Protein stability: Electrostatics and compact denatured states

(molten globule/statistical mechanics/protein folding/heteropolymers)

DIRK STIGTER, DARWIN O. V. ALONSO, AND KEN A. DILL

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143

Communicated by Marshall Fixman, February 1, 1991 (received for review October 1, 1990)

Globular proteins can be denatured by ABSTRACT changing pH and ionic strength. Much recent evidence has led to the surprising conclusion that there are two acid-denatured states: one highly unfolded and the other more compact, sometimes called the "molten globule." Here we describe a molecular theory for electrostatic stability of globular proteins based on the properties of the constituent amino acids: oil/water partition coefficients, pK values of the titratable groups, and their temperature dependences. Predicted denaturation temperatures vs. pH are in good agreement with experiments of other workers on myoglobin. The theory also predicts two populations of denatured species, one open and the other more compact, with densities in the range found experimentally for molten globular states. In addition, it predicts a phase diagram (stability vs. pH, ionic strength) in good agreement with experiments of Goto and Fink [Goto, Y. & Fink, A. L. (1989) Biochemistry 28, 945-952; and Goto, Y. & Fink, A. L. (1990) J. Mol. Biol. 214, 803-805]. The well-known salt destabilization of myoglobin has been generally considered evidence for ion pairing, but the present theory, based on smeared charge repulsion, explains the salt destabilization at low pH without ion pairing. In addition, for myoglobin the theory predicts salt stabilization at high pH, as observed for β -lactamase by Goto and Fink.

Among the earliest known agents affecting protein stability and aggregation were salts (1) and acids and bases (2). The earliest understood electrostatic effect on proteins was the pH titration behavior of native molecules (3). Better theoretical models are now available for understanding the electrostatic interactions in native proteins (4–6). However, because protein stability involves the free energy difference between folded and unfolded states, theory for electrostatic effects on stability must also treat the unfolded conformations. This is the purpose of the present work.

In this paper we develop a model for the stability of globular proteins that takes into account the hydrophobic interaction, chain conformational entropy, and electrostatic effects. We apply it to the prediction of the phase diagram of the stable states of proteins as a function of temperature, pH, and ionic strength. Our model is constructed from two components. The first component is a mean-field lattice model recently developed to describe the stabilities of uncharged proteins; details are given elsewhere (7, 8). It assumes a protein is a random heteropolymer, in which compact states are favored by the hydrophobic interaction and disfavored by chain conformational entropies. The second component, described in more detail below and in refs. 9 and 10, is a smeared-charge Poisson-Boltzmann theory of the electrostatics of folded and unfolded states. We combine here these two components to calculate electrostatic effects on protein stability, and we test these predictions against experiments.



FIG. 1. Imaginary reversible process of protein folding for the purpose of evaluating the free energy difference per molecule, $g_u - g_c$ between unfolded and folded state.

Free Energy Evaluation

To evaluate the free energy difference between native and denatured states, we use the fictitious reversible thermodynamic cycle shown in Fig. 1. The process of folding the charged protein involves first discharging the denatured protein, then collapsing it into a compact state with random arrangement of the nonpolar and polar residues throughout the core, then ordering the residues to form a nonpolar core, then finally charging the ionic residues at the surface of the folded protein.

Two steps of this thermodynamic cycle do not involve electrostatics. For those steps, the free energy depends on the number of chain (lattice) segments, the fraction of residues, Φ , that are nonpolar, and the thermodynamic quantities for the transfer of a typical nonpolar amino acid side chain from a hydrophobic environment to water, for which we use the values $\Delta H_0 = 0$, $\Delta S_0 = -6.7 \text{ cal}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$, $\Delta C_p = 55$ cal·K⁻¹·mol⁻¹ at 25°C (1 cal = 4.18 J) (7). This part of the theory predicts the stabilities of uncharged proteins as a function of temperature, in reasonable agreement with experiments on small single-domain proteins (see figure 5 of ref. 7).

We assume that the free energies are sums of the electrostatic contributions to the chain, described in this section, and the hydrophobic contributions referred to above. The charged native protein is described by a solid sphere with radius R_f , dielectric constant 3.5, and a uniform surface charge. The Poisson-Boltzmann equation is applied to the surrounding salt solution with the provision that the ionization of the groups at the protein surface be consistent with the surface potential, ψ , of the protein. This consistency for

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

groups of type *i* with intrinsic proton binding constant $k_i = 10^{-pK_i}$ and degree of ionization α_i is given by

$$q_i \ln \frac{\alpha_i}{1 - \alpha_i} = \ln[\mathrm{H}^+] + \ln k_i - \frac{e\psi}{kT}, \qquad [1]$$

where $q_i = +1$ for basic groups and -1 for acidic groups, and $[H^+] = 10^{-pH}$ is the proton concentration in the bulk solution. Eq. 1 is well known from the theory of polyelectrolyte titration (11). The first two terms on the right-hand side simply describe Langmuir type binding of protons to independent sites. The binding is biased by electrostatic interactions between ionized groups, as given by the last term of Eq. 1.

With n_i groups of type *i* per molecule, the electrostatic free energy, $g_f(el)$, per molecule of the native protein is

$$g_f(el) = \int_0^Z e\psi dz - Ze\psi + kT \sum_i n_i \ln(1 - \alpha_i), \quad [2]$$

where the total charge per protein molecule is

$$Ze = e \sum_{i} q_{i} \alpha_{i} n_{i}.$$
 [3]

The integral term in Eq. 2 represents the electrostatic interaction, the term $-Ze\psi$ is the chemical free energy of proton binding, and the last term derives from the entropy of distribution of charged and uncharged sites. As discussed elsewhere (9) in detail, Eq. 2 can be derived as the work required to reversibly charge the native protein, the last step in Fig. 1. It is noteworthy that the results for $g_f(el)$ from Eq. 2 are numerically the same, within errors of computation, as those from a completely different expression derived with the polynomial method of proton binding (9).

The electrostatics of the unfolded protein is treated by modification of the Hermans-Overbeek porous sphere model of polyelectrolytes (12). The unfolded protein has radius R_{μ} . The dielectric constant and the concentration of ionic residues and of foreign salt inside the porous sphere are averages that depend linearly on the uniform chain density, $(R_f/R_u)^3$. The distributions inside and outside the porous sphere of the mobile small ions are treated through the Poisson-Boltzmann equation. Eq. 1 is applied at each radial distance r between the center and the surface of the porous sphere for each type *i* of ionizable group, with the appropriate values of $\alpha_i(r)$ and $\psi(r)$. The free energy is obtained as in Eq. 2, where ψ and α_i are now suitable averages of $\psi(r)$ and $\alpha_i(r)$ over the porous sphere volume. The experimental input required here is the ionic strength, pH, temperature of the solution, the radii R_{f} and R_{μ} of the model spheres, the number n_i of the ionizable groups, and their pK_i values as a function of temperature. Details of the numerical solution of the Poisson-Boltzmann equation and of the electrostatic free energy calculations have been given earlier (10).

Folding of Myoglobin

More thermodynamic and titration information is available for myoglobin than for other proteins, so here we report the comparison of the model with experiments on myoglobin. We have taken the pK_i values from the titration experiments of Breslow and Gurd (13) at 25°C, shown in Table 1. The decrease of the pK_i values with increasing temperature is obtained from the reported heats of ionization, 10.7 kcal·mol⁻¹ for amino groups, 7.5 kcal·mol⁻¹ for histidine, and 2.5 kcal·mol⁻¹ for carboxylic acids (13). The numbers n_i of the various ionic groups in Table 1 correspond to the composition of myoglobin, except for histidine. Since six histidines in native myoglobin are buried in the hydrophobic core and not

Table 1. Ionizable groups of metmyoglobin at 25°C

Group	n _i	pK _i
ε-Amino	19	10.60
α-Amino	1	7.80
Histidine (globular)	6	6.62
Histidine (unfolded)	12	6.48
Hemic acid	1	8.90
Carboxyl	23	4.40

Data are from Breslow and Gurd (13).

titrated (see Table 1), we exclude those from the count of titratable groups in the folded state.

The folding equilibrium depends on the difference between the (total) free energy g_{μ} of the unfolded state and g_f of the folded state, where g_{μ} is a function of the variable radius R_{μ} of the unfolded protein. We take $R_f = 18$ Å for native myoglobin (13). At 25°C, pH 4, and ionic strength 0.01 M, Fig. 2 illustrates the various contributions to the free energy of unfolding, $g_u - g_f$, as a function of the radius of the unfolded state, from the minimum value $R_u = 18$ Å for the most compact form to 90 Å for very open coils. Since g_f is constant, an individual curve in Fig. 2 shows how the free energy of the unfolded state, g_u , depends on R_u . A minimum in the free energy difference $g_u - g_f$ marks the radius of a stable unfolded state. Curve a in Fig. 2 shows the configurational entropy contribution to the free energy difference between unfolded and folded states. It gives the free energy difference for a hypothetical noninteracting coil, as for a chain in a good solvent, where denatured myoglobin would have a radius of about 55 Å. The free energy curve has a shallow minimum around this radius. The double-arrowed horizontal line marks the range of R_{μ} over which g_{μ} varies by 1 kT. This indicates the large fluctuations in coil size predicted under these hypothetical conditions. Curve b shows the charged coil. From the shift of the free energy minimum, from $R_u = 55$ Å to about $R_u = 75$ Å, we conclude that the coil has expanded, as expected from electrostatic repulsion between the charges fixed to the protein chain. On the other hand, if instead of charges we add hydrophobic interactions to the coil of curve a, we obtain curve c, and we observe that attraction contracts the coil to a radius of only about 21 Å. with much smaller fluctuations. Finally, when we add both hydrophobic and charge interactions, we arrive at curve d, the prediction for the total free energy of unfolding at pH 4. Here the predicted radius of the unfolded state is about 65 Å.



FIG. 2. Contributions to the free energy of the unfolded state per molecule, g_u , relative to the folded state, g_f , as a function of the radius R_u of the unfolded state for myoglobin at 25°C, ionic strength 0.01 M, and pH 4. Curve a, from conformational entropy. Curve b, from conformational entropy and electrostatic free energy. Curve c, from conformational entropy and hydrophobic effects. Curve d, from total free energy = conformational entropy, hydrophobic effects, and electrostatic free energy.

Fig. 3 presents a series of curves for the (total) free energy of unfolding at different pH values, including curve d in Fig. 2 for pH 4, now on a different scale. Any two curves differ in the value of $g_{\mu} - g_{f}$ because the pH affects the free energy of the native state as well as the denatured states. A positive value of $g_{\mu} - g_{f}$ at the stable radius of the minimum in the curve implies the folded state is most stable; a negative value implies the unfolded state is most stable; $g_{\mu} - g_f = 0$ of the stable state identifies a denaturation midpoint. Fig. 3 shows that at pH 7, near the isoelectric pH, the native state is predicted to be stable by about 13 kT. At pH 4, the most stable state is an unfolded molecule of large radius (65 Å). The midpoint of denaturation at 0.01 M ionic strength and 25°C is at pH 4.662; at that pH the native state is in equilibrium with a compact unfolded state of radius 23 Å. At pH 4.28, the theory predicts two stable unfolded states: one with $R_{\mu} = 24$ Å and the other with $R_{\mu} = 55$ Å. We discuss the folding equilibrium first and return to the nature of the denatured states later.

We have determined denaturation as a function of pH, ionic strength, and temperature by fixing two of these variables and computing the value of the third from the condition $g_{\mu} - g_f = 0$ by an iteration procedure. Fig. 4 shows the predicted denaturation temperature, T_m , as a function of pH for myoglobin in aqueous solutions of ionic strength 0.01 M. Denaturation of the relatively uncharged molecule is predicted at 82°C and at -20°C (7). In the intermediate pH range, from about 6 to 8, the protein has maximum stability because it has the least charge repulsion. In acids or bases, the protein is more highly charged, which favors unfolding. The melting loop is inclined with respect to the T_m and pH coordinates; this asymmetry arises from the temperature dependence of the pK_i values. The melting loop also depends on the hydrophobic composition, Φ ; results of a hypothetical change in composition are shown by the dotted line in Fig. 4. The burial of titratable groups into the protein core also affects the melting loop (compare broken and unbroken curves in Fig. 4). Burying histidines destabilizes the folded protein because they are unable to ionize, so g_f increases. This has no effect on the alkaline side of the melting loop, since the basic histidines are not ionized at high pH, irrespective of whether they are buried. Therefore the histidines make no contribution to electrostatic free energy at high pH. The theory is in reasonable agreement with the experimental data of Privalov et al. (14), also shown in Fig. 4. At low pH, there is a difference of nearly one pH unit; this difference is probably attributable to the neglect of the heme group in the calculations. The heme must increase the stability of the folded protein, since it spontaneously combines (see below).



FIG. 3. Free energy of the unfolded state, g_u , relative to the folded state, g_f , as a function of the radius R_u of the unfolded state for myoglobin at 25°C, ionic strength 0.01 M, and pH as indicated.



FIG. 4. Denaturation temperature T_m of myoglobin in aqueous solutions versus pH. Unbroken curve, theory for ionic strength 0.01 M with parameters described in text. Dotted curve, theory with hydrophobic fraction of myoglobin chain reduced by 10%. Broken curve, theory with 12 instead of 6 histidines at surface of folded state. •, Experiments of Privalov *et al.* (14) on sperm whale metmyoglobin in 0.1 M glycine/NaOH (high pH) and in 0.01 M piperazine/HCl (low pH).

Fig. 5 shows the predicted phase diagram of ionic strength vs. pH for myoglobin at 25°C, compared with the experimental data of Friend and Gurd (15) and Goto and Fink (18). These two experiments differ by the presence of the heme group; the experiments of Goto and Fink are on apomyoglobin (without the heme), corresponding to our calculations. The agreement is good. A surprise is that the observed positive slope of the phase boundary of ionic strength with pH is predicted by the theory. The traditional view (15, 17) has held that at fixed pH, if a protein is destabilized by increased salt, this implies the electrostatics is dominated by ion pairing. In that view, ion pairs should stabilize the folded protein and shielding by salt should weaken this stabilization. However, the present model shows that salt-induced destabilization can arise also from nonspecific charge repulsions and changes in ionization. In this case, however, the salt effects are larger on the denatured than the native state. Eq. 2 shows that many different terms contribute to the electrostatic free energy. The influence of salt on each separate term is readily understood (9). The salt effect on the sum of all (positive and negative) contributions to $g_u - g_f$, however, is a complex balance. An unexpected prediction of the model is that both acid and alkaline phase boundaries for myoglobin should have positive slope (see Fig. 5). The positive slope for denaturation at high pH is in agreement with data of Goto and Fink on β -lactamase (16); ion pairing would incorrectly predict a negative slope. Thus, the present model appears to be consistent with experimental results that would not be predicted by either of the two classical expectations. The model predicts that both acid and alkaline slopes of the phase



FIG. 5. Folding transitions of myoglobin in the ionic strength-pH plane at 25°C. Curves, present theory; ●, experiments on sperm whale ferrimyoglobin by Friend and Gurd (15); ■, experiments on horse apomyoglobin by Goto and Fink (18).

diagram should be positive for myoglobin. Ion pairing would predict that the acid phase boundary should have a positive slope but the alkaline boundary should have negative slope. Alternatively, from simpler models of classical charge repulsions, a negative acid slope and a positive alkaline slope would be expected.

Denatured States

The theory predicts two stable unfolded populations under some conditions (see Fig. 3, pH 4.28, for example), arising from a subtle balance between long-range and short-range interactions. The open denatured state is predicted to be stabilized by long-range interactions and the compact denatured state is predicted to be stabilized by short-range interactions. The highly unfolded state is restrained from further expansion by the chain elastic entropy and from further contraction by increased charge repulsion. The compact unfolded state is restrained from expansion by the hydrophobic free energy and from contraction by the excludedvolume chain configurational entropy. Therefore, at low ionic strength, denatured states of intermediate compactness are unstable because hydrophobic attractions make them smaller or electrostatic repulsions make them larger. The pH 4.28 curve in Fig. 3 shows the free energy barrier between the two states at the transition. When this barrier is larger than kT the transition between the compact and the highly unfolded states is first order. The height of the transition barrier depends on temperature and on ionic strength. Under certain conditions the barrier disappears altogether and the two denatured states merge at an intermediate (average) coil radius. The Boltzmann-averaged radius, $\langle R \rangle$, is shown in Fig. 6. At pH 5.5, most molecules are native (point 3 on the curve), with radius 18 Å. For 0.1 M salt, decreasing the pH to 4.8 leads to a plateau radius $\langle R \rangle = 24$ Å, and to still lower pH leads to further increases of the unfolded radius.

The predicted denatured state radii can be compared with measurements of the intrinsic viscosity of apomyoglobin by Griko *et al.* (19). This is a very sensitive test because the theory predicts that the free energy minima are broad: small changes in free energy correspond to relatively large variations in unfolded radius. Griko *et al.* (19) have determined the intrinsic viscosity of apomyoglobin in 10 mM sodium acetate buffer solutions between pH 2 and 5.5 at 2°C, 27°C, and 50°C. Their $[\eta]$ vs. pH curves show two transitions similar to our $\langle R \rangle$ vs. pH curve for 0.1 M in Fig. 6. From their high plateau values of $[\eta]$ for pH < 3 we derive approximate coil radii of 35–43 Å, about half the predicted value for $\langle R \rangle$ of about 80 Å



FIG. 6. Average radius of protein, $\langle R \rangle$, versus pH for myoglobin at 25°C and ionic strength as indicated. State 1, low density unfolded; state 2, high density unfolded; state 3, folded.

in Fig. 6. Thus the present model may overestimate the electrostatic coil expansion. There are two obvious errors inherent in the present smeared charge model: the inclusion of the self-energy of the smeared charges in g_u and g_f , and the omission of the attraction between discrete protein charges of opposite sign. Both errors increase the predicted electrostatic coil expansion.

Fig. 7 shows the predicted phase diagram of myoglobin (Lower) compared with the experimental results of Goto and Fink (Upper) (16, 18). Goto and Fink have identified three states: N (native), A (compact denatured), and U_A (acid unfolded). The phase boundaries they observe spectroscopically are relatively broad, centered where the points are shown in the figure. The theory shows a series of different radii. We constructed the lines of constant $\langle R \rangle$ in Fig. 7 as follows. From a series of curves of the type shown in Fig. 6, for many different salt concentrations, a horizontal line was used to identify the series of (salt concentration, pH) combinations that lead to a given $\langle R \rangle$. From these points we constructed curves of constant $\langle R \rangle$ in the ionic strength-pH plane in Fig. 7. We note only that the shapes of the A to U_A phase boundaries closely resemble those of the experiments; as noted above, the predicted quantitative values of $\langle R \rangle$ are probably too large.

The compact unfolded state predicted by the theory bears considerable resemblance to that recently identified experimentally, sometimes referred to as the "molten globule" state (16, 18, 20–22): the radius is only slightly greater than that of the native state and, because the amount of secondary structure in proteins increases strongly with their compactness (23, 24), these compact molecules should have considerable helix and sheet.

The model predicts an interesting asymmetry for aciddenatured states compared to alkaline-denatured states. The radius of a denatured state should be greater at the nativedenatured transition than under other denaturing conditions. Fig. 8 shows the density of the denatured state at the transition, plotted vs. ionic strength for acid transitions and base transitions at 25°C. Because the different ionic groups have different titration behavior at low and high pH, the radii of the denatured molecules at the transition are predicted to have quite different salt dependence, depending on whether the protein is acid denatured or base denatured. At low salt,



FIG. 7. Phase diagram of apomyoglobin in the ionic strength-pH plane at 25°C. (*Upper*) Experiments on horse apomyoglobin by Goto and Fink (18). Native folded state is labeled N, acid unfolded state U_A , and intermediate state A. (*Lower*) Theory. Curves of constant density of the unfolded state are shown in the lower left.



FIG. 8. Density $(R_f/R_u)^3$ of unfolded state at transition versus ionic strength for acid and alkaline denaturation of myoglobin at 25°C. Density of B state of alkaline denatured β -lactamase (16) is shown by \bullet .

base-denatured molecules should be much more expanded than acid-denatured molecules. However, at high salt, both forms of denaturation should lead to relatively compact denatured molecules. The one experimental data point on Fig. 8 is from Goto and Fink (16) for state B (compact denatured state) of β -lactamase at high pH and high ionic strength; it is consistent with our predictions for myoglobin.

Conclusions

We have developed a model for protein stability. It has three components: (i) a temperature-dependent hydrophobic contact interaction, parameterized on small molecule oil/water transfer studies. (ii) conformational entropies due to elastic (local in the chain sequence) and excluded volume interactions (nonlocal in the sequence), and (iii) Poisson-Boltzmann treatment of the electrostatic interactions of the protein as a porous charged sphere. In the model, the free energy of folding depends on the protein chain length n, fraction of nonpolar residues Φ , and the numbers, pK_i values, and temperature dependences of the titratable groups. Predictions are in reasonable agreement with experiments. The theory predicts the denaturation temperature/pH and the pH/ionic strength phase diagrams for apomyoglobin. It correctly predicts positive slopes for the ionic strength/pH phase boundaries for both base- and acid-denatured apomyoglobin; neither ion-pairing nor classical models predict this. This shows the importance of the considerable dependence of the denatured state on electrostatics and the importance of an electrostatic entropy of charge distribution, often neglected. Finally, the theory predicts the existence of two classes of denatured states: one is highly unfolded, the other is compact. This model predicts that the compact denatured state is stabilized by a high degree of hydrophobic clustering, but it differs from the native state in that it has no hydrophobic "core."

We thank Prof. Anthony Fink (University of California, Santa Cruz) for providing his data prior to publication, and the National Institutes of Health, Defense Advanced Research Planning Agency University Research Initiative, and Pew Foundation for financial support.

- 1. Hofmeister, F. (1888) Naunyn-Schmiedebergs Arch. Exp. Pathol. Pharmakol. 24, 247-260.
- 2. Chick, H. & Martin, C. J. (1911) J. Physiol. 43, 1-27.
- 3. Linderstrom-Lang, K. U. (1924) C. R. Trav. Lab. Carlsberg 15, 73-98.
- Tanford, C. & Kirkwood, J. G. (1957) J. Am. Chem. Soc. 79, 5333-5339.
- 5. Gilson, M. K. & Honig, B. H. (1988) Proteins 4, 7-18.
- 6. Matthew, J. B. & Gurd, F. R. N. (1986) Methods Enzymol. 130, 413-436.
- Dill, K. A., Alonso, D. O. V. & Hutchinson, K. (1989) Biochemistry 28, 5439-5449.
- 8. Dill, K. A. (1985) Biochemistry 24, 1501-1509.
- 9. Stigter, D. & Dill, K. A. (1989) J. Phys. Chem. 93, 6737-6743.
- 10. Stigter, D. & Dill, K. A. (1990) Biochemistry 29, 1262-1271.
- Katchalsky, A. & Gillis, J. (1949) Rec. Trav. Chim. Pays-Bas 68, 879-897.
- 12. Hermans, J. J. & Overbeek, J. Th. G. (1948) Rec. Trav. Chim. Pays-Bas 67, 761-776.
- 13. Breslow, E. & Gurd, F. R. N. (1962) J. Biol. Chem. 237, 371-381.
- Privalov, P. L., Griko, Yu. V., Venyaminov, S. Yu. & Kutyshenko, V. P. (1986) J. Mol. Biol. 190, 487–498.
- 15. Friend, S. H. & Gurd, F. R. N. (1979) Biochemistry 18, 4612-4619.
- 16. Goto, Y. & Fink, A. L. (1989) Biochemistry 28, 945-952.
- 17. Kauzmann, W. (1959) Adv. Protein Chem. 14, 1-63.
- 18. Goto, Y. & Fink, A. L. (1990) J. Mol. Biol. 214, 803-805.
- Griko, Yu. V., Privalov, P. L., Venyaminov, S. Yu. & Kutyshenko, V. P. (1988) J. Mol. Biol. 202, 127–138.
- 20. Ptitsyn, O. B. (1987) J. Protein Chem. 6, 273-293.
- 21. Kuwajima, K. (1989) Proteins 6, 87-103.
- 22. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989) Biochemistry 28, 7-13.
- 23. Chan, H. S. & Dill, K. A. (1990) J. Chem. Phys. 92, 3118-3135.
- 24. Chan, H. S. & Dill, K. A. (1990) Proc. Natl. Acad. Sci. USA 87, 6388-6392.