac{-(r_{ij} - r_{ij}^{nat})^2}{2|i-j|^{0.3}}\right].$$

The summation runs over all C^α residues, r_{ij}^{nat} is the distance between C^αs in the crystal structure, and N refers to the number of residues. The measure Q_{charge} is analogous to Q but the summation is restricted to charge-charge contacts.

A quantitative comparison of the agreement for the $P(r)$ distributions is shown using an information theoretic surprisal measure:

$$Surprisal(\text{residue } i) = \sum_r -P_{exp}(r) log(P_{sim}(r)/P_{exp}(r)).$$

Surprisal is a measure of the similarity between two distributions and is related to entropy. If two distributions are the same then the surprisal is zero; surprisal increases as the similarity between the two distributions decreases.

ACKNOWLEDGMENTS. We thank Garyk Papoian for advice on making the funnel images. This work was supported by National Institutes of Health Grants GM044557 (to P.G.W.), GM068461 (to J.R.W.), and DK019038 (to H.B.G.) and by National Science Foundation Center for Theoretical Biological Physics Grant PHY-0822283.

1. Kohn JE, *et al.* (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101:12491–12496.
2. Stigter D, Alonso D, Dill K (1991) Protein stability - electrostatics and compact denatured states. *Proc Natl Acad Sci USA* 88:4176–4180.
3. Kundrotas PJ, Karshikoff A (2002) Modeling of denatured state for calculation of the electrostatic contribution to protein stability. *Protein Sci* 11:1681–1686.
4. Kundrotas PJ, Karshikoff A (2003) Effects of charge-charge interactions on dimensions of unfolded proteins: A Monte Carlo study. *J Chem Phys* 119:3574–3581.
5. Oliveberg M, Vuilleumier S, Fersht A (1994) Thermodynamic study of the acid denaturation of barnase and its dependence on ionic strength: Evidence for residual electrostatic interactions in the acid/thermally denatured state. *Biochemistry* 33:8826–8832.
6. Cho JH, Raleigh DP (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.
7. Pace CN, Alston RW, Shaw KL (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci* 9:1395–1398.
8. Trefethen JM, Pace CN, Scholtz JM, Brems DN (2005) Charge-charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa. *Protein Sci* 14:1934–1938.
9. Lattman E, Fiebig K, Dill K (1994) Modeling compact denatured states of proteins. *Biochemistry* 33:6158–6166.
10. Dill K, Shortle D (1991) Denatured states of proteins. *Annu Rev Biochem* 60:795–825.
11. Mccarney ER, *et al.* (2005) Site-specific dimensions across a highly denatured protein; a single molecule study. *J Mol Biol* 352:672–682.
12. Ackerman MS, Shortle D (2002) Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. *Biochemistry* 41:13791–13797.
13. Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293:487–489.
14. Hodsdon ME, Frieden C (2001) Intestinal fatty acid binding protein: The folding mechanism as determined by NMR studies. *Biochemistry* 40:732–742.
15. Klein-Seetharaman J, *et al.* (2002) Long-range interactions within a nonnative protein. *Science* 295:1719–1722.
16. Flanagan J, Kataoka M, Shortle D, Engelman D (1992) Truncated staphylococcal nuclease is compact but disordered. *Proc Natl Acad Sci USA* 89:748–752.
17. Flanagan J, Kataoka M, Fujisawa T, Engelman D (1993) Mutations can cause large changes in the conformation of a denatured protein. *Biochemistry* 32:10359–10370.
18. Sosnick T, Trewhella J (1992) Denatured states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry* 31:8329–8335.
19. Segel DJ, Fink AL, Hodgson KO, Doniach S (1998) Protein denaturation: A small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome c. *Biochemistry* 37:12443–12451.
20. Akiyama S, *et al.* (2002) Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle X-ray scattering. *Proc Natl Acad Sci USA* 99:1329–1334.
21. Pletneva EV, Gray HB, Winkler JR (2005) Snapshots of cytochrome c folding. *Proc Natl Acad Sci USA* 102:18397–18402.
22. Pletneva EV, Gray HB, Winkler JR (2005) Nature of the cytochrome c molten globule. *J Am Chem Soc* 127:15370–15371.
23. Pletneva EV, Gray HB, Winkler JR (2005) Many faces of the unfolded state: Conformational heterogeneity in denatured yeast cytochrome c. *J Mol Biol* 345:855–867.
24. Chahine J, Nymeyer H, Leite VBP, Socci ND, Onuchic JN (2002) Specific and nonspecific collapse in protein folding funnels. *Phys Rev Lett* 88:168101.
25. Levy Y, Wolynes PG, Onuchic JN (2004) Protein topology determines binding mechanism. *Proc Natl Acad Sci USA* 101:511–516.
26. Chavez LL, Gosavi S, Jennings PA, Onuchic JN (2006) Multiple routes lead to the native state in the energy landscape of the β-trefoil family. *Proc Natl Acad Sci USA* 103:10254–10258.
27. Clementi C, Jennings PA, Onuchic JN (2000) How native-state topology affects the folding of dihydrofolate reductase and interleukin-1β *Proc Natl Acad Sci USA* 97:5871–5876.
28. Yang SC, *et al.* (2004) Domain swapping is a consequence of minimal frustration. *Proc Natl Acad Sci USA* 101:13786–13791.
29. Weinkam P, Zong C, Wolynes PG (2005) A funneled energy landscape for cytochrome c directly predicts the sequential folding route inferred from hydrogen exchange experiments. *Proc Natl Acad Sci USA* 102:12401–12406.
30. Eastwood MP, Wolynes PG (2001) Role of explicitly cooperative interactions in protein folding funnels: A simulation study. *J Chem Phys* 114:4702–4716.
31. Lyubovitsky JG, Gray HB, Winkler JR (2002) Structural features of the cytochrome *c* molten globule revealed by fluorescence energy transfer kinetics. *J Am Chem Soc* 124:14840–14841.
32. Cho SS, Weinkam P, Wolynes PG (2008) Origins of barriers and barrierless folding in BBL. *Proc Natl Acad Sci USA* 105:118–123.
33. Simonson T, Perahia D, Brunger AT (1991) Microscopic theory of the dielectric properties of proteins. *Biophys J* 59:670–690.
34. Pierce MM, Nall BT (2000) Coupled kinetic traps in cytochrome c folding: His-heme misligation and proline isomerization. *J Mol Biol* 298:955–969.
35. Colon W, Wakem LP, Sherman F, Roder H (1997) Identification of the predominant non-native histidine ligand in unfolded cytochrome c. *Biochemistry* 36:12535–12541.
36. Chen EF, Goldbeck RA, Kliger DS (2004) The earliest events in protein folding: A structural requirement for ultrafast folding in cytochrome c. *J Am Chem Soc* 126:11175–11181.
37. Wandschneider E, Bowler BE (2004) Conformational properties of the iso-1-cytochrome c denatured state: Dependence on guanidine hydrochloride concentration. *J Mol Biol* 339:185–197.
38. Hammack BN, Smith CR, Bowler BE (2001) Denatured state thermodynamics: Residual structure, chain stiffness and scaling factors. *J Mol Biol* 311:1091–1104.
39. Maity H, Maity M, Englander SW (2004) How cytochrome c folds, and why: Submolecular foldon units and their stepwise sequential stabilization. *J Mol Biol* 343:223–233.