J|A|C|S
COMMUNICATIONS

Published on Web 01/31/2003

Direct Test of the Gaussian-Chain Model for Treating Residual Charge−Charge Interactions in the Unfolded State of Proteins

Huan-Xiang Zhou*

Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306

Received December 20, 2002; E-mail: hxzhou@csit.fsu.edu

Recent experimental studies have indicated significant residual charge−charge interactions in unfolded proteins.[1] These have prompted the development of theoretical models for accounting for such electrostatic effects.[2] Thus far these models have been tested against experimental data for the pH dependence of the total charges on unfolded protein chains.[2,3] Recently Kay and co-workers[4] measured pK_a's of individual Asp, Glu, and His side chains in the unfolded state of the N-terminal SH3 domain of the *Drosophila* protein drk by NMR. Here these measurements were used to critically test the Gaussian-chain model.

The difficulty in characterizing the unfolded state lies in the fact that the unfolded protein chain samples numerous different conformations. However, the conformational sampling also means that not every detail of the unfolded state is significant and the key is to capture important average properties. The hallmark of the Gaussian-chain model (see Figure 1) is that residues closer along the sequence have higher probabilities of sampling short distances, and hence, residual electrostatic effects are attributed to nonspecific interactions dominated by charges close along the sequence.[2b,3]

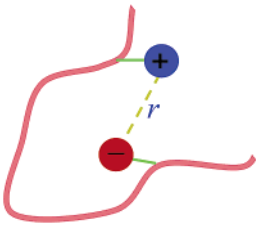
In this model (see Figure 1), the average electrostatic interaction energy between two charged residues i and j is[2b]

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

where $x = \kappa d/6^{1/2}$ and erfc(x) is the complementary error function. These interactions perturb the propensities of ionizable groups to protonate/unprotonate. Through a Monte Carlo simulation, the average protonation of each ionizable group at a given pH was determined.[2b] The pH at which the average protonation was 50% was taken as the pK_a value of that group.

The drkN SH3 domain has 59 residues with 26 ionizable groups (7 Asp, 6 Glu, 5 Lys, 2 Arg, 2 His, 2 Tyr, N-terminal, and C-terminal; see Table 1). Kay and co-workers[4] determined the pK_a's for 12 of the 13 Asp and Glu residues (excluding Asp59) and the 2 His residues in the unfolded SH3 domain. Their two main conclusions were: (1) overall, pK_a shifts for the Asp and Glu residues from model-compound values were small, and (2) Glu2 and Asp8 had the largest pK_a shifts, but even these were only 0.32 and 0.25 pH units, respectively. The pK_a's predicted by the Gaussian-chain model (see Table 1) were in good accord with these findings. Specifically, the predicted pK_a's for the 13 Asp and Glu residues had an average downward shift of just 0.09 pH units. Glu2 and Asp8 had the largest predicted shifts, which were 0.20 and 0.34 pH units, respectively. Predicted and measured results for the pK_a of His7 were reasonably close. For His58, an upward shift in pK_a was correctly predicted, but the shift was much smaller than the measured result.

Kay and co-workers[4] suggested that the downward pK_a shifts of Glu2 and Asp8 were due to neighboring positive charges on the N-terminal and His7, respectively. Gaussian-chain model calculations could provide fuller explanations for these two pK_a shifts and for the minimal shifts of the other 11 Asp and Glu residues. For Glu2 and Asp8, the N-terminal and His7 (respectively) are the closest charges along the sequence. Their interactions with Glu2 and Asp8, which are nonspecific within the Gaussian-chain model, may therefore dominate the contributions to the latter residues' pK_a



$$u_{ij} = q_i q_j e^{-\kappa r}/\epsilon r$$
$$p(r) = 4\pi r^2 (3/2\pi d^2)^{3/2} \exp(-3r^2/2d^2)$$
$$d = b_{\mathrm{eff}} |j - i|^{1/2} + s$$

Figure 1. The Gaussian-chain model.[2b] At a particular distance r, residues i and j (with charges q_i and q_j) interact via a Debye−Hückel potential u_{ij} (κ: Debye screening length; ϵ: dielectric constant of water). The chain samples numerous conformations, as a result r has a Gaussian distribution $p(r)$. The root-mean-square distance d increases with the chain separate $|j - i|$ ($b_{\mathrm{eff}} = 7.5$ Å: effective bond length; $s = 5$ Å: constant shift distance). The average of u_{ij} over the Gaussian distribution of r gives W_{ij} in eq 1.

Table 1. Effects of Residual Charge−Charge Interactions

	pK_a		
group	predicted[a]	experimental[b]	G_u^{intc}
N-terminal	7.70−7.64		
Glu2	4.18−4.23	4.08 ± 0.02	−0.14
Lys6	10.8−10.7		−0.14
His7	6.77−6.71	7.07 ± 0.05	−0.15
Asp8	3.61−3.71	3.75 ± 0.03	−0.15
Asp14	3.90−3.94	3.97 ± 0.02	0.27
Asp15	3.90−3.95	3.99 ± 0.02	0.32
Glu16	4.42−4.43	4.44 ± 0.03	0.23
Arg20	12.5−12.3		−0.10
Lys21	10.6−10.5		−0.07
Lys26	10.8−10.6		−0.23
Glu31	4.49−4.48	4.37 ± 0.03	0.32
xxAsp32	3.97−3.99	4.05 ± 0.02	0.41
Asp33	3.95−3.97	4.13 ± 0.03	0.35
Tyr37	9.88−9.77		
Arg38	12.8−12.6		−0.47
Glu40	4.35−4.37	4.42 ± 0.03	0.15
Asp42	3.81−3.87	4.05 ± 0.02	0.17
Lys44	11.1−10.9		−0.55
Glu45	4.27−4.29	4.45 ± 0.02	0.05
Tyr52	9.91−9.81		
Glu54	4.29−4.32	4.28 ± 0.03	0.02
Lys56	10.9−10.8		−0.34
His58	6.94−6.85	7.83 ± 0.07	−0.35
Asp59	3.83−3.87		0.07
C-terminal	3.28−3.34		

[a] The two numbers were calculated for ionic strengths of 0.05 and 0.1 M. The solvent temperature was 5 °C (at which $\epsilon = 86$ and $\kappa = I^{1/2}/3.08$ Å^{-1} if the ionic strength I is in units of M). Model-compound pK_a's were: Asp, 4.0; Glu 4.4; His, 6.6; Tyr, 9.6; Lys 10.4; Arg, 12.0; N-terminal, 7.5; and C-terminal, 3.5. [b] From Kay and co-workers.[4] The buffer was 0.05 M sodium phosphate. [c] Calculated for pH 6.0 and ionic strength of 0.05 M by the Gaussian-chain model.[2b]

10.1021/ja0298491 CCC: $25.00

shifts. When the N-terminal (or His7) was neutralized, the predicted pK_a shift of Glu2 (or Asp8) decreased from 0.20 to 0.08 (or from 0.34 to 0.17). However, along the sequence Lys6 is the second closest charge to both Glu2 and Asp8 and could also contribute to their pK_a shifts. If Lys6 were neutral, the predicted pK_a shifts of Glu2 and Asp8 decreased from 0.20 to 0.14 and from 0.34 to 0.23, respectively.

The Gaussian-chain model predicts significant downward pK_a shifts for Asp and Glu residues by positive charges close along the sequence.[5] The minimal pK_a shifts of the remaining 11 Asp and Glu residues may be attributed to the lack of close neighboring positive charges or the countering acts of positive charges and other Asp or Glu at neighboring positions. Of the remaining 11 Asp and Glu residues, six (Asp14, Asp15, Glu16, Glu31, Asp32, and Asp33) have another Asp or Glu as their nearest neighboring charges along the sequence, and three (Glu40, Asp42, and Asp59) have both another Asp or Glu or the C-terminal and an Arg, Lys, or His as their nearest neighboring charges with equal sequence separations. Glu45 has Lys44 as the nearest neighboring charge but has Asp42 and Glu40 as next nearest neighbors. Glu54 has Lys56 and His58 as nearest neighbors but has Asp59 and the C-terminal as next nearest neighbors. It should also be noted that the drkN SH3 domain has four more Asp and Glu residues than Lys, Arg, and His residues. The more distant excess Asp and Glu residues may also counter the effects of neighboring positive charges on the pK_a shifts of Glu45 and Glu54.

The minimal pK_a shifts found for the Asp and Glu residues on unfolded drkN are to be contrasted with significantly larger shifts found on other proteins in the unfolded state. For example, Fersht and co-workers[1b] observed an average shift of 0.3−0.4 pH units for the 12 Asp and Glu residues in unfolded barnase, which was reproduced by both the Gaussian-chain model[2b] and the nativelike model of Elcock.[2a] For unfolded staphylococcal nuclease (SNase), the Gaussian-chain model predicted an average downward shift of 0.34 pH units for the 20 Asp and Glu residues, and the resulting pH dependence of the total charge was in good agreement with experimental measurements of Whitten and Garcia-Moreno (ref 1e).[3b] drkN can be differentiated from barnase and SNase by the sequence features noted above, namely the clustering of Asp and Glu along the sequence (most notably, Asp14/Asp15/Glu15 and Glu31/Asp32/Asp33) and the excess of Asp and Glu over Lys, Arg, and His. Most Asp and Glu residues in barnase and SNase are adjacent to positively charged Lys, Arg, and His residues along the sequence, and these two proteins have four and twelve, respectively, more Lys, Arg, and His residues than Asp and Glu residues. The contrasting results on pK_a shifts for drkN and for barnase and SNase can both be explained by nonspecific charge−charge interactions that decrease with sequence separation.

In the folded structure, Asp14, Asp15, Glu16, Glu31, Asp32, and Asp33 are in close proximity and can potentially interact with Arg residues of target peptides.[6] The proximity of the negative charges destabilizes the folded structure. However, by having these charges clustered along the sequence, the same charges also destabilize the unfolded state. In a previous study of a thermophilic cold shock protein,[7] we have proposed destabilization of the unfolded state by unfavorable charge−charge interactions as a mechanism for increasing folding stability. Had the negative charges assembled in the folded structure of the SH3 domain come from distant parts of the sequence, the folding stability would be significantly lower. Table 1 lists the contributions (G_u^{int}) to the free energy of the unfolded SH3 by electrostatic interactions between each ionizable group and the rest of the protein at pH 6. Each of the six Asp and Glu residues in the two contiguous sequence stretches was found to destabilize the unfolded state by 0.23 to 0.41 kcal/mol.[8] By assembling the negative charges into proximity in the folded structure from two contiguous stretches of the sequence, the desired binding surface for target peptides can be formed while potential adverse effects on folding stability are minimized.

Experiments have indicated that the unfolded drkN SH3 domain has residual structures.[9] The nonspecific charge−charge interactions on which the Gaussin-chain model is based do not necessarily conflict with the presence of residual structures. Residual side-chain contacts mostly involve local hydrophobic interactions.[9,10] As the unfolded protein rapidly interconverts between different conformations, clustering of certain hydrophobic side chains can form repeatedly (giving rise to residual structures), but there is no evidence that specific interactions between charged residues occur (perhaps due to desolvation and entropic costs). However, because of their long-range nature, even nonspecific interactions between charged residues can give rise to significant electrostatic effects. Consideration of such electrostatic effects will lead to fuller characterizations of the unfolded state.

In summary, the individual pK_a's of Asp and Glu residues in the unfolded drkN SH3 domain predicted by the Gaussian-chain model were in good agreement with experimental results. The clustering of Asp and Glu charges along the sequence has been implicated in limiting pK_a shifts and in contributing to the folding stability by destabilizing the unfolded state. The present critical test and previous extensive tests demonstrate that the Gaussian-chain model yields reasonable quantitative results for and suggests a realistic picture of electrostatic effects in the unfolded state.

Acknowledgment. This work was supported in part by NIH Grant GM58187.

References

(1) (a) Swint-Kruse, L.; Robertson, A. D. *Biochemistry* **1995**, *34*, 4724−4732. (b) Oliverberg, M.; Arcus, V. L.; Fersht, A. R. *Biochemistry* **1995**, *34*, 9424−9433. (c) Tan, Y. J.; Oliverberg, M.; Davis, B.; Fersht, A. R. *J. Mol. Biol.* **1995**, *254*, 980−992. (d) Kuhlman, B.; Luisi, D. L.; Young, P.; Raleigh, D. P. *Biochemistry* **1999**, *38*, 4896−4903. (e) Whitten, S. T.; Garcia-Moreno E. B. *Biochemistry* **2000**, *39*, 14292−14304. (f) Pace, C. N.; Alston, R. W.; Shaw, K. L. *Protein Sci.* **2000**, *9*, 1395−1398.

(2) (a) Elcock, A. H. *J. Mol. Biol.* **1999**, *294*, 1051−1062. (b) Zhou, H.-X. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 3569−3574. (c) Kundrotas, P. J.; Karshikoff, A. *Protein Sci.* **2002**, *11*, 1681−1686. Previously Dill and co-workers have developed a porous-sphere model (Stigter, D.; Alonso, D. O.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 4176−4180).

(3) (a) Zhou, H.-X. *Biochemistry* **2002**, *41*, 6538−6533. (b) Zhou, H.-X. *Biophys. J.* **2002**, *83*, 2981−2986.

(4) (a) Tollinger, M.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 5714−5717. (b) Tollinger, M.; Crowhurst, K. A.; Kay, L. E.; Forman-Kay, J. D. *P. Natl. Acad. Sci. U.S.A.* **2003**, in press.

(5) pK_a perturbations of an Asp residue at the *i*th position by an Asp or Lys at the $(i + 1)$th position in an unfolded chain are different both in direction and in magnitude. A neighboring Asp contributes an upward shift of 0.09 pH units, whereas a neighboring Lys contributes a downward shift of 0.19 pH units.

(6) Zhang, O.; Forman-Kay, J. D. *Biochemistry* **1995**, *34*, 6784−6794.

(7) Zhou, H.-X.; Dong, F. *Biophys. J.* **2003**, in press.

(8) Note also that, except for Glu2 and Asp8, all Asp and Glu residues show unfavorable interactions with the rest of the unfolded chain, whereas all Lys, Arg, and His show favorable interactions.

(9) Crowhurst, K. A.; Tollinger, M.; Forman-Kay, J. D. *J. Mol. Biol.* **2002**, *322*, 163−178.

(10) (a) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw; S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719−1722. (b) Baldwin, R. L. *Science* **2002**, *295*, 1657−1658.

JA0298491