Three-State Analysis of Sperm Whale Apomyoglobin Folding†

Doug Barrick and Robert L. Baldwin*

Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, California 94305

Received November 17, 1992; Revised Manuscript Received February 1, 1993

ABSTRACT: We give a quantitative description of the urea- and acid-induced transitions of apomyoglobin at 0 °C and 2 mM sodium citrate. Our data consist of two series of unfolding curves: (1) acid-induced unfolding carried out in the presence of various concentrations of urea and (2) urea-induced unfolding at various pH values. A three-state equation is derived which relates the stability of three different conformations of apomyoglobin (native [N], unfolded [U], and intermediate [I]) as a function of urea and of pH. This equation fits our data reasonably well. The parameters which give the best fit have both thermodynamic and structural implications for N, I, and U. Specifically, I is closer in Gibbs energy to U than to N, indicating that side-chain packing results in much of the stability of native protein structure. The equilibria between N and I and between I and U are equally sensitive to urea, suggesting that much of the surface of I is inaccessible to solvent. The acid-induced transition in which N unfolds can be described as the result of titration of approximately two histidines with low pK_as in N. Under physiological conditions (neutral pH, no urea) I is the most stable non-native conformation.

Sperm whale apomyoglobin (ApoMb[1]; myoglobin with its heme group removed) adopts a partly folded intermediate conformation (I) in mild denaturing conditions (Kirby & Steiner, 1970; Griko et al., 1988; Hughson et al., 1990; Goto & Fink, 1990). At low temperature and ionic strength, a decrease in pH from 7 to 4 results in a transition from the native (N) state to I; I contains less helical structure than N ($[\Theta]_{222}$ = −15 000 and −19 000 deg·cm^2·dmol^{-1}, respectively) as judged by circular dichroism spectroscopy (CD). The unfolded conformation (U) is not populated until the pH is decreased further. Intrinsic viscosity measurements indicate that I is nearly as compact as the native conformation (Griko et al., 1988). Fluorescence spectroscopy studies have shown an emission maximum at 336 nm for I, midway between that seen in N (331 nm) and U (340 nm, Kirby & Steiner, 1970; Irace et al., 1981), suggesting an intermediate degree of solvent exposure of tryptophan and tyrosine side chains. Hydrogen-exchange studies indicate that three of the helices (A, G, and H) are formed in I (Hughson et al., 1990). Although these helices are closely packed in the crystal structure (Hansen & Schoenborn, 1981), site-directed mutagenesis studies indicate that this close packing is absent in the intermediate conformation and, along with studies of peptide fragments, suggest that some other loose interaction stabilizes the helices (Hughson et al., 1991).

Our objective here is to determine the stability of I relative to N and U. Our data consist of pH titrations in the presence of varying amounts of urea and of urea-induced unfolding transitions at various pHs. We fit a three-state equation, in which urea and pH are treated as independent variables, to our data using nonlinear least-squares methods. The best-fitting parameters, when interpreted within the framework of the three-state model, have implications for both the stability and the structure of I.

EXPERIMENTAL PROCEDURES

ApoMb Preparation. The apoMb used in this study is produced and purified from the *Escherichia coli* strain TB-1 (JM83 *hsd*R (r_k− m_k+) harboring the plasmid pMb413b (Hughson et al., 1991). pMb413b carries a modified version (Hughson et al., 1991) of the synthetic sperm whale Mb gene (Springer & Sligar, 1987), which was kindly provided by B. A. Springer and S. G. Sligar. Expression of sperm whale Mb is carried out essentially as described (Hughson et al., 1991). Typically, 4 L of L-broth (10 g of tryptone, 5 g of yeast extract, and 5 g of NaCl/L) containing 200 mg/L carbenicillin is divided between three 6-L flasks. Each flask is inoculated with 1 mL of an early log-phase culture of pMb413b/TB-1. Flasks are vigorously aerated by shaking for approximately 12 h at 37 °C. As the cultures approach stationary phase, Mb production is monitored by the visible whole-cell carbon monoxide difference spectrum method of Springer and Sligar (1987). When myoglobin production peaks (usually resulting in a difference peak of 0.06–0.07 absorbance units at 420 nm), cells are harvested by centrifugation and stored at −80 °C. Purification is carried out as described (Hughson et al., 1991), with the addition of a sonication step immediately after the 3-h treatment with lysozyme and with the omission of the sonication step of the cell debris following lysate clarification. For the experiments described here, 8 L of cell culture (two growths on consecutive days) is purified at once. Heme is removed from purified Mb by extraction with 2-butanone as described (Teale, 1959; Hughson et al., 1991). Apoprotein concentration is determined by UV absorbance in 6 M guanidine hydrochloride using the method of Edelhoch (1967); stock solutions are between 0.15 and 0.5 mM in apoprotein and are stored at 4 °C for no longer than 36 h prior to use.

CD Studies. Helix content of apoMb is measured at 222 nm ($[\Theta]_{222}$) in an Aviv 60DS spectropolarimeter, using a path length of 10.0 mm. Samples are 1 μM in apoMb. All CD measurements are made at 0 °C, as low temperature destabilizes N (Griko et al., 1988) and allows I to be well resolved. CD samples are made 2 mM in citrate. pH is adjusted by mixing appropriate volumes of 2 mM citric acid and 2 mM sodium citrate, each containing the appropriate amount of urea (U.S. Biochemicals). pH values less than 3

† This work was supported by NIH Grant GM19988. D.B. is a Howard Hughes Medical Institute Predoctoral Fellow.

[1] Abbreviations: apoMb, apomyoglobin; Mb, myoglobin; CD, circular dichroism; $[\Theta]_{222}$, mean residue ellipticity at 222 nm; N, I, and U, the native, intermediate, and unfolded forms of apomyoglobin.

0006-2960/93/0432-3790$04.00/0

require the addition of a small amount of 5 M HCl. For each measurement, apoMb is added to the ice-cold buffered urea solution. The sample is then allowed to equilibrate for 5 min at 0 °C. No change in CD signal is observed after this time. The pH of each sample is determined after measuring ellipticity, using a Metrohm combined glass pH electrode (Model 6.0216.100). When this protocol is used, folding can be fully reversed from strongly denaturing conditions (8 M urea or pH 2–3); however, we find that in refolding by dilution from 4 to 5 M urea, as little as 70% of native ellipticity is recovered (D. Barrick, Ph.D. thesis, in preparation). Since the unfolding transitions we present here are essentially complete under these conditions, we believe our thermodynamic analysis is cogent despite this partial irreversibility.

Fitting the Data. We first construct a model of how acid and urea affect the conformational stabilities of N, I, and U. From this model, a three-state equation is derived (Appendix), which we use to fit our data. Much of the model and the subsequent equations are based on those of Tanford (Aune & Tanford, 1969a,b; Tanford, 1970) and of the multiple ligand binding theory described by Cantor and Schimmel (1980).

Our model for apoMb unfolding contains three conformations: N, I, and U. We adopt the linear extrapolation model of Pace (Greene & Pace, 1974; Schellman, 1978; Pace, 1986; Santoro & Bolen, 1992) to describe the effect of urea on Gibbs energy differences between N, I, and U. The effects of urea are given by two coefficients, $m_{N\rightarrow I}$ and $m_{I\rightarrow U}$, relating Gibbs energy change to urea concentration. The reference conditions we have chosen are 0 urea concentration and high pH (pH greater than the pK_as which are linked to the observed unfolding transitions).

The two observed acid-induced transitions are modeled as the result of two classes of titratable groups. The number of groups within each class is designated n_{H1} or n_{H2}. Here, 1 refers to the high-pH transition, in which N is converted to I and U, while 2 refers to the low-pH transition, which produces U from I. Within each class, and for a given conformation, all groups have identical proton affinities given by, for example, $pK_{a1,N}$ or $pK_{a2,U}$, where 1 and 2 indicate the class. All groups are modeled to titrate independently of each other.

The model then links proton binding of groups in classes 1 and 2 to conformational equilibria in the following way. The first transition, in which N is lost, must result from increased proton binding to I and U compared to N ($pK_{a1,N} < pK_{a1,I}$, $pK_{a1,U}$). Since our data does not provide information on differences between $pK_{a1,I}$ and $pK_{a1,U}$, our model takes $pK_{a1,I}$ and $pK_{a1,U}$ as equal. Since transition 1 is observed in the absence of urea, differential class 1 proton binding persists down to the lowest pH values that N is detected. $pK_{a1,N}$ is assumed to be below this pH limit; thus, our model is constructed so that class 1 proton binding does not occur in N. Although class 1 groups may be protonated in N at pH values lower than those resulting in this first transition, we wish only to extrapolate our data from the transition to neutral pH.

In transition 2, U is formed from N and I as pH is decreased. Thus, U has a higher affinity for protons at class 2 groups than do N and I ($pK_{a2,U} > pK_{a2,N}$, $pK_{a2,I}$). For convenience, we construct our model so that class 2 groups do not bind protons in N and I. Although our model provides a useful working description of the pH dependence of transition 2, it is not likely to be mechanistically correct. Since the structure of I is suspected to be loose and fluctuating (Hughson et al., 1991), transition 2 is more likely to result from a larger number of groups with only small pK_a differences between I and U. Our treatment of the effects of acid on transition 2 are formally consistent with the model of Goto, in which electrostatic repulsion between positively charged groups results in unfolding of I (Goto & Nishikori, 1991).

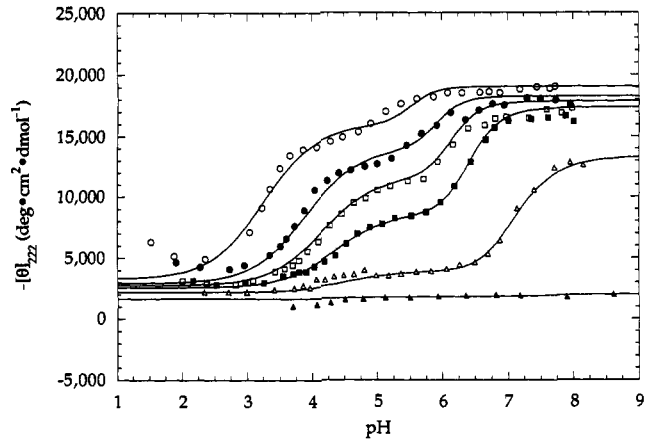

FIGURE 1: pH-induced unfolding transitions of apoMb in urea at 0 °C and 2 mM sodium citrate. Data are from experiment 1 (see text and Table I). Titrations were carried out in 0 (○), 1.0 (●), 1.5 (□), 2.0 (■), 3.0 (△), and 4.5 M urea (▲). Curves result from evaluation of the three-state stability equation (eq 15) at the urea concentrations listed above and using best-fitting parameters listed in Table II, experiment 1.

In our model, effects of urea and of pH are treated as independent. This assumption requires that urea does not change proton binding (pK_as) and that pH changes do not affect binding the of urea to the folded and unfolded conformations. This assumption has been made previously by Aune and Tanford (1969a,b) in analyzing the two-state unfolding of hen egg-white lysozyme and by Kuwajima in his three-state analysis of α-lactalbumin unfolding (Kuwajima et al., 1976).

According to the model, the parameters which describe the relative stability of the apoMb N, I, and U conformations with urea and pH are $m_{N\rightarrow I}$, $m_{I\rightarrow U}$, n_{H1}, n_{H2}, $pK_{a1,I}$ (which is taken as equal to $pK_{a1,U}$), $pK_{a2,U}$, $K_{N\rightarrow I}$, and $K_{I\rightarrow U}$. To fit our data, we also determine the $[\Theta]_{222}$ values for N, I, and U, along with any measurable dependence on urea. In Figure 2, both N and U show measurable changes in $[\Theta]_{222}$ with urea; we model these as linear changes. Since I is not well resolved in urea-induced unfolding transitions (Figure 2), the urea dependence of $[\Theta]_{222}$ of I is omitted.

A three-state equation, based on the model above, (eq 15, Appendix) is fitted to the data using nonlinear least-squares adjustment (Johnson & Frasier, 1985) with a beta version of the program *NonLin for Macintosh* (Brenstein, 1991). Confidence intervals are evaluated at 67% with the NonLin program, using methods described by Johnson (1983).

RESULTS

Urea Dependence of Acid-Induced Unfolding. Figure 1 shows the effect of urea on acid-induced unfolding transitions. The existence of I is apparent at all urea concentrations except 4.5 M, where the protein is fully unfolded (although the amount of intermediate appears to decrease as urea is added; compare the effects of 0 versus 3 M urea on observed $-[\Theta]_{222}$ values in the pH range 4.5–5.5 of Figure 1). The addition of urea to the pH titrations destabilizes N relative to I and U, as evidenced by a progressive shift of the first unfolding transition to higher pHs (from pH 5 at 0 M urea to pH 7.2 at 3 M urea). The addition of urea also destabilizes I relative to U: the second transition shifts from about pH 3.3 at 0 M urea to pH 4.4 at 2 and 3 M urea. This sensitivity to urea allows the unfolding transitions to be studied over a larger range of pH,

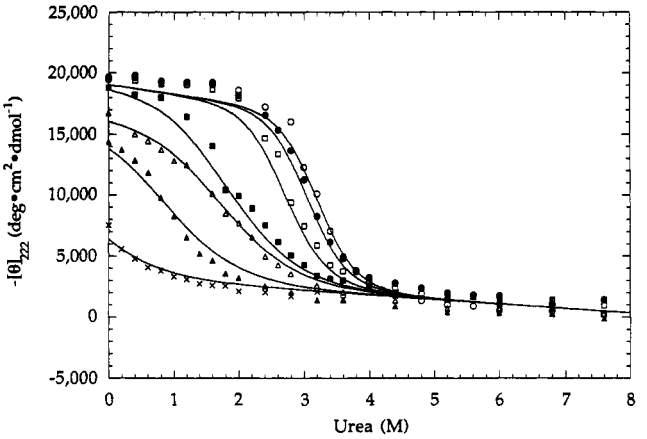

FIGURE 2: Urea-induced unfolding of apoMb at various values of pH at 0 °C and 2 mM sodium citrate. Data are from experiment 1 (see text and Table I). Symbols indicate pH values of approximately 7.8 (○), 7.4 (●), 6.9 (□), 5.9 (■), 5.0 (△), 3.9 (▲), and 2.8 (×). The solid lines are determined using the best-fitting parameters to data from experiment 1. In these urea-induced transitions we observe minor, but in some cases significant, pH differences throughout a given transition. Therefore, rather than evaluating eq 15 at a single averaged pH to generate expected urea transitions, we evaluate eq 15 at combinations of urea and pH corresponding to the actual conditions at each data point. The solid lines are interpolated curves through these expected $[\Theta]_{222}$ values.

making possible the determination of pK_as linked to conformational stability, as well as the extrapolation of different conformational equilibria to common reference conditions.

pH Dependence of Urea-Induced Unfolding. Figure 2 shows the effect of pH on urea-induced unfolding transitions. As the pH is decreased, unfolding transitions shift to lower urea concentrations. The existence of I is less obvious in urea-induced unfolding than in the acid-induced transitions of Figure 1. The urea unfolding transitions in Figure 2 appear to be single, sigmoidal transitions, even at pHs where N and I coexist (pH 5 and 6; see Figure 1). This may reflect different mechanisms in acid- and urea-induced unfolding: acid is a highly specific unfolding agent, which acts at a limited number of high-affinity proton-binding sites, whereas urea is relatively nonspecific and is less likely to destabilize one particular folded conformation selectively.

Fitting the Urea- and Acid-Induced Unfolding Data. We fitted the three-state model described above to our pH- and urea-induced unfolding data. We applied nonlinear least-squares fitting procedures (Johnson & Frasier, 1985; Brenstein, 1991) to fit eq 15 to the data. The fitting procedure was carried out on the data from Figures 1 and 2 simultaneously. This data set contains 299 independent measurements of apoMb at different pH values and urea concentrations and will be referred to as "experiment 1". The parameters resulting from fitting eq 15 to experiment 1 are given in Table I, along with 67% confidence intervals (Johnson, 1983).

The agreement between the experimental data and the model can be seen in Figures 1 and 2. In Figure 1, solid lines show the result of evaluating eq 15 as a function of pH, using best-fitting parameters (Table I, experiment 1) along with the urea concentrations given in the figure caption. The solid lines in Figure 2 are also determined using the best-fitting parameters to data from experiment 1. The model reproduces the observed dependence of $[\Theta]_{222}$ on pH and urea reasonably well. Use of the least-squares minimization procedure on the data in experiment 1 (Figures 1 and 2) results in a residual of 630 deg·cm²·dmol⁻¹, while fitting the experiment 2 data (an independent data set; see below) gives a residual of 1100 deg·cm²·dmol⁻¹.

The 67% confidence intervals for parameters determined by the fitting procedure are sufficiently small (Table I). In most cases, the interval of 67% confidence for each parameter has a width of around 10% of the parameter value. Confidence intervals are larger for the reference equilibrium constants (especially $K_{N\rightarrow I}$); when these are transformed to free energies, however, the confidence intervals again have widths of 10% of best estimates. These confidence intervals report on how well a parameter has been determined from the data used to generate the parameter. In order to determine the reproducibility of these parameters, we made independent measurements on a different protein preparation. This data set (experiment 2) contains 245 separate measurements of $[\Theta]_{222}$ at different pH values and urea concentrations (data not shown). We then repeated the fitting procedure for the data from experiment 2. The best-fitting parameters are given in Table I. In all cases except for $m_{N\rightarrow I}$, the 67% confidence intervals overlap between the two experiments, indicating that these intervals are reasonable indicators of uncertainty. Data from experiments 1 and 2 (544 measurements) were then combined; parameters resulting from fitting to this combined data set (Table I, 1 + 2) are discussed below.

DISCUSSION

Assumptions of the Model. We make three assumptions when fitting our model to the data. (I) There are only three different conformations. Our model treats I as a single species and does not allow for gradual unfolding to U. A test of this assumption can be made by comparing independent probes of folding, such as far-UV CD and fluorescence. Fluorescence changes resulting from acid-induced unfolding of apoMb qualitatively agree with CD changes at 222 nm, but a quantitative analysis of fluorescence as a function of pH and urea is not possible because the pH dependence of the native baseline is complicated [unpublished results; also see Kirby and Steiner (1970) and Irace et al. (1981)]. (II) Each unfolding transition is caused by abnormal pK_as of just one type of group. We are testing this assumption for transition 1 by mutating histidine residues to see if replacement of n_1 histidines eliminates the pH dependence of the N → I reaction. (III) The effects of urea and acid on the conformational equilibria are independent. This assumption requires that pK_as are not influenced by urea and that pH does not change m-values. We are testing the effect of urea on pK_as by titrating with pH short peptides containing histidine in the presence and absence of urea. Studies on RNase A and RNase T1 have shown that m-values may change with pH (Pace et al., 1990). The largest effect was found for RNase A, where m increases by approximately 50% between pH 8 and 3. RNase T1 shows only a 30% increase (Pace et al., 1990). This effect cannot be predicted a priori, nor can it be extracted reliably from our data.

The failure of any of these assumptions would compromise the meaning of the parameters that have been determined. A failure of (II) would give an oversimplistic accounting of the titratable groups linked to folding, but the determined pH dependence would still be useful in extrapolating conformational equilibria. A failure of (III) would still give a reasonable description of solvent effects at each transition, but would introduce error in extrapolation. A failure of (I) could strip thermodynamic meaning from all parameters.

The model has the advantage of being simple and in most respects physically reasonable. Moreover, the model is based on well-understood phenomena, such as acid–base equilibria, and urea-induced conformational transitions. We have used these phenomena to build what we believe is the simplest model,

Table I: Best-Fit Parameters for Apomyoglobin Unfolding

parameter	experiment 1[a] value	experiment 1[a] lower confidence limit[b]	experiment 1[a] upper confidence limit[b]	experiment 2[a] value	experiment 2[a] lower confidence limit[b]	experiment 2[a] upper confidence limit[b]	1 + 2[a] value
$pK_{a1,I}$	7.14	6.96	7.31	7.00	6.81	7.14	7.05
n_{H1}	2.27	2.02	2.52	2.37	2.00	2.73	2.26
$pK_{a2,U}$	4.57	4.42	4.74	4.50	4.29	4.69	4.58
n_{H2}	1.08	0.96	1.19	1.43	1.18	1.70	1.31
$m_{N\rightarrow I}$[c]	0.98	0.87	1.07	1.27	1.12	1.37	1.03
$m_{I\rightarrow U}$[c]	0.94	0.90	0.99	0.86	0.79	0.93	1.01
$K_{N\rightarrow I,reference}$	1.6×10^{-4}	9.0×10^{-5}	2.7×10^{-4}	9.9×10^{-5}	4.4×10^{-5}	1.8×10^{-4}	2.3×10^{-4}
$\Delta G°_{N\rightarrow I,reference}$[d]	4.71	4.35[e]	4.97[e]	4.98	4.53[e]	5.33[e]	4.55
$K_{I\rightarrow U,reference}$	3.9×10^{-2}	3.0×10^{-2}	4.8×10^{-2}	3.5×10^{-2}	2.3×10^{-2}	5.0×10^{-2}	2.2×10^{-2}
$\Delta G°_{I\rightarrow U,reference}$[d]	1.75	1.63[e]	1.87[e]	1.81	1.58[e]	2.00[e]	2.07
$-[\Theta]_{222,N}$[f]	19 000	18 700	19 300	19 900	19 300	20 500	19 260
$-[\Theta]_{222,I}$[f]	16 500	16 000	17 100	17 400	15 900	18 800	16 200
$-[\Theta]_{222U}$[f]	3300	3000	3600	2660	2230	3190	3250

[a] Experiments 1 and 2 were done on different protein preparations. CD measurements for the two experiments were made on different months. Data from the two experiments were fitted separately using nonlinear least-squares adjustments. The data set in experiment 1 contained 299 separate measurements; that in experiment 2, 245. Entries 1 + 2 result from fitting to both data sets simultaneously. [b] Evaluated at the 67% confidence level. [c] In kcal·mol^{-1}·M^{-1} (1 cal = 4.187 J). [d] Calculated using the formula $\Delta G = -RT \ln K$, where R is the gas constant in kcal·mol^{-1}·K^{-1} and T is the absolute temperature. Reference conditions are 0 M urea and pH $\gg pK_{a1,I}$. [e] Uncertainties in free energy are calculated as $-RT[\partial(\ln K)/\partial K] \times \sigma_K$, where σ_K is the confidence limit of K. [f] Circular dichroism, in deg·cm^2·dmol^{-1}, of N, I, and U in the absence of urea. In addition, for N and U a linear dependence of $[\Theta]_{222}$ on urea was fitted for experiment 1, $\partial(-[\Theta]_{222,N})/\partial[\text{urea}] = -750$ deg·cm^2·mol^{-1}·(M urea)$^{-1}$, and $\partial(-[\Theta]_{222,U})/\partial[\text{urea}] = -370$ deg·cm^2·mol^{-1}·(M urea)$^{-1}$; for experiment 2 $\partial(-[\Theta]_{222,N})/\partial[\text{urea}] = -770$ deg·cm^2·mol^{-1}·(M urea)$^{-1}$, and $\partial(-[\Theta]_{222,U})/\partial[\text{urea}] = -170$ deg·cm^2·mol^{-1}·(M urea)$^{-1}$; for the combined fitting 1 + 2 $\partial(-[\Theta]_{222,N})/\partial[\text{urea}] = -220$ deg·cm^2·mol^{-1}·(M urea)$^{-1}$, and $\partial(-[\Theta]_{222,U})/\partial[\text{urea}] = -310$ deg·cm^2·mol^{-1}·(M urea)$^{-1}$.

containing a minimal number of parameters, that can account for the data, namely, two separate acid-induced unfolding reactions, each dependent on urea. The model can be tested in two related ways. First, we ask whether the model adequately reproduces the data. Second, we ask if the fitted parameters are physically reasonable. The model does fit the data reasonably well (see Results), and the parameters determined by fitting are reproducible. There are, however, two regions where systematic errors between the model and the data exist. One region is at low pH and urea concentration (Figure 1): higher CD values are observed than predicted by fitted curves. This probably results from the increase in ionic strength which accompanies pH decrease below pH 2, as described by Goto and Fink (1990). Omission of these outlying points from the data set in experiment 1 did not significantly change the parameters (data not shown). Another region where systematic deviations are seen is at low urea concentration and between pH 3.5 and 5.5. Here transition 1 is predicted to occur at a higher pH than observed. This discrepancy may result from an oversimplification of our model. Specifically, one or more class 1 proton-binding groups may be beginning to bind protons in the native conformation around pH 5, decreasing the sensitivity of transition 1 to pH.

Stability of I. I is closer in Gibbs energy to U than to N. The conformational stability of N, I, and U can be obtained from the reference equilibrium constants $K_{N\rightarrow I}$ and $K_{I\rightarrow U}$. Our best estimates for these values are 2.3×10^{-4} and 2.2×10^{-2} (Table I, experiments 1 + 2), respectively. These give a Gibbs energy difference between N and I of 4.5 kcal·mol^{-1} and between I and U of 2.1 kcal·mol^{-1}. This is consistent with the common observation that equilibrium intermediates are not found: if conditions harsh enough to destabilize N cause even a modest destabilization of I, I will become less stable than U, and only a transition from N to U will be observed.

The primary structural difference between N and I is global loss of tertiary structure (Hughson et al., 1991); I retains substantial helical structure (see Table I: at 0 M urea, $-[\Theta]_{222,N}$ = 19 260, $-[\Theta]_{222,I}$ = 16 200). Thus, the Gibbs energy differences determined here indicate that side-chain packing plays a dominant role in stabilizing the folded protein.

The stability of I can be compared with that of other equilibrium intermediates. Kuwajima and co-workers (Ikeguchi et al., 1986) have measured a difference in Gibbs energy between the native and the molten globule conformations of bovine α-lactalbumin of 2.34 kcal·mol^{-1}; they found a difference in Gibbs energy between the molten globule and U forms of 1.42 kcal·mol^{-1}, at pH 7.0 and 25 °C. As is the case for apoMb, the intermediate conformation of α-lactalbumin shows only modest stability compared with the fully unfolded form. Rather different values have been reported, however, from a recent calorimetric study of bovine α-lactalbumin (Xie et al., 1991): when enthalpy and entropy values at 25 °C and pH 8 are used (Table I, Xie et al., 1991), the Gibbs energy difference between N and the molten globule is reported to be 0.6 kcal·mol^{-1}, and that between the molten globule and U, 1.6 kcal·mol^{-1}.

Our determination of the stability of I is consistent with the protection factors measured by amide hydrogen exchange (Hughson et al., 1990). The stability of I relative to U, evaluated at 0 M urea and pH 4.2, is −1.1 kcal·mol^{-1}. This corresponds to a protection factor of 8, if exchange occurs by complete unfolding and by the EX2 mechanism [see the review by Englander and Kallenbach (1984)]. Protection factors for apoMb I were measured at pH 4.2 and 5 °C by Hughson et al. (1990). The most strongly protected helices in I gave protection factors ranging from 5 to 90, with an average value of 24, corresponding to a Gibbs energy of −1.6 kcal·mol^{-1} in the EX2 limit. [To determine the average, we have divided protection factors for the two Val and three Ile residues in helicies A, G, and H, Table 1 of Hughson et al. (1990), by 4, because the intrinsic exchange rates used to calculate these protection factors (Molday et al., 1972) are too high by a factor of 4 (Robertson & Baldwin, 1991)]. Exchange may occur directly from I without complete unfolding to U. If this happens, exchange will be faster and the measured protection factors will be smaller than predicted for complete unfolding to U. Our data are consistent with exchange occurring by unfolding to U.

A similar comparison can be made from exchange measurements on N at pH 6 (Hughson et al., 1990): the free energy calculated from protection factors is 5.6 kcal·mol^{-1}; that from the data presented here is calculated to be 3.4 kcal·mol^{-1}. This discrepancy may be caused in part by the different temperatures of the CD and hydrogen exchange

measurements (0 and 5 °C, respectively). ApoMb at low temperature is in the cold-denaturation regime (Griko et al., 1988) and shows increased stability with temperature increase.

Urea Dependence of the Three-State Equilibrium. The urea dependences (m-values) of the N → I and I → U equilibria are 1.03 and 1.01 kcal·mol^{-1}·M^{-1}. The similarity of values for $m_{N\rightarrow I}$ and $m_{I\rightarrow U}$ shows that the increase in urea binding sites resulting from full unfolding of I is roughly the same as the increase upon formation of I from N. If the urea binding sites are equally distributed on the surface of the unfolded protein, then the m-value is proportional to the increase in solvent-accessible surface area on unfolding. This interpretation indicates that only half of the buried surface area in N becomes exposed to solvent in the N to I transition; the other half remains sequestered from solvent in I. This view is supported by fluorescence studies in which the emission maximum for tyrosine and tryptophan fluorescence is found midway between those of N and U (Kirby & Steiner 1970) and is consistent with stabilization of I by interactions among hydrophobic groups (Hughson et al., 1991).

The results obtained here for apoMb can be compared with those for other proteins. The m-values for the two unfolding transitions of α-lactalbumin found by Ikeguchi et al. (1986) are not the same: 0.7 kcal·mol^{-1}·(M guanidine hydrochloride)$^{-1}$ for the transition from molten globule to U, compared with 1.6 kcal·mol^{-1} (M guanidine hydrochloride)$^{-1}$ for the transition from N to molten globule. Kuwajima and co-workers (1977) found, however, a substantial positive heat capacity increment on unfolding of the molten globule of α-lactalbumin, indicating that the molten globule has substantial nonpolar surface area that is shielded from solvent. Similar findings have recently been reported by Xie et al. (1991). A similar conclusion has been reached about a molten globule intermediate of horse cytochrome c (referred to as IIc; Kuroda et al., 1992).

pH Dependence of the Three-State Equilibrium. The pH dependence of transition 1 is best described by 2.26 proton-binding sites with pK_as of 7.05 in I. This pK_a is close to that of the imidazole side chain of histidine, which was measured in small peptides to have a value of 6.8–6.9 at 25 °C (Takahashi et al., 1992) or 7.2–7.3 at 0 °C using the van't Hoff relation and an ionization enthalpy of 6.9 kcal·mol^{-1} (Edsall & Wyman, 1958). This interpretation of the data suggests that there are approximately two histidines linked to transition 1 and that in I (and U) they have pK_as around 7.0. The assumption we make in our model that these groups are not protonated in N at pH values at or above transition 1 seems in good agreement with the data (Figures 1 and 2). Abnormally low pK_as in N could result either from burial or from the presence of nearby positive charges. Proton-binding curves and histidine studies suggest that one or more histidines in apomyoglobin remain buried and deprotonated at low pH (Breslow, 1964). Recently, Cocco et al. (1992) have reported several low pK_as for histidine residues in apoMb. We are currently testing the involvement of histidine residues in transition 1 by mutating several histidines in sperm whale myoglobin, both individually and in combination. The pH dependence of transition 2 is described here by 1.31 binding sites with a pK_a of 4.58. Although this pK_a is near that of glutamic acid (4.33 at 5 °C; Izatt & Christensen, 1968), it is unlikely that this unfolding transition results from a single nonprotonatable group. Rather, the observed pH dependence may result from several groups with small, conformation-linked pK_a shifts.

Phase Diagram of apoMb. Relative populations of N, I, and U are calculated from eq 16 as a function of pH and urea (Figure 3). As pH is decreased, the population shifts from N to I; I shows its maximal stability at pH 4.6, where it is

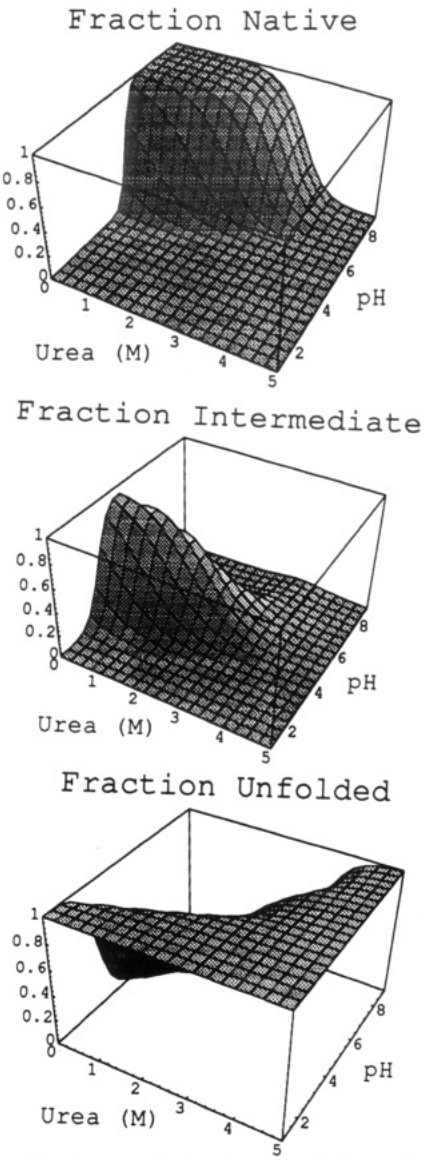

FIGURE 3: Population of the N, I, and U conformations of apomyoglobin as a function of urea and pH. Best-fitting parameters (experiments 1 + 2, Table I) are used to evaluate the fraction of N, I, and U over urea- and pH-space, according to eq 16.

populated to 94%. Figure 3 also indicates that at some pHs urea should increase the population of I. Although urea destabilizes I relative to U, it stabilizes I relative to N: at pHs where N (but not I) is destabilized (from pH 5 to 7), low urea concentrations will convert N to I.

The results here allow the stabilities of N, I, and U to be compared under different solution conditions. This is useful for comparison with other work which suggests that I has some characteristics of a molten globule (Griko et al., 1988; Goto et al., 1990; Hughson et al., 1991). Our data indicate that under physiological conditions (neutral pH, in the absence of urea) I is the predominant non-native conformation. Thus, if I is formed rapidly from U, I is likely to be an early intermediate on the kinetic pathway of folding. This suggests that I may be important for the process of folding in the cell. A major motivation in these studies is to apply the fitting procedures to determine the energetic consequence of mutation on N, I, and U. Considerable data for the pH- and urea-induced transitions of mutant proteins are already available (Hughson et al., 1991), and more data of this kind are now being acquired (D. Barrick, unpublished).

ACKNOWLEDGMENT

We are very grateful to Steve Sligar and Barry Springer for the gift of pMb413a and to Fred Hughson for developing much of the apomyoglobin system. We thank Melanie Cocco and Juliette Lecomte for sharing results prior to publication and also for useful discussion. We also thank the Stanford Department of Statistics and especially Josee Dupuis for useful discussions and advice regarding data analysis. Fred Hughson and Marty Scholtz are thanked for critical reading of this manuscript.

APPENDIX: A THREE-STATE STABILITY EQUATION

Here equations are derived which describe the relative stabilities of the N, I, and U forms of apoMb to pH and to urea concentration. From these stability equations, an expression can be written which relates the CD signal to solution conditions. It is this equation which we fit to our data. The model we adopt to determine this equation, as well as the underlying rationale, is described in the text.

We begin by considering the equilibrium between two of the three conformations, for instance, N and I, as a function of urea and pH. We write the equilibrium constant between the two forms as

$$K_{N\rightarrow I} = K_{N\rightarrow I,\text{reference}} f_{N\rightarrow I}(\text{urea}) g_{N\rightarrow I}(\text{pH}) \quad (1)$$

The same equation can be written for the I to U equilibrium. In (1) it is assumed that the effects of urea and of pH ($f_{N\rightarrow I}$(urea) and $g_{N\rightarrow I}$(pH), respectively) on conformational stability are independent of each other. This formalism has been used by Aune and Tanford (1969a,b) in their two-state analysis of hen lysozyme unfolding. $K_{N\rightarrow I,\text{reference}}$ is a reference equilibrium constant. We have chosen high pH and 0 M urea as our reference point as described in the Results section. Below we determine the explicit forms for the urea and the pH dependence and substitute them into (1).

We obtain the urea dependence of $K_{N\rightarrow I}$ by adopting the linear extrapolation model of Pace (Greene & Pace 1974; Pace, 1986),

$$\Delta G^{\circ}_{N\rightarrow I}(\text{urea}) = -RT \ln K_{N\rightarrow I}(\text{urea}) = \Delta G^{\circ}_{N\rightarrow I}(H_2O) - m_{N\rightarrow I}[\text{urea}] \quad (2)$$

where m is assumed to be independent of urea and pH. The same model is used to describe the I to U reaction.

The function $g_{N\rightarrow I}$(pH) is derived by assuming identical and independent binding sites (Cantor & Schimmel, 1980). For a conformation with n proton-binding sites (N, for example) the dissociation of a single proton from N bound with i protons (NH_i) is described by a macroscopic equilibrium constant,

$$K_{i,N} = \frac{[NH_{i-1}][H]}{[NH_i]} \quad (3)$$

Similar expressions can be written for proton dissociation from $[NH_{i-j}]$ ($j < i$); substitution of these relations into (3) followed by rearrangement gives

$$[NH_i] = [N]\prod_{j=1}^{i}\left(\frac{[H^+]}{K_{j,N}}\right) \quad (4)$$

The total concentration of native protein, regardless of the degree of protonation, is given by

$$[N_{\text{total}}] = [N]\left(1 + \sum_{i=1}^{n}\left\{\prod_{j=1}^{i}\left(\frac{[H^+]}{K_{j,N}}\right)\right\}\right) \quad (5)$$

In the model we develop here, all groups of a particular class have identical and independent pK_as for a given conformation and, therefore, identical microscopic acid dissociation constants. Considering a single class of proton-binding groups, for instance, those in class 1 (see text), the macroscopic constant K_i can be related to the microscopic acid dissociation constant through the relation

$$K_{i_1,N} = \frac{\Omega_{n_1,i-1}}{\Omega_{n_1,i}} \times K_{a_1,N} \quad (6)$$

where the statistical factor $\Omega_{n_1,i}$ is the number of distinct ways to arrange the i protons on n_1 sites of the species NH_i and is given by

$$\Omega_{n_1,i} = \frac{n_1!}{(n_1 - i)!i!} \quad (7)$$

In this discussion, we neglect other sites not linked to transition 1, as they will be identical in N and I and will not affect the equilibrium. Writing the statistical factors in (6) as prescribed by (7) and substituting into (5) yields

$$[N_{\text{total}}] = [N]\left(1 + \sum_{i=1}^{n_1}\left(\frac{[H^+]}{K_{a_1,N}}\right)^i\left(\frac{n_1!}{(n_1 - 1)!i!}\right)\right) \quad (8)$$

Since the term in parentheses on the right-hand side of (8) is the binomial expansion of $(1 + [H^+]/K_{a_1,N})^{n_1}$, eq 13 can be written as

$$[N_{\text{total}}] = [N]\left(1 + \frac{[H^+]}{K_{a_1,N}}\right)^{n_1} \quad (9)$$

The same relations can be written for I and U, where the acid dissociation constants change as prescribed by the model. We can then use eq 9 and analogous ones to obtain conformational equilibrium constants as a function of pH, as, for example,

$$K_{N\rightarrow I}(\text{pH}) = \frac{[I_{\text{total}}]}{[N_{\text{total}}]} = \frac{[I]}{[N]}\frac{\left(1 + \frac{[H^+]}{K_{a_1,I}}\right)^{n_1}}{\left(1 + \frac{[H^+]}{K_{a_1,N}}\right)^{n_1}} = K_{\text{pH}\gg\text{p}K_{aN,I}}\frac{\left(1 + \frac{[H^+]}{K_{a_1,I}}\right)^{n_1}}{\left(1 + \frac{[H^+]}{K_{a_1,N}}\right)^{n_1}} \quad (10)$$

A similar relation describes $K_{I\rightarrow U}$(pH). A further simplification comes from the condition we impose in our model that in the native form, class 1 proton-binding sites (those linked to transition 1) do not protonate. This means that for class 1 proton-binding sites $K_{a_1,N} \gg [H^+]$, which simplifies (10) to

$$K_{N\rightarrow I}(\text{pH}) = K_{\text{pH}\gg\text{p}K_{a1}}\left(1 + \frac{[H^+]}{K_{a_1,I}}\right)^{n_1} \quad (11)$$

Thus $g_{N\rightarrow I}$(pH) can be taken as the term in parentheses on the right-hand side of (11). Again, a similar relation can be written for transition 2, where for class 2 proton-binding sites $K_{a_2,I} \gg [H^+]$.

The pH and urea dependences of the conformational equilibria between N and I and between I and U can then be written as

$$K_{N\rightarrow I} = K_{\text{reference},N\rightarrow I}10^{(m_{N\rightarrow I}[\text{urea}]/2.303RT)}\left(1+\frac{[H^+]}{K_{a_1,I}}\right)^{n_1} \quad (12)$$

$$K_{I\rightarrow U} = K_{\text{reference},I\rightarrow U}10^{(m_{I\rightarrow U}[\text{urea}]/2.303RT)}\left(1+\frac{[H^+]}{K_{a_2,U}}\right)^{n_2} \quad (13)$$

From these equilibrium constants an expression can be constructed which describes the observed $[\Theta]_{222}$ as a function of both urea and pH. Assuming that only three conformations (N, I, and U) are significantly populated,

$$[\Theta]_{222,obs} = [\Theta]_{222,N}f_N + [\Theta]_{222,I}f_I + [\Theta]_{222,U}f_U \quad (14)$$

where $[\Theta]_{222,N}$, $[\Theta]_{222,I}$, and $[\Theta]_{222,U}$ are the CD values at 222 nm for N, I, and U, respectively. Our data show that, for N and U, $[\Theta]_{222,N}$ depends approximately linearly on urea concentration; therefore, we include terms linear in urea for $[\Theta]_{222,N}$ and $[\Theta]_{222,U}$. The terms f_N, f_I, and f_U represent the fraction of [N], [I], and [U] and can be rewritten in (14) to give

$$[\Theta]_{222,obs} = \frac{[\Theta]_{222,N} + [\Theta]_{222,I}K_{N\rightarrow I} + [\Theta]_{222,U}K_{N\rightarrow I}K_{I\rightarrow U}}{1 + K_{N\rightarrow I} + K_{N\rightarrow I}K_{I\rightarrow U}} \quad (15)$$

The equilibrium constants in (15) are represented as in (12) and (13) and give both urea and pH dependence to $[\Theta]_{222,obs}$. Additional effects of urea result from the baseline dependencies on $[\Theta]_{222,N}$ and $[\Theta]_{222,U}$ as described above. It is eq 15 which we fit to the data in experiments 1 and 2, described in the text.

The fraction of N, I, and U can be expressed in terms of urea and pH as

$$f_N = \frac{1}{1 + K_{N\rightarrow I} + K_{N\rightarrow I}K_{I\rightarrow U}} \quad (16)$$

$$f_I = \frac{K_{N\rightarrow I}}{1 + K_{N\rightarrow I} + K_{N\rightarrow I}K_{I\rightarrow U}}$$

$$f_U = 1 - f_N - f_I$$

REFERENCES

Aune, K. C., & Tanford, C. (1969a) *Biochemistry 8*, 4579–4585.

Aune, K. C., & Tanford, C. (1969b) *Biochemistry 8*, 4586–4590.

Brenstein, R. J. (1991) *NonLin for Macintosh*, Robelko Software, Carbondale, IL.

Breslow, E. (1964) *J. Biol. Chem. 239*, 486–496.

Cantor, C. R., & Schimmel, P. R. (1980) Ligand Interactions at Equilibrium, in *Biophysical Chemistry Part III: The behavior of biological macromolecules*, pp 849–886, W. H. Freeman and Company, New York.

Cocco, M. J., Kao, Y.-H., Phillips, A. T., & Lecomte, J. T. J. (1992) *Biochemistry 31*, 6481–6491.

Edelhoch, H. (1967) *Biochemistry 6*, 1948–1954.

Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*: 521–665.

Edsall, J. T., & Wyman, J. (1958) *Biophysical Chemistry*, Vol. 1, Academic Press, New York.

Goto, Y., & Fink, A. L. (1990) *J. Mol. Biol. 214*, 803–805.

Goto, Y., & Nishikiori, S. (1991) *J. Mol. Biol. 222*, 679–686.

Goto, Y., Calciano, L. J., & Fink, A. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 573–533.

Greene, R. F. J., & Pace, C. N. (1974) *J. Biol. Chem. 249*, 5388–5393.

Griko, Y. V., Privalov, P. L., Venyaminov, S. Y., & Kutyshenko, V. P. (1988) *J. Mol. Biol. 202*, 127–138.

Hansen, J. C., & Schoenborn, B. P. (1981) *J. Mol. Biol. 153*, 117–146.

Hughson, F. M., & Baldwin, R. L. (1989) *Biochemistry 28*, 4115–4222.

Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544–1548.

Hughson, F. M., Barrick, D., & Baldwin, R. L. (1991) *Biochemistry 30*, 4113–4118.

Ikeguchi, M., Kuwajima, K., & Sugai, S. (1986) *J. Biochem. (Tokyo) 99*, 1191–1201.

Irace, G., Balestrieri, C., Parlato, G., Servillo, L., & Colonna, G. (1981) *Biochemistry 20*, 792–799.

Izatt, R. M., & Christensen, J. J. (1968) *CRC Handbook of Biochemistry: Selected Data for Molecular Biology*, pp J49-J139, CRC Press, Inc., Boca Raton, FL.

Johnson, M. L. (1983) *Biophys. J. 44*, 101–106.

Johnson, M. L., & Frasier, S. G. (1985) *Methods Enzymol. 117*, 301–342.

Kirby, E. P., & Steiner, R. F. (1970) *J. Biol. Chem. 245*, 6300–6306.

Kuroda, Y., Kidokoro, S.-I., & Wada, A. (1992) *J. Mol. Biol. 223*, 1139–1153.

Kuwajima, K. (1977) *J. Mol. Biol. 114*, 241–258.

Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87–103.

Kuwajuma, K., Katsutoshi, N., Yoneyama, M., & Sugai, S. (1976) *J. Mol. Biol. 106*, 359–373.

Molday, R. S., Englander, S. W., & Kallen, R. G. (1972) *Biochemistry 11*, 150–158.

Pace, C. N. (1986) *Methods Enzymol. 131*, 266–280.

Pace, C. N., Laurents, D. V., & Thomson, J. A. (1990) *Biochemistry 29*, 2564–2572.

Robertson, A. D., & Baldwin, R. L. (1991) *Biochemistry 30*, 9907–9914.

Santoro, M. M., & Bolen, D. W. (1992) *Biochemistry 31*, 4901–4907.

Schellman, J. A. (1978) *Biopolymers 17*, 1305–1322.

Springer, B. A., & Sligar, S. G. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8961–8965.

Takahashi, T., Nakamura, H., & Wada, A. (1992) *Biopolymers 32*, 897–909.

Tanford, C. (1970) *Adv. Protein Chem. 124*, 1–95.

Teale, F. W. J. (1959) *Biochim. Biophys. Acta 35*, 543.

Xie, D., Bhakuni, V., & Freire, E. (1991), *Biochemistry 30*, 10673–10678.

Yutani, K., Ogasahara, K., & Kuwajima, K. (1992) *J. Mol. Biol. 228*, 347–350.