Residual Electrostatic Effects in the Unfolded State of the N-Terminal Domain of L9 Can Be Attributed to Nonspecific Nonlocal Charge−Charge Interactions †

ARTICLE *in* BIOCHEMISTRY · JUNE 2002

Impact Factor: 3.02 · DOI: 10.1021/bi025580m · Source: PubMed

1 AUTHOR:

Huan-Xiang Zhou
Florida State University
259 PUBLICATIONS **10,190** CITATIONS

SEE PROFILE

Residual Electrostatic Effects in the Unfolded State of the N-Terminal Domain of L9 Can Be Attributed to Nonspecific Nonlocal Charge−Charge Interactions†

Huan-Xiang Zhou*

Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104

Received January 23, 2002; Revised Manuscript Received April 1, 2002

ABSTRACT: Residual electrostatic interactions in the unfolded state of the N-terminal domain of L9 (NTL9) were found by Kuhlman et al. [(1999) *Biochemistry 38*, 4896−4903]. These residual interactions are analyzed here by the Gaussian-chain model [Zhou, H.-X. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 3569−3574]. The original model is made more realistic by replacing "standard" model-compound pK_a values for ionizable groups by those measured by Kuhlman et al. in peptide fragments of NTL9. The predicted pH dependence of the unfolding free energy is in agreement with experiment over the pH range of 1−7 at ionic strengths of 100 and 750 mM. This indicates that the residual electrostatic effects in the unfolded state of NTL9 can be attributed to nonspecific nonlocal charge−charge interactions.

The pH dependence of the unfolding free energy, ΔG_{unfold}, provides a unique opportunity for gaining insight into the unfolded state of a protein under physiological conditions. This dependence is governed by (*1*)

$$\Delta G_{unfold}(\mathrm{pH}) - \Delta G_{unfold}(\mathrm{pH}_0) = (k_B T \ln 10)\int_{\mathrm{pH}_0}^{\mathrm{pH}} Q_u \, d\mathrm{pH} - (k_B T \ln 10)\int_{\mathrm{pH}_0}^{\mathrm{pH}} Q_f \, d\mathrm{pH} \quad (1)$$

where k_BT is the product of the Boltzmann constant and the absolute temperature and Q_u (Q_f) is the total charge on the protein at a given pH in the unfolded (folded) state. ΔG_{unfold} can be obtained by measuring the unfolding free energy under decreasingly denaturing conditions and then extrapolating to physiological conditions, whereas Q_f can be obtained either from pH titration of the folded protein or pK_a values of all ionizable groups measured in the folded state. The total charge on the protein in the unfolded state, which is ordinarily inaccessible under physiological conditions, can then be obtained from eq 1. Q_u in turn sheds light on the nature of the unfolded state.

Historically, the unfolded state has been thought of as a "random coil" that is devoid of any residual interactions (*2*, *3*). Then the total charge in the unfolded state can be estimated from

$$Q_u = -\sum_i 1/(1 + 10^{\mathrm{p}K_{i,0} - \mathrm{pH}}) + N_+ \quad (2)$$

where p$K_{i,0}$ are the pK_a values of model compounds for the ionizable groups and N_+ is the number of ionizable groups that become charged upon protonation (e.g., Arg and Lys). A number of recent experimental studies have found that, when Q_u estimated from eq 2 is used in eq 1, the predicted ΔG_{unfold} shows substantial deviations from actual measurements (*4*−*9*). These demonstrate that there are residual charge−charge interactions in the unfolded state.

The importance of residual charge−charge interactions on protein stability has now been recognized (*10*−*12*). Elcock modeled the unfolded state by one structure partially unfolded from the folded structure (*10*). We have just proposed a Gaussian-chain model to account for the residual charge−charge interactions (*12*). The interaction energy between two ionizable groups at a distance r is given by the Debye−Hückel theory

$$U = \pm 332 \exp(-\kappa r)/\epsilon r \quad (3)$$

where ϵ is the dielectric constant of water, $\kappa = (8\pi I e^2/\epsilon k_B T)^{1/2}$ ($= I^{1/2}/3.04$ Å^{-1} at room temperature), and I is the ionic strength. The distance between two groups is assumed to have a Gaussian distribution

$$p(r) = 4\pi r^2 (3/2\pi d^2)^{3/2} \exp(-3r^2/2d^2) \quad (4)$$

where d is the root-mean-square distance. This mean distance depends on l, the number of peptide bonds separating the two residues, and is given by

$$d = b l^{1/2} + s \quad (5)$$

where the effective bond length b was set to 7.5 Å and the shift distance s (to account for the fact that the distance of interest is between two side chains) was set to 5 Å. The mean interaction energy then has the magnitude

$$W_{ij} = 332(6/\pi)^{1/2}[1 - \pi^{1/2} x \exp(x^2)\,\mathrm{erfc}(x)]/\epsilon d \quad (6)$$

where $x = \kappa d/6^{1/2}$ and erfc(x) is the complementary error function. The residual interactions shift the pK_a values of the ionizable groups in the unfolded state. As a result, the total charge Q_u on the protein is significantly different from what is predicted by eq 2. This model was found to give accurate predictions for the pH dependence of ΔG_{unfold} of

† This work was supported in part by Grant GM58187 from the National Institutes of Health.

* After July 1, 2002, address correspondence to this author at the Institute of Molecular Biophysics and Department of Physics, Florida State University, Tallahassee, FL 32306 [telephone (215) 895-2716; fax (215) 895-5934; e-mail zhou@sb.fsu.edu].

10.1021/bi025580m CCC: $22.00
Published on Web 04/27/2002

Table 1: pK_a Values of the Acidic Residues in the Folded and Unfolded States of NTL9 and in Peptide Fragments at $I = 100$ and 750 mM

	$I = 100$ mM			$I = 750$ mM			
redisue	p$K_{i,f}$[a]	p$K_{i,u}$[b]	p$K_{i,frag}$[a,c]	p$K_{i,f}$[a]	p$K_{i,u}$[b]	p$K_{i,frag}$[a,c]	fragment[d]
D8	2.99 ± 0.05	3.66	3.84 ± 0.06 3.67	3.17 ± 0.07	3.78	3.82 ± 0.03 3.80	1−11: NH_3−MKVIFLKDVKG−NH_2
E17	3.57 ± 0.05	3.99	4.11 ± 0.17 4.10	3.80 ± 0.12	4.13	4.15 ± 0.02 4.21	12−23: Ac−KGKKGEIKNVAD−NH_2
D23	3.05 ± 0.04	3.86	4.11 ± 0.11 4.0[e]	3.18 ± 0.05	3.78	3.83 ± 0.01 3.9[e]	21−27: Ac−VADGYAN−NH_2
E38	4.04 ± 0.05	4.46	4.63 ± 0.14 4.4[e]	4.12 ± 0.14	4.27	4.30 ± 0.04 4.3[e]	35−42: Ac−LAIEATPA−NH_2
E48	4.21 ± 0.08	4.27	4.31 ± 0.12 4.18	4.35 ± 0.05	4.35	4.35 ± 0.08 4.24	40−56: Ac−TPANLKALEAQKQKEQR−NH_2
E54	4.21 ± 0.08	4.28	4.32 ± 0.14 4.06	4.23 ± 0.09	4.18	4.19 ± 0.11 4.20	40−56

[a] Measured by Kuhlman et al. (*8*). The Hill coefficients vary from 0.85 to 1.22 at $I = 100$ mM and from 0.83 to 1.20 at $I = 750$ mM. [b] Calculated with the Gaussian-chain model by using the fragments as "model compounds". The Hill coefficients vary from 0.95 to 1.00 at $I = 100$ mM and from 0.98 to 1.00 at $I = 750$ mM. [c] The second number for each residue was calculated with the original Gaussian-chain model on the fragment listed in the last column. [d] The full sequence of NTL9 is NH_3−MKVIFLKDVKGKGKKGEIKNVADGYANNFLFKQGLAIEATPANLKALEAQ-KQKEQR−NH_2. [e] pK_a values for model compounds were assumed because the fragments do not have other charges.

barnase, chymotrypsin inhibitor 2, ovomucoid third domain, and ribonucleases A and T1 over wide pH ranges.

As we cautioned when the Gaussian-chain model was first proposed (*12*), the use of the model does not suggest that an unfolded protein actually samples conformations expected of a Gaussian chain. Rather, as a polymer chain, the unfolded protein is expected to have mean residue−residue distances that increase with sequence separation. The Gaussian chain is just the simplest model to account for this increase of mean residue−residue distances. This model thus will fail if the unfolded state is dominated by specific nonlocal charge−charge interactions.

Recently Kuhlman et al. (*8*) carried out a particularly illuminating study on the pH-dependent stability of the N-terminal domain of L9 (NTL9).[1] They determined ΔG_{unfold} by thermal and urea denaturation and obtained the pK_a values of the six acidic residues of NTL9 in the folded state by NMR. Importantly, they also determined the pK_a values (to be called p$K_{i,frag}$) of the acidic residues in peptide fragments of the protein. The experimental results at $I = 100$ and 750 mM are listed in Table 1. When p$K_{i,frag}$ values were used (in place of p$K_{i,0}$) in eq 2 to predict Q_u, there were substantial discrepancies between ΔG_{unfold} calculated from eq 1 and experimental results at $I = 100$ mM (see Figure 1). Kuhlman et al. suggested that the discrepancies serve as "evidence that there is nonrandom structure in the denatured state of NTL9". They concluded "in the denatured state there may be transiently formed structures that bring charges distant in primary sequence together".

In this paper, we use the Gaussian-chain model to predict the total charge on NTL9 in the unfolded state. The availability of p$K_{i,frag}$ removes the two most suspect aspects of the Gaussian-chain model. First, the "standard" model-compound pK_a values can be replaced by the experimentally determined p$K_{i,frag}$. Second, because local interactions (along the sequence) are already reflected in p$K_{i,frag}$, only interactions with charges outside the peptide fragment need to be modeled. The Gaussian-chain model is more suitable for the distribution of these nonlocal distances.

We show that the pH dependence of ΔG_{unfold} predicted by the Gaussian-chain model (as modified by the use of p$K_{i,frag}$) is in agreement with experiment. This indicates that the residual electrostatic effects in the unfolded state of NTL9 can be attributed to nonspecific nonlocal charge−charge interactions.

[1] Abbreviations: NTL9, the N-terminal domain of protein L9.

MODELS AND METHODS

Original Gaussian-Chain Model. In the model proposed earlier (*12*), we assumed that, in the absence of residual charge−charge interactions, the pK_a values of the ionizable groups take "standard" model-compound values p$K_{i,0}$. The distribution of the protonation states x_i of the ionizable groups is governed by the Hamiltonian (*13*)

$$H = (k_B T \ln 10)\sum_i (\mathrm{pH} - \mathrm{p}K_{i,0})x_i + (1/2)\sum_{i\neq j} W_{ij}(x_i - x_{i0})(x_j - x_{j0}) \quad (7)$$

where $x_i = 0$ (1) when group i is unprotonated (protonated) and x_{i0} is the protonation state when the group is charge neutral (1 for Asp and Glu and 0 for Arg and Lys). The average protonation of group i at a given pH is

$$\bar{x}_i = \sum_{\{x_i\}} x_i \exp(-H/k_B T)/\sum_{\{x_i\}} \exp(-H/k_B T) \quad (8)$$

which was evaluated by a Monte Carlo simulation. The total charge on the protein is

$$Q_u = \sum_i (\bar{x}_i - x_{i0}) \quad (9)$$

This was then integrated numerically over pH to yield pH dependence of the unfolding free energy (see eq 1). The total charge Q_f on the folded protein, required in eq 1, was calculated from the experimental pK_a values of the ionizable groups in the folded state (listed under p$K_{i,f}$ in Table 1) and the associated Hill coefficients.

The pH titration of an ionizable group can be described by a pK_a value. This was obtained by fitting the pH dependence of $\bar{x}_i$ to the Hill equation:

$$\log \frac{\bar{x}_i}{1 - \bar{x}_i} = n_i(\mathrm{p}K_{i,u} - \mathrm{pH}) \quad (10)$$

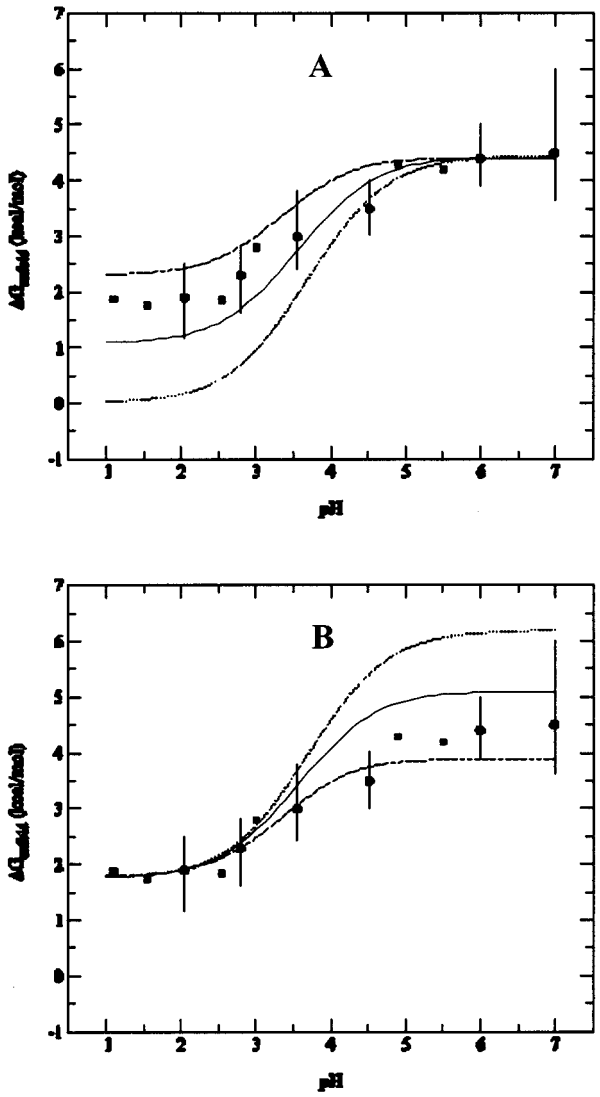

FIGURE 1: Comparison of experimental and calculated pH dependence of the unfolding free energy at $I = 100$ mM. Circles with error bars are from Kuhlman et al. (*8*), and squares are additional data from Luisi and Raleigh (*14*). Solid and dashed curves display predictions of the modified and original Gaussian-chain models, respectively, and the dotted curve displays the prediction without nonlocal charge−charge interactions (with Q_u calculated from experimental $pK_{i,frag}$). Predictions are shifted vertically so they coincide with experiment at (A) pH 6 or (B) pH 2.

The standard model-compound pK_a values were assumed to be 4.0 for Asp and 4.4 for Glu at $I = 100$ mM. Higher salt concentrations help to stabilize the ionized state, and this affects the pK_a values of model compounds. The difference in solvation energy between an unprotonated Asp (or Glu) residue and the protonated form was found by the UHBD program to decrease by 0.12 kcal/mol when the ionic strength increased from 100 to 750 mM. This corresponds to a decrease of 0.1 unit in the model-compound pK_a values. Using the extended Debye−Huckle law, Kuhlman et al. (*8*) also estimated a downward shift of 0.1. The model-compound pK_a values of Asp and Glu at $I = 750$ mM therefore were taken to be 3.9 and 4.3, respectively.

Modified Gaussian-Chain Model with Use of $pK_{i,frag}$. The original Gaussian-chain model was modified with $pK_{i,0}$ replaced by the pK_a values of ionizable groups measured in peptide fragments. To be self-consistent, for each ionizable group whose model-compound pK_a was set to $pK_{i,frag}$, the interaction terms between this group and other ionizable groups within the fragment were set to zero at the same time.

NTL9 has 56 residues. Its sequence is shown in Table 1. The five peptide fragments on which the pK_a values of the six acidic groups of NTL9 were measured by Kuhlman et al. (*8*) are also listed in Table 1. Calculations were done at $I = 100$ and 750 mM and a temperature of 25 °C. For $I = 100$ mM, the more recent work of Luisi and Raleigh (*14*) has expanded the pH range of the ΔG_{unfold} data from 2−7 to 1.1−10.5. However, the pK_a values of the N-terminal, Y25, and other basic residues in the folded state are not known. We therefore limited the calculations at $I = 100$ mM to pH 1−7. The calculations at $I = 750$ mM were done for pH 2−7.3. Because the pK_a of the N-terminal was not known in the folded state, the charge on the N-terminal was not included in the calculation of either Q_f or Q_u. This is equivalent to assuming that the pK_a of the N-terminal is unchanged upon unfolding.

Prediction of $pK_{i,frag}$. The values of $pK_{i,frag}$ were predicted by applying the original Gaussian-chain model to the five peptide fragments. That is, we assumed that each fragment is a polymer chain with distances between charged residues distributed according to eq 4.

Parameters in all of the calculations on NTL9 were not adjusted. They were simply taken from our earlier study of residual charge−charge interactions in five other unfolded proteins (*12*).

RESULTS AND DISCUSSION

Calculated pK_a Values in the Unfolded State. The pK_a values calculated by the modified Gaussian-chain model for the six acidic groups in the unfolded state at $I = 100$ mM are shown in Table 1. They are all downward shifted from the measured $pK_{i,frag}$ and are between those measured in the folded state and $pK_{i,frag}$. The downward shifts of the calculated $pK_{i,u}$ range from 0.25 for D23 to 0.04 for E48 and E54.

To understand these shifts, it is useful to recognize that NTL9 has 13 positive charges (11 lysines, 1 argine, and the N-terminal). Fragments 1, 2, and 5 each have four positive charges, with the remaining charge, K32, in the small gap between fragments 3 and 4. According to the Gaussian-chain model, the mean distance between two residues increases with their sequence separation. The relatively large downward shift of $pK_{i,u}$ for D23 (located in fragment 3) can therefore be accounted for by the fact that it is right next to fragments 2 and 1 (with four positive charges each) and the small gap (containing K32). The intermediate shifts of $pK_{i,u}$ for D8, E17, and E38 can be related to the fact each of these residues is next to just one fragment with a large positive charge (fragments 2, 1, and 5, respectively). The small shifts of $pK_{i,u}$ for E48 and E54 are due to the large sequence separations between these residues and the positively charged fragments 1 and 2.

It should be noted that the measured $pK_{i,f}$ values show downward shifts from $pK_{i,frag}$ with exactly the same rank order as the calculated $pK_{i,u}$. In particular, $pK_{i,f}$ values shift down by 1.06 for D23 but by merely 0.1 for E48 and E54. Although the congruence of the pK_a rank ordering may be construed as suggesting that the unfolded state is native-

like, the proper interpretation is that, both in the unfolded state and the folded state, charges closer along the sequence will exert somewhat stronger influences on the pK_a value. Even in the folded state, charges closer along the sequence will more often have shorter distances than those far apart along the sequence.

The experimental and calculated pK_a values at $I = 750$ mM are also shown in Table 1. Here p$K_{i,f}$ values measured by Kuhlman et al. (*8*) for E38, E48, and E54 are nearly the same as p$K_{i,frag}$, indicating that electrostatic interactions have been screened to a large extent at the high ionic strength. Not surprisingly, the calculated pK_a values for the six acidic groups in the unfolded state differ very little from p$K_{i,frag}$. They are still downward shifted, but by <0.06.

Predicted pH Dependence of ΔG_{unfold}. The total charge Q_u on NTL9 in the unfold state, calculated with the Gaussian-chain model using the measured p$K_{i,frag}$, is put in eq 1, along with the experimental Q_f as calculated from the measured p$K_{i,f}$ (and the associated Hill coefficients), to predict the pH dependence of ΔG_{unfold}. The results at $I = 100$ mM are compared in Figure 1 with the experimental data of Kuhlman et al. (*8*) and Luisi and Raleigh (*14*). It can be seen that, with the inclusion of residual charge−charge interactions by the Gaussian-chain model, ΔG_{unfold} is now predicted to have much weaker pH dependence. This weaker pH dependence of ΔG_{unfold} is in agreement with experiment over the pH range of 1−7.

We also tested the prediction of the original Gaussian-chain model (using "standard" model-compound pK_a values). The pH dependence of ΔG_{unfold} is found to be somewhat weaker than what is found by using p$K_{i,frag}$ in the Gaussian-chain model. A major contribution to the difference of the two predictions is the somewhat large pK_a values, 4.11 and 4.63, measured for D23 and E38 in fragments 3 and 4, respectively. Neither of the two fragments has any additional charges, thus the pK_a values of D23 and E38 in the fragments were predicted to have "standard" model-compound values (4.0 and 4.4, respectively). The experimental data for the pH dependence of ΔG_{unfold} are bounded by the two Gaussian-model predictions. Both predictions appear to fall just within the relatively large error ranges of the experimental data.

At $I = 750$ mM, Kuhlman et al. (*8*) hardly detected any residual charge−charge interactions in the unfolded state. In line with this, the predicted pH dependence of ΔG_{unfold} by the modified Gaussian-chain model differs only slightly from that predicted by setting residual charge−charge interactions to zero (see Figure 2). As Figure 2 shows, for this ionic strength, the predictions of the modified and the original Gaussian-chain models are barely distinguishable.

Predicted pK_a Values in the Peptide Fragment. To further test the original Gaussian-chain model, we applied it on the five peptide fragments to predict p$K_{i,frag}$. The results for $I = 100$ and 750 mM are shown in Table 1. For D23 and E38, the peptide fragments containing them do not have other charges, so their pK_a values are predicted to take standard model-compound values (4.0 and 4.4, respectively, at $I = 100$ mM). Deviations of the measured pK_a values from model-compound values at $I = 100$ mM have been noted already. At $I = 750$ mM, the measured pK_a values of D23 and E38, 3.83 and 4.30, respectively, are actually close to the model-compound values corrected for ionic strength dependence: 3.9 and 4.3. For D8 and E17, the predicted

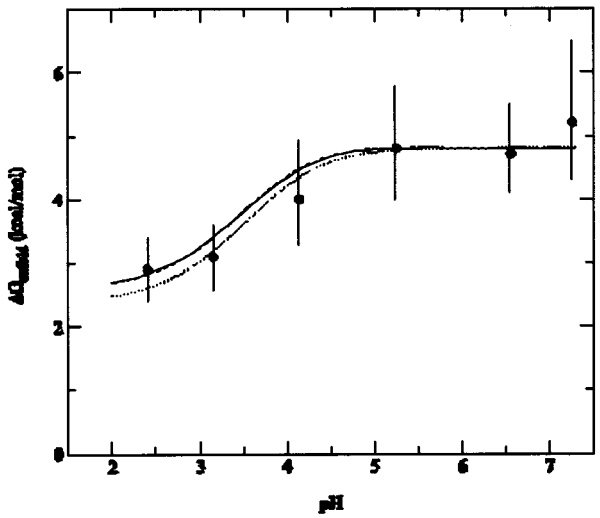

FIGURE 2: Comparison of experimental and calculated pH dependence of the unfolding free energy at $I = 750$ mM. Circles with error bars are from Kuhlman et al. (*8*). Solid and dashed curves (barely distinguishable) display predictions of the modified and original Gaussian-chain models, respectively, and the dotted curve displays the prediction without nonlocal charge−charge interactions.

p$K_{i,frag}$ values are downward shifted from model-compound values by ~0.3 at $I = 100$ mM and by 0.1 at $I = 750$ mM, in reasonable agreement with experiment. The downward shift for D8 (brought by the three Lys residues within the fragment) at the lower ionic strength is somewhat overestimated. An increase of 0.13 unit is predicted for the pK_a of D8 when the ionic strength is increased from 100 to 750 mM, whereas experimentally the pK_a does not change. NMR and CD have indicated the fragments containing D8 and E17 are predominantly unstructured (*15*), which may explain the fair performance of the Gaussian-chain model.

E54 at $I = 100$ mM has the worst predicted pK_a. The measured value, 4.32, is very close to the assumed model-compound value (4.4) and is higher than the result at $I = 750$ mM. The Gaussian-chain model predicts a downward shift of 0.34 unit at $I = 100$ mM, primarily due to the presence of a Lys residue (K53) in the preceding position, and is lower than the corresponding result at the higher ionic strength. It is possible that, in the hypothetical situation where interactions with other charges in the fragment are turned off, E54 has a pK_a higher than the assumed model-compound value of 4.4 (as illustrated by D23 and E38). If the pK_a of E54 in this hypothetical situation actually takes the assumed value of 4.4 and the small downward shift of the measured pK_a in the fragment by 0.08 unit is solely attributed to K53, then, according to eq 3, that positive charge must be as far away as $r = 11.6$ Å from E53. The mean distance d for two neighboring residues in the present Gaussian-chain model is set to 12.5 Å (see eq 5). Although d is slightly larger than r, it is the distances much less than d in the Gaussian distribution that dominate the contributions to the interaction energy. The presumed small perturbation on the measured pK_a of E54 by K53 can be explained either by a narrow distribution of charge−charge distances around ~11.6 Å or by a broad distribution (such as the Gaussian distribution) that has a mean distance substantially >12.5 Å. Given that the maximum distance between these two charges connected by a single peptide bond is ~15 Å, the first scenario is much more likely. A narrow distance distribution is obtained if

Table 2: Effective Bond Lengths Deduced from Measured Radii of Gyration of Denatured Proteins

protein	N	R_g (Å)	b_{eff} (Å)	reference
B1 domain of protein G	56	23	7.6	Smith et al., 1996 (*32*)
protein L	62	26	8.2	Plaxco et al., 1999 (*33*)
bovine ubiquitin	76	26.3	7.4	Kamatari et al., 1999 (*34*)
horse cytochrome *c*	104	30.3	7.3−7.8	Segel at al., 1998 (*35*)
		32.4		Kataoka et al., 1993 (*36*)
neocarzinostatin	113	26	6.0	Perez et al., 2001 (*37*)
α-lactalbumin	123	30	6.7	Kataoka et al., 1997 (*38*)
bovine ribonuclease A	124	28	6.2	Sosnick and Trewhella, 1992 (*39*)
hen lysozyme	129	23.5	5.1−6.2	Segel et al., 1999 (*40*)
		28.7		Kamatari et al., 1998 (*41*)
staphylococcal nuclease	149	35	7.0−8.7	Flanagna et al., 1992 (*42*)
		43		Panick et al., 1998 (*43*)

the peptide fragment has a stable conformation. The fragment containing E54 (and E48 as well) is actually found to be 40% helical at room temperature (*15*). In the X-ray structure of the full-length protein L9 (PDB entry 1div), the distance between NZ of K53 and CD of E54 is only 4.9 Å, but this distance will likely be different in the fragment (and, by extension, in the unfolded state of NTL9).

The measured pK_a values for the peptide fragments serve two purposes. They offer an opportunity to test the original Gaussian-chain model, and their incorporation into the model also allows it to be more realistic.

How Realistic Is the Effective Bond Length b? An important parameter in the Gaussian-chain model is the effective bond length b. Although we have treated b as a model parameter and downplayed its physical significance, the reader may still wonder whether such a "bond length" parameter leads to realistic dimensions of unfolded proteins.

We have investigated the dimensions of denatured proteins by generating chain conformations with (ϕ, ψ) angles sampled from a structural database and explicitly accounting for excluded-volume effects (*16*). The magnitude of the excluded-volume effects was determined by having calculated Stokes radius and intrinsic viscosity reproduce experimental results. The resulting effective bond lengths range from 6 to 8.5 Å. This range brackets the 7.5 Å value of b chosen in the Gaussian-chain model. Even though excluded-volume effects bias local conformations sampled by residues (*16*−*18*), the distributions of residue−residue distances are found to be nearly Gaussian (*16*). The Gaussian distribution is not particularly surprising; indeed, the distributions of residue−residue distances in polymers with excluded volumes often are modeled as Gaussian, although the mean distances are scaled by an expansion factor (*19*, *20*).

Another property of denatured proteins that can be easily probed by small-angle X-ray (or neutron) scattering is the radius of gyration R_g. For a chain with Gaussian distributions of residue−residue distances, R_g is related to the effective bond length b_{eff} by (*19*)

$$R_g^2 = Nb_{eff}^2/6 \qquad (11)$$

where N is the number of residues. In Table 2, we list the experimental results of R_g for nine proteins. The resulting values of b_{eff} calculated according to eq 11 range from 5.1 to 8.7 Å, again bracketing the 7.5 Å value chosen for b. In certain cases (such as hen lysozyme), b_{eff} derived from R_g is clearly smaller than b (the relatively small b_{eff} value of lysozyme is perhaps related to the fact this protein has four disulfide bonds). Another caveat is that the results for R_g were obtained for fully unfolded proteins under strong denaturing conditions. Unfolded proteins under physiological conditions (which typically cannot be probed directly) may have smaller dimensions.

In short, the chosen value of b is not inconsistent with dimensions of denatured proteins as determined experimentally, but we again note that b should really be treated as a model parameter (as opposed to a physical parameter).

Alternative Models for Treating Residual Electrostatic Interactions. Stigter et al. (*21*) have treated the electrostatics of the unfolded state by a porous sphere model. Tanford (*20*) has given a very useful discussion of the difference between the bases of the Gaussian-chain and porous sphere models. In the latter model, the unfolded chain is approximated by a sphere inside which charges on the protein chain and mobile ions from the solution can coexist. The radius of the porous sphere was actually predicted (in conjunction with the heteropolymer collapse model), and two populations of unfolded species, one open and the other more compact, were predicted. In contrast, the size parameters (b and s) in our Gaussian-chain model are not predetermined, although we have not adjusted these parameters ($b = 7.5$ Å and $s = 5.0$ Å) in either the present study of NTL9 or our earlier study of five other proteins (*12*). The residual charge−charge interactions exhibited by these proteins may serve as a simple test of the porous sphere model. Kundrotas and Karshikoff (*22*) have proposed a variation of the porous sphere model.

In contrast to the sampling of residue−residue distances in the Gausian-chain model, Elcock (*10*) modeled the unfolded state by one structure partially unfolded from the folded structure. This was obtained by molecular dynamics simulations with expanded van der Waals radii for all protein atoms. It appears that the model could inherit too much of the nonlocal interactions in the folded state, leading to excessively perturbed pK_a values in the unfolded state (*10*). The Gaussian-chain model and the native-like model of Elcock are qualitatively different. The former relies on nonspecific interactions, which explicitly diminish as sequence separation increases, whereas the latter relies on nonlocal interactions that are present in the folded state. Further tests of the two models are warranted.

The limitations of the Gaussian-chain model are twofold. First, all charge−charge interactions in the unfolded state are treated as nonspecific. Thus, the model will fail in cases when the unfolded state is dominated by specific nonlocal charge−charge interactions. Conceivably, the Gaussian-chain model, with a relatively large mean distance, may fortuitously reproduce the interaction energy between two charged residues that actually have a fixed and small distance (see discussion regarding the predicted pK_a of E54 at $I = 100$ mM). Specific local interactions can be incorporated by using peptide fragments as model compounds, as is done in the present work.

Second, the Gaussian-chain model deals with only electrostatic interactions in the unfolded state. For an unfolded state that is characterized by a hydrophobic cluster (native-like or non-native-like), the electrostatic interactions might still be accounted for by the Gaussian-chain model. However, in this situation the model provides only limited information on the true nature of the unfolded state.

Peptide Fragments as Windows on the Unfolded State. The unfolded protein may be viewed as an ensemble of peptide fragments. In this ensemble the conformations of the fragments and the distances between fragments are changing all of the time. However, there may be recurring local structure elements such as a preferred region on the (ϕ, ψ) map for individual residues or a preferred secondary structure for a stretch of residues. Peptide fragments allow these local structure elements to be captured.

These local structure elements should perturb the pK_a values of ionizable groups. The pK_a values measured on fragment 5 for E48 and E54 may be examples. Insofar as peptide fragments provide good representations of local structures, the use of measured p$K_{i,frag}$ in the Gaussian-chain model makes the model much more realistic.

Do Residual Electrostatic Effects Necessarily Indicate Nonrandom Interactions? There has been intensive interest in the nature of the unfolded state under physiological conditions in recent years (*2*, *3*, *23*−*31*). The persistence of local structures has been widely demonstrated, and investigations into the nature of nonlocal (i.e., tertiary) interactions in the unfolded state have begun (*27*, *28*, *30*, *31*). In some cases, the unfolded state has been characterized as "compact" (*24*, *28*, *31*). Residual charge−charge interactions, by virtue of their relatively ready availability and the relatively weak distance dependence (r^{-1} as opposed to the r^{-6} dependence of NOEs), may prove to be valuable in the characterization of the unfolded state.

Kuhlman et al. attributed the residual electrostatic interactions in the unfolded state of NTL9 to transiently formed (tertiary) structures that bring charges distant in primary sequence together. Our analysis by the Gaussian-chain model indicates that the residual electrostatic effects may not be necessarily due to nonrandom interactions. Indeed, they are consistent with nonspecific charge−charge interactions. This proposition is in line with the more recent finding that the folding rate of NTL9 has a rather weak dependence on pH (*14*), indicating that electrostatic interactions are only weakly formed in the transition state.

REFERENCES

1. Tanford, C. (1970) *Adv. Protein Chem. 24*, 1−95.
2. Dill, K. A., and Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795−825.
3. Shortle, D. (1996) *FASEB J. 10*, 27−34.
4. Oliveberg, M., Vuilleumier, S., and Fersht, A. R. (1994) *Biochemistry 33*, 8826−8832.
5. Swint-Kruse, L., and Robertson, A. D. (1995) *Biochemistry 34*, 4724−4732.
6. Oliverberg, M., Arcus, V. L., and Fersht, A. R. (1995) *Biochemistry 34*, 9424−9433.
7. Tan, Y.-J., Oliverberg, M., Davis, B., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 980−992.
8. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) *Biochemistry 38*, 4896−4903.
9. Whitten, S. T., and Garcia-Moreno E. B. (2000) *Biochemistry 39*, 14292−14304.
10. Elcock, A. H. (1999) *J. Mol. Biol. 294*, 1051−1062.
11. Pace, C. N., Alston, R. W., and Shaw, K. L. (2000) *Protein Sci. 9*, 1395−1398.
12. Zhou, H.-X. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 3569−3574.
13. Zhou, H.-X., and Vijayakumar, M. (1997) *J. Mol. Biol. 267*, 1002−1011.
14. Luisi, D. L., and Raleigh, D. P. (2000) *J. Mol. Biol. 299*, 1091−1100.
15. Luisi, D. L., Wu, W.-J., and Raleigh, D. P. (1998) *J. Mol. Biol. 287*, 395−407.
16. Zhou, H.-X. (2002) *J. Phys. Chem. B* (in press).
17. Chan, H. S. and Dill, K. A. (1990) *J. Chem. Phys. 92*, 3118−3135.
18. Pappu, R. V., Srinivasan, R., and Rose, G. D. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 12565−12570.
19. Doi, M., and Edwards, S. F. (1986) *The Theory of Polymer Dynamics*, Clarendon, Oxford, U.K.
20. Tanford, C. (1961) *Physical Chemistry of Macromolecules*, Wiley, New York.
21. Stigter, D., Alonso, D. O., and Dill, K. A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 4176−4180.
22. Kundrotas, P. J., and Karshikoff, A. (2002) *Phys. Rev. E 65*, 011901.
23. Bierzynski, A., and Baldwin, R. L. (1982) *J. Mol. Biol. 162*, 173−186.
24. Neri, D., Billeter, M., Wider, G., and Wuthrich, K. (1992) *Science 257*, 1559−1563.
25. Logan, T. M., Theriault, Y., and Fesik, S. W. (1994) *J. Mol. Biol. 236*, 637−648.
26. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) *Biochemistry 36*, 8977−8991.
27. Gillespie, J. R., and Shortle, D. (1997) *J. Mol. Biol. 268*, 170−184.
28. Mok, Y.-K., Kay, C. M., Kay, L. E., and Forman-Kay, J. (1999) *J. Mol. Biol. 289*, 619−638.
29. Wong, K.-B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M. V., Fersht, A. R., and Daggett, V. (2000) *J. Mol. Biol. 296*, 1257−1282.
30. Yi, Q., Scalley-Kim, M. L., Alm, E. J., and Baker, D. (2000) *J. Mol. Biol. 299*, 1341−1351.
31. Choi, W.-Y., and Forman-Kay, J. D. (2001) *J. Mol. Biol. 308*, 1011−1032.
32. Smith, C. K., Bu, Z., Anderson, K. S., Sturtevant, J. M., Engelman, D. M., and Regan, L. (1996) *Protein Sci. 5*, 2009−2019.
33. Plaxco, K. W., Millet, I. S., Segel, D. J., Doniach, S., and Baker, D. (1999) *Nat. Struct. Biol. 6*, 554−556.
34. Kamatari, Y. O., Ohji, S., Konno, T., Seki, Y., Soda, K., Kataoka, M., and Akasaka, K. (1999) *Protein Sci. 8*, 873−882.
35. Segal, D. J., Fink, A. L., Hodgson, K. O., and Doniach, S. (1998) *Biochemistry 37*, 12443−12451.
36. Kataoko, M., Hagihara, Y., Mihara, M., and Goto, Y. (1993) *J. Mol. Biol. 229*, 591−596.
37. Perez, J., Vachette, P., Russo, D., Desmadril, M., and Durand, D. (2001) *J. Mol. Biol. 308*, 721−743.
38. Kataoka, M., Kuwajima, K., Tokunaga, F., and Goto, Y. (1997) *Protein Sci. 6*, 422−430.
39. Sosnick, T. R., and Trewhella, J. (1992) *Biochemistry 31*, 8329−8335.
40. Segal, D. J., Bachmann, A., Hofrichter, J., Hodgson, K. O., Doniach, S., and Kiefhaber, T. (1999) *J. Mol. Biol. 288*, 489−499.
41. Katamari, Y., Konno, T., Kataoka, M., and Akasaka, K. (1998) *Protein Sci. 7*, 681−688.
42. Flanagan, J. M., Kataoka, M., Shortle, D., and Engelman, D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 748−752.
43. Panick, G., Malessa, R., Winter, R., Rapp, G., Frye, K. J., and Royer, C. A. (1998) *J. Mol. Biol. 275*, 389−402.

BI025580M