

# Structural Details, Pathways, and Energetics of Unfolding Apomyoglobin

Alexey Onufriev, David A. Case and Donald Bashford\*

Department of Molecular  
Biology, The Scripps Research  
Institute, TPC15, 10550 N  
Torrey Pines Road, La Jolla, CA  
93027, USA

Protein folding is often difficult to characterize experimentally because of the transience of intermediate states, and the complexity of the protein–solvent system. Atomistic simulations, which could provide more detailed information, have had to employ highly simplified models or high temperatures, to cope with the long time scales of unfolding; direct simulation of folding is even more problematic. We report a fully atomistic simulation of the acid-induced unfolding of apomyoglobin in which the protonation of acidic side-chains to simulate low pH is sufficient to induce unfolding at room temperature with no added biasing forces or other unusual conditions; and the trajectory is validated by comparison to experimental characterization of intermediate states. Novel insights provided by their analysis include: characterization of a dry swollen globule state forming a barrier to initial unfolding or final folding; observation of cooperativity in secondary and tertiary structure formation and its explanation in terms of dielectric environments; and structural details of the intermediate and the completely unfolded states. These insights involve time scales and levels of structural detail that are presently beyond the range of experiment, but come within reach through the simulation methods described here. An implicit solvation model is used to analyze the energetics of protein folding at various pH and ionic strength values, and a reasonable estimate of folding free energy is obtained. Electrostatic interactions are found to disfavor folding.

© 2003 Elsevier Science Ltd. All rights reserved

*Keywords:* protein folding; apomyoglobin; energetics; molecular dynamics; continuum solvent

\*Corresponding author

## Introduction

Protein folding generally occurs on time scales well beyond the current range of conventional, fully atomistic, explicit-solvent molecular dynamics simulations. Some of the smallest proteins are thought to fold in a few microseconds,<sup>1</sup> while millisecond or longer time scales are more typical. Therefore, efforts to simulate folding have often employed simplified physical models, such as lattice models of the peptide chain<sup>2–5</sup> or continuum solvent, and in many cases remain confined to unusually small systems. As an alternative, one can generate non-native states of a protein by simulating its unfolding from the native state.<sup>6–8</sup> In general, one would expect that the milder the simulated unfolding

conditions, the better the resulting states would approximate those occurring on the natural folding or unfolding pathways. However, to reach substantially unfolded states within feasible computation times, vacuum simulations, high temperatures, or biasing potentials are often used.<sup>6,9–11</sup> There is some controversy over the applicability of high-temperature simulations to lower-temperature unfolding, with some test simulations, suggesting that the processes pass through similar states at both higher and lower temperature,<sup>6,12</sup> while other such tests show pathway differences.<sup>13</sup>

Apomyoglobin (myoglobin without the heme group) is well suited for both theoretical and experimental studies of folding, as it folds through a set of well-defined intermediate states.<sup>14</sup> Experimental studies reveal<sup>15–17</sup> that the protein is structurally very similar to the native (holo) myoglobin, retaining most of its secondary, and most likely, tertiary structure. With the addition of acid, apomyoglobin undergoes a multi-phase unfolding, first to a molten globule intermediate (I-state) at

Abbreviations used: PB, Poisson–Boltzmann; GB, generalized Born.

E-mail address of the corresponding author:  
bashford@scripps.edu

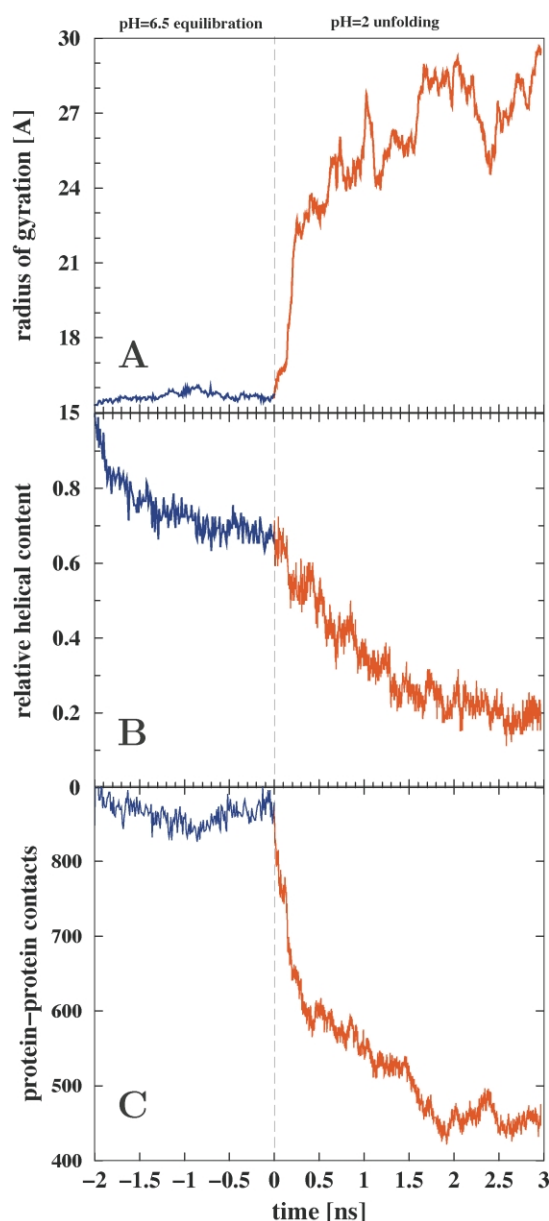
about pH 4, then to a mostly unfolded state (but with measurable residual structure) at pH 2, and finally to a more nearly completely unfolded state at pH 2 in the presence of urea. In kinetic unfolding experiments at pH 2.7, the native state unfolds through the I-state,<sup>18</sup> and in re-folding experiments at neutral pH, the I-state is shown to be an obligatory folding intermediate,<sup>18–20</sup> suggesting strong similarity between the acid-unfolding and re-folding pathways of apomyoglobin.<sup>14</sup>

To generate an unfolding trajectory of apomyoglobin we have carried out a molecular dynamics simulation, including all protein atoms and ~13,000 water molecules, starting from an initial state closely approximating the known features of apomyoglobin at pH 6.5. This initial state has been prepared from an experimental structure of holo-myoglobin using MD simulation at neutral pH. To model the conditions of acid shock (pH 2) we protonate all titratable groups in the protein, and continue the simulation for 3 ns. No other conditions to enforce unfolding, such as a biasing potential or high temperature, are employed. Furthermore, simulation under identical conditions, but without the above protonation produces a stable trajectory that remains in the native (apo) state, indicating that the simulation protocol does not produce an undue bias toward unfolding. The behavior of the partially unfolded states in the simulation matches, in many respects, features seen experimentally, including the radius of gyration, the average helical content for individual helices, and the presence of residual native contacts. An analysis of the ensemble of structures seen in the simulation provides new insights into the structures and energetics that probably characterize the acid-induced folding/unfolding transition.

## Results and Discussion

### Modeling of apomyoglobin at neutral pH

The apomyoglobin model is generated by removing the heme group from the native structure of the holo-protein and subsequent equilibration using molecular dynamics at neutral pH. The simulated structural changes are summarized in Figure 1 (blue trace, from  $t = -2$  ns to  $t = 0$  ns). After 2 ns, the myoglobin native fold is only slightly distorted, (Figure 3, 0.0 ns, lower panel), mainly due to collapse of helix F into the space previously occupied by the heme group. This results in appearance of a few new contacts between protein residues (red dots, 0.0 ns upper panel), which are not present in holo-myoglobin (green squares). However, as is evident from the corresponding contact map, at this stage the simulated protein is structurally very similar to holo-myoglobin, in agreement with general experimental observations.<sup>15–17</sup> Its overall characteristics, given in Table 1 as apo calc, such as radius of gyration,



**Figure 1.** Structural changes in apomyoglobin during equilibration at pH 6.5 (blue trace) and simulated unfolding at pH 2 (red trace). The zero time point is defined as the beginning of the acid-unfolding trajectory. (A) Radius of gyration. Calculated using all atoms. ( $R_g$ ). (B) Calculated  $\alpha$ -helical content relative to that of holo-myoglobin crystal structure. (C) Number of protein-protein (residue-residue) contacts in the protein. Detailed definitions of the computed quantities are given in the Methods section.

helical content and the average separation between helices A and G are similar to those observed experimentally<sup>16,21</sup> (apo exp) in apomyoglobin. The calculated distribution of helicity among major helices, A: 0.69, B: 0.97, E: 0.81, G: 0.75, and H: 0.70, is in reasonable agreement with experiment.<sup>16</sup> We find the F helix considerably unstructured, with only 56% of the native helicity left. Although no experimental number exists for the helical population of F helix in apomyoglobin,

**Table 1.** Calculated *versus* experimental structural signatures

Signature	Holo		Apo		I		U	
	Exp	Calc	Exp	Calc	Exp	Calc	Exp	Calc
$R_g$	$18 \pm 1$	15.3	$19 \pm 1$	15.8	$23 \pm 2$	21.6	$30 \pm 2$	29.1
$ \Delta G $ (Å)	$24 \pm 2$	22.7	$24 \pm 2$	24.9	$27 \pm 3$	24.3	>50	50.4
Rel. helicity	1	1	0.77	0.67	0.54	0.53	0.14-0.21	0.20

pH 6.5 holomyoglobin (holo), apomyoglobin (apo), acid unfolding intermediate state (I), and completely unfolded state at pH 2 (U).  $|\Delta G|$  is the separation between helices A and G, calculated as the arithmetic mean of distances between  $C_\alpha$  atoms of residues 7 and 103, and 14 and 103. Calculated values are averages over the following: for apo—last 0.5 ns of pH 6.5 MD simulation; for I—from  $t = 0.2$  to 0.25 ns of pH 2 acid-unfolding simulation; and for U—from  $t = 2.866$  to 2.964 ns of the same simulation. Structural signatures of holo-myoglobin are computed from its crystallographic coordinates. Experimental results are taken from Refs. 16,21,54. The experimental radius of gyration, obtained from SAXS experiments, is expected to be about 2 Å larger than that calculated directly from crystallographic coordinates.<sup>55</sup>

this region was suggested<sup>16</sup> to participate in slow exchange between unfolded and partially folded structures. Enhanced structural fluctuations in this region were also observed in earlier models<sup>7,22</sup> of apomyoglobin on the basis of molecular dynamics simulations. The two small helices, C and D, are predicted by our calculations to lose almost all of their holo-protein helical population in the apomyoglobin, at least according to the criteria we have used (on the basis of 1–4 hydrogen bonds). Eliezer *et al.* observed considerable  $\alpha$ -helical population in this region based on the secondary chemical shift data. We attribute the discrepancy to uncertainties in the definition of helicity for such small helices (helix C, for example, is only six residues long), especially in the presence of significant structural fluctuations. On the other hand, our observation that the C–D region loses most of its native structure in apomyoglobin is consistent with the data of Lecomte *et al.*, who observed considerable structural fluctuations in the region, and is in agreement with earlier molecular dynamics simulations of Tirado-Rives & Jorgensen.<sup>7</sup> The above observations allow us to use the simulated structure as a reasonable model for apomyoglobin. Furthermore, the stability of the structural features of this 2 ns neutral-pH, simulation, and its agreement with experiment suggests that our droplet-based simulation protocol (see Methods) does not introduce undue bias away from the folded state.

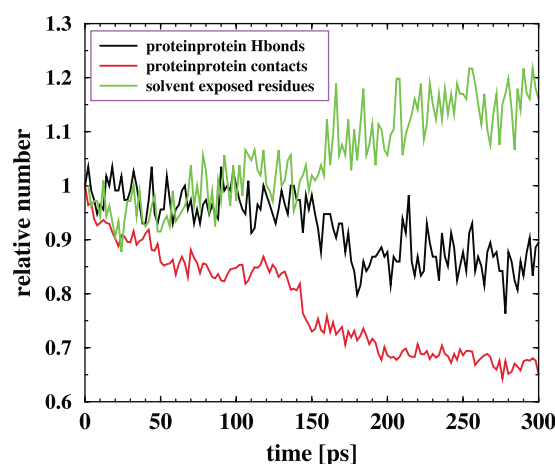
### Simulated acid-unfolding

The simulated structural changes in the protein as a function of time after the acid-shock at  $t = 0$  are presented in Figures 1 (red trace), 2, and 3. Compared to experimental acid-induced unfolding which occurs on sub-millisecond time-scales,<sup>18</sup> the simulated unfolding is complete in  $\sim 3$  ns, reflecting the fact that in the simulation all titratable groups are instantly protonated at  $t = 0$ , while in a pH-jump experiment this process would occur on a finite time-scale through diffusion, and in the case of buried groups, structural fluctuations. The methodology is initially validated by agreement between the computed average characteristics of the starting, intermediate, and completely

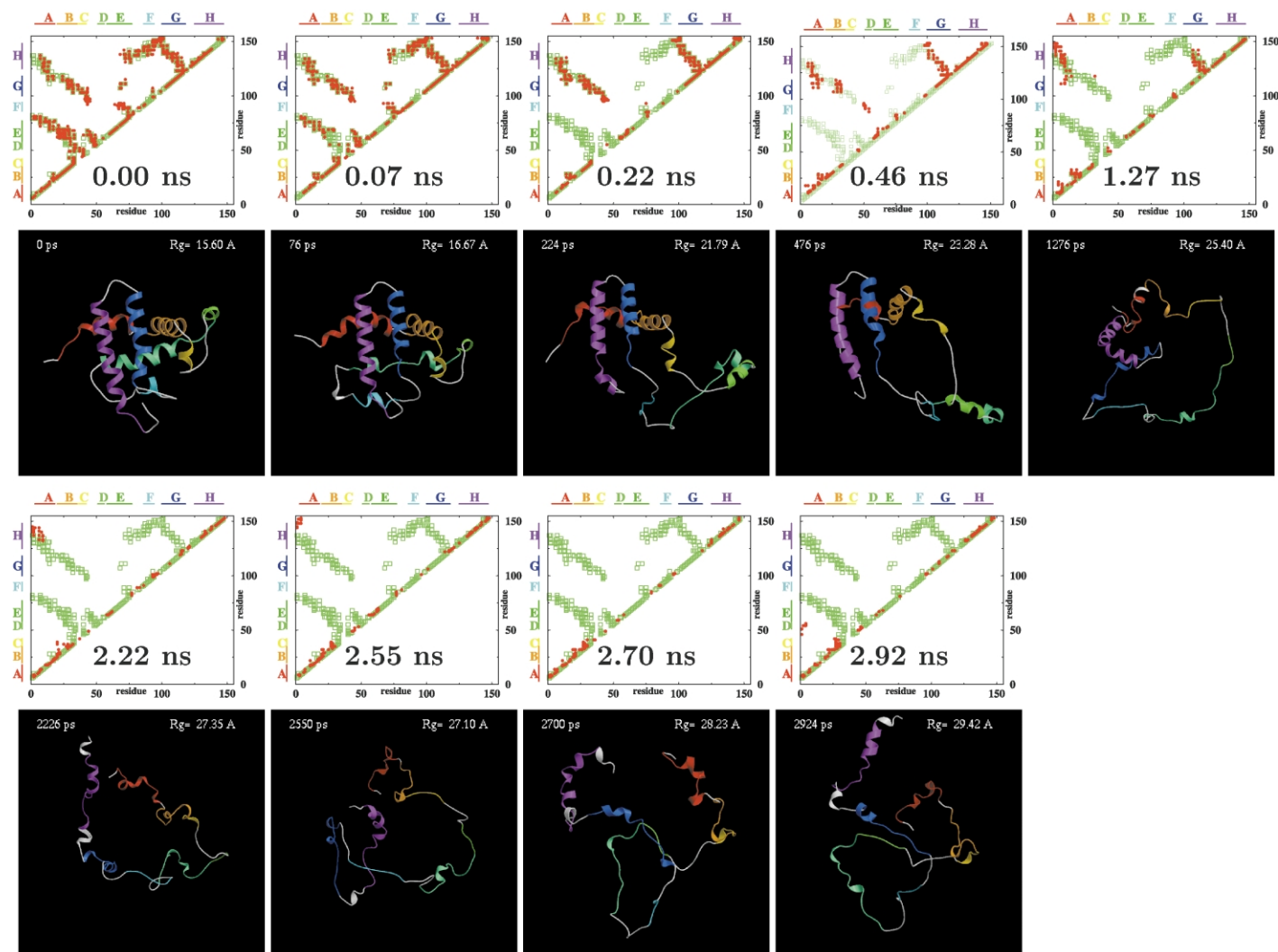
unfolded states and available experimental data; these include protein radius ( $R_g$ ), helix formation, and tertiary interactions in the folded, acid-unfolded and the molten globule intermediate states (Table 1). More detailed analysis of selected stages of the simulation, and where appropriate, comparison with experimental characterizations of the intermediate and acid-unfolded states are presented below.

### Initial stages of unfolding

During the first 0.15 ns of the simulated acid-unfolding trajectory the protein stays in near-native conformations, the globule being only slightly swollen ( $R_g \sim 16.5$  Å, Figure 1). A number of native contacts are disrupted at this stage (Figure 2), but the native fold is practically



**Figure 2.** Initial stages of unfolding. During the first 0.15 ns of the simulated acid-unfolding trajectory a number of native contacts are disrupted, but most hydrogen bonds remain intact (on average), and water does not enter the protein interior, since at this stage there is no increase in the number of solvent exposed residues. A decrease in the number of H-bonds at  $t > 0.15$  ns indicates the disruption of secondary structure, while the corresponding increase in the number of solvent exposed residues can be attributed to water entering the interior of the globule.



**Figure 3.** Predicted non-native states of apomyoglobin during time course of simulated acid-unfolding at pH 2. Colors in the ribbon diagram correspond to consecutive helical regions in the native holo-myoglobin: red (A), orange (B), yellow (C), light green (D), navy (E), cyan (F), blue (G), violet (H). Regions corresponding to native coil are white. The same color scheme is used to highlight helical regions along the axis of contact maps. Contact map (red circles) is computed using a set of 25 consecutive snapshots representing a 0.05 ns interval centered on the time point indicated on each panel (for  $t = 0$  ns the snapshots are taken from the end of pH 6.5 MD simulation). Residues  $i$  and  $j$  are represented by a red dot on the  $(ij)$  plane if they form a contact (see Methods) with probability higher than 0.5 in the given set of structures. Due to obvious symmetry, only the half-plane  $i > j$  is shown; short-range contacts  $|i - j| \leq 4$  are excluded. For comparison, the contact map of holomyoglobin computed from its crystallographic coordinates is superposed as open green squares. Thus, a red dot in a green square is a native contact, an empty green square is a lost native contact, and a red dot alone is a non-native contact.



**Table 2.** Fluctuations of  $\chi_2$  in key residues of native apomyoglobin and its molten globule state

State	Trp7	I15	Phe123	Met131
Apomyoglobin pH 6.5	13.4 ± 1.5	9.0 ± 2.7	13.8 ± 2.7	12.2 ± 2.5
Molten globule pH 2	26.8	10.9	12.4	60

$\chi_2$  torsion angle fluctuations ( $\langle\chi_2^2\rangle - \langle\chi_2\rangle^2$ )<sup>1/2</sup> (deg.), in pH 6.5 apomyoglobin are computed over each of the four 0.5 ns consecutive intervals making up the 2 ns MD trajectory which has been used to produce the apomyoglobin model. The values for the apomyoglobin at neutral pH reported above correspond to the last 0.5 ns interval, and the standard deviation between fluctuations computed in each of the 4 intervals is shown. The corresponding fluctuations in the pH 2 molten globule state are computed over 0.05–0.55 ns interval (which includes both the dry swollen globule and the I-state) of the 3 ns acid-unfolding trajectory. For Phe123 and Leu115, fluctuations in both the native apo and the molten globule states are within the error margin from each other, while for Trp7 and Met131 the fluctuations are substantially larger in the molten globule state. A relatively large value of the fluctuation for Met131 indicates a rotamer transition.

unchanged (Figure 3, 0.07 ns). Most hydrogen bonds remain intact (on average), and the number of residues exposed to water does not significantly increase, indicating that water does not enter the protein interior. Thus, the disruption of energetically favorable native contacts from 0 ns to ~0.15 ns is not compensated by an increase in favorable protein–water contacts. These observations over the ~0.05–0.15 ns interval are consistent with the idea of a “dry swollen globule” as a barrier to initial unfolding and final stages of re-folding. This state was originally postulated theoretically,<sup>23</sup> and later observed in experimental studies of unfolding of ribonuclease A,<sup>24</sup> DHFR,<sup>25</sup> and recently of a small all  $\beta$ -protein *tendamistat*.<sup>26</sup> In the latter study, direct volume measurements have suggested that exclusion of water precedes the formation of the native state during re-folding, implying that the dry swollen globule is the final transition state in folding.

Experimental characterization of such a barrier state is very difficult because of its extremely low population, but the presence of globally intact hydrogen bonds alongside locally disrupted packing interactions in the early unfolding of apomyoglobin<sup>27</sup> and DHFR<sup>28</sup> has been observed, and is consistent with the present simulation results, as well as with earlier MD simulations on BPTI<sup>29</sup> that used thermal unfolding. In their studies of partially unfolded conformations of barnase by molecular dynamics at various temperatures, Caflisch & Karplus<sup>13</sup> observed a delay in the entrance of water into the protein interior upon unfolding at 360 K and low pH, but no such delay at 600 K. In light of these results, our simulations at 300 K are of special interest.

### Intermediate stages

A rapid transition to more open conformations ( $R_g \sim 23 \text{ \AA}$ ), occurs at  $t \approx 0.15 \text{ ns}$  (Figure 1). The protein is now partially unfolded, with large parts of the A, G, H, and B helices forming a compact core, while the rest of the protein is considerably unfolded (Figure 3, 0.22 ns). Compared to the preceding dry swollen globule stage, a number of residues have become exposed to water (see Methods, Calculation of structural signatures), namely: A19, A22, I28, L29, T39, F49, A53, L61,

G65, V68, A74, I75, L76, A84, L89, A90, A94, I99, I101, S108, V114, A134. These residues lie on the interface of E and F helices and EF loop with helices B, C, G and H, which opens upon “swinging” of the E–F subdomain (Figure 3, transition from 0.07 ns to 0.22 ns). The key residues L115 and A130 on A/H interface, M131 on A/G/H interface, and F123 on G/H loop—all shown<sup>30</sup> to stabilize the I state—remain buried. This intermediate appears to be similar to the experimentally observed<sup>16</sup> equilibrium molten globule state at pH 4 (I-state) and the obligatory re-folding intermediate.<sup>19</sup> The core helices retain a substantial fraction of their native helicity, A: 0.47, B: 0.53, G: 0.71, and H: 0.78, while the relative helical content of the longest helix outside the core, E, is only 0.25. These values are computed over a 50 ps interval around  $t = 0.225 \text{ ns}$  (Figure 3), and are in the best agreement with experimental data<sup>16</sup> on the equilibrium I-state, where the corresponding numbers are A: 0.7, B: 0.45, G: 0.6, H: 0.55, and E: 0.3, respectively. Overall, approximately half of the native contacts and helical population (relative to the native holo-myoglobin) are lost, but almost all of the remaining contacts are native (Figure 3, 0.22 ns). Topologically similar I-state-like structures with significant portions of the A, G, H, and B helices forming a compact core are observed from ~0.2 to ~0.55 ns along the unfolding trajectory (Figure 3, 0.46 ns). Tirado-Rives & Jorgensen<sup>7</sup> also observed I-like helical populations in the final stages of both pH 6 and pH 4.2 (all His protonated) simulations at 358 K. However, they observed only a small relative increase in radius of gyration, and less loss of helicity than reported here; in particular, they observed a higher population of the E helix. The overall topology of the I-state predicted by Tirado-Rives & Jorgensen is close to that of apomyoglobin, in contrast to our findings (Figure 3, 0.22 ns).

The question of exactly how molten is the molten globule<sup>31</sup> has been a long standing one; it turns on the extent of structural fluctuations in this state, which are not easily quantified by experiment. A comparison of fluctuations of  $\chi_2$  torsional angles of some key residues in the molten globule portion of the unfolding trajectory with those of the pH 6.5 native-state apomyoglobin trajectory identifies some regions of the compact core as remaining essentially native-like, while others have significant

fluctuations indicative of a rotamer transition (Table 2). Specifically, the G/H loop around Phe123, and the A/G helix interface around Leu115 remain in a tight, virtually native conformation in the I-like stage of the simulation, while the A/G/H interface around Met131, and the A/H interface around Trp7 are characterized by significant fluctuations indicative of a molten character. Interestingly, Met131 still retains its hydrophobic contacts with a number of residues at this stage (up to  $t = 0.55$  ns): Leu9, Val10, Leu115, and Phe123.

During the next stage of unfolding (Figure 3, 1.27 ns), the protein's  $R_g$  increases further to  $\sim 25$  Å, and only  $\sim 35\%$  of the native helical structure remains (Figure 1). Parts of the A, G, and H helices still form a compact core, with most of the remaining contacts being native, although some non-native contacts appear in the A/G interface. A noteworthy feature is the persistence of the native contacts around the GH loop. The highly conserved Phe123 extends into this loop, forming contacts with residues on both the A and H helices. A compact intermediate with a stable core formed by parts of the A, G, and H helices is observed in the earliest steps of apomyoglobin re-folding currently accessible to experiment,<sup>32</sup> indicating yet another common point on the simulated unfolding and experimental folding pathways. As the unfolding simulation proceeds to around 2.22 ns, the G/H contacts disappear, and only a few contacts on the A/H interface remain, most of which are non-native (Figure 3). At this stage, the protein's radius of gyration,  $\sim 27$  Å, and relative helical content,  $\sim 20\%$ , begin to reach the experimentally observed<sup>16</sup> values for the completely unfolded state at pH 2 (Figure 1). The contacts between the N terminus and the C terminus of the H helix appear to be the last ones lost (Figure 3, 2.55 ns).

#### The acid-unfolded stage

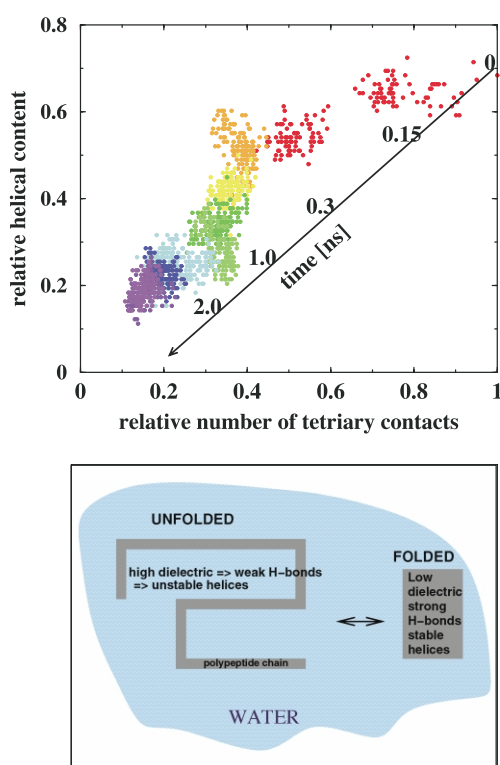
The predicted average characteristics of the mostly unfolded stage near the end of the trajectory, which are exemplified by two snapshots in Figure 3 (2.70 and 2.92 ns), agree quite well with experimental<sup>16</sup> data on apomyoglobin equilibrated at pH 2 (Table 1), as does the distribution of the helical population among the major helical regions: the calculated values for A, B, E, F, G and H helices are 0.22, 0.14, 0.1, 0.0, 0.1, and 0.27, respectively, while the corresponding experimental<sup>16</sup> numbers are 0.2, 0.1, 0.07, 0.0, 0.07, and 0.25. Our simulations provide further insights into the unfolded state—a state which is hard to investigate experimentally due to large scale structural fluctuations.

The acid-unfolded structures seen at  $t \gtrsim 2.7$  ns are far from being random coil. They retain some ( $\sim 20\%$ ) helical structure. As is evident from Figure 3, no native tertiary (off-diagonal) contacts remain, but a number of new non-native contacts appear. In particular, local clusters are formed in the B helix region, which is not surprising given that it is rich in hydrophobic amino-acids: V21, A22,

G23, G25, I28, L29, I30, L32, P33. Local hydrophobic clusters in this region have also been detected experimentally.<sup>17</sup> Interestingly, we also observe formation of non-native, albeit transient, contacts between the N-terminal region and CD loop–D helix regions (Figure 3, 2.92 ns). We attribute this to the fairly hydrophobic nature of the N-terminal region (VLSEG), especially at low pH when glutamic acid is neutral. Curiously, myoglobins from very different organisms all appear to possess this “sticky tail”, although the particular amino-acid sequences are different: GLSDG in horse, pig, human, elephant; VLSEG in sperm whale, rat; VAFTE in soybean. Note that the same “tail” participates in hydrophobic contacts with the H helix during the final stages of unfolding (Figure 3, 2.22 and 2.55 ns). In light of the similarities between apomyoglobin acid-unfolding and folding pathways, it would be interesting to see how alterations in the sequence of the N-terminal region that render it much less hydrophobic would affect the folding kinetics.

#### Origins of cooperativity between tertiary and secondary structure formation

The simulated trajectory provides great detail on the amount of native-like secondary structure and tertiary contacts at various stages of unfolding, and thus, a probe of the extent of cooperativity in the dissolution/formation of these aspects of protein structure. As seen from Figure 4, loss of helical structure and loss of native tertiary contacts tend to go hand-in-hand during most of the unfolding trajectory, that is, at times beyond the first 0.15 ns. Assuming that the folding process qualitatively resembles the reverse of the unfolding process, this implies a cooperative model of secondary and tertiary structure folding/unfolding, as opposed to a sequential model in which the folding/unfolding of one sort of structure occurs in a time range distinct from that of the other sort of structure. Although a number of different mechanisms may contribute to the coupling between the secondary and tertiary structure formation, such as compaction and steric interactions,<sup>33</sup> the energetic origin of the trends throughout Figure 4 can be rationalized in terms of macroscopic dielectric models. These predict that the hydrogen-bonding interactions that stabilize secondary structure are greatly weakened in the high-dielectric environment of water, but that the presence of lower-dielectric peptidic material (even if it does not directly reduce the water-accessibility of the H-bonding groups), can significantly strengthen the H-bonding interactions and promote the formation of  $\alpha$ -helices.<sup>34</sup> A similar correlation between helix formation and dielectric constant has been seen in experiments on  $\beta$ -lactoglobulin in organic solvents.<sup>35</sup> Therefore, the formation of tertiary structure, which introduces more low-dielectric material into the environment of the backbone amides, strengthens the H-bonds responsible for

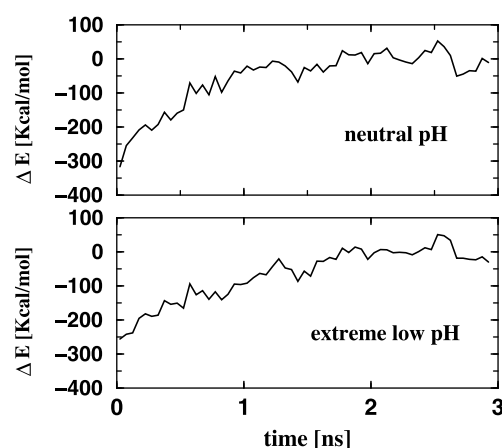


**Figure 4.** Top: correlation between relative helical content and number of tertiary contacts present in apomyoglobin during its unfolding. Ten rainbow colors from red to violet represent times along the unfolding trajectory that fall, in 0.3 ns segments from  $t = 0$  (red) to  $t = 3$  ns (violet). The trivially correlated intra-helical contacts are excluded by showing only contacts between residues four-apart or more in the sequence. Bottom: a rationalization based on continuum dielectric models for the coupling between secondary and tertiary structure formation. The formation of tertiary structure, which introduces more low-dielectric material into the environment of the backbone amides, strengthens the H-bonds responsible for stabilizing secondary structure.

stabilizing secondary structure. This explains the cooperativity observed in most of Figure 4 (e.g. for times greater than 0.15 ns). As for the first 0.15 ns of the unfolding trajectory, during which tertiary contacts are not accompanied by much change in relative helicity (top right corner of Figure 4), this is the time range of the native-to-dry swollen globule transition in which the degree of water penetration into the interior is not much changed (Figure 2), and so the dielectric environment of the H-bonding groups, and therefore, the  $\alpha$ -helical content is also not much changed. The mechanism outlined above is expected to be most effective for medium or large size proteins where the differences in dielectric properties between the “inside” and the “outside” are most pronounced.

### Estimates of folding free energy

We can also use continuum solvent models to estimate the free energies of apomyoglobin folding



**Figure 5.** Overall energy profile of unfolding apomyoglobin as a function of time calculated using charges characteristic of neutral pH (top) or strongly acidic (bottom) conditions, averaged over consecutive intervals of 50 ps along the unfolding trajectory. The completely unfolded state is taken to be the reference state ( $\Delta E = 0$ ).

at neutral and acidic pH. In general, theoretical calculations of the folding free energy using all-atom models for proteins and solvent are extremely difficult, and have, so far, been done only for small systems, such as a 46-residue three-helix bundle.<sup>11</sup> The major difficulty here lies in the need to sample the enormous conformational space of protein/solvent system. Another problem usually arises from the uncertainty about the structure of the unfolded state, but here the unfolding trajectory provides a reasonable ensemble of structural models. To avoid the first difficulty, we use the Poisson–Boltzmann/surface area approach<sup>36,37</sup> in which solvent is represented implicitly as continuum with the dielectric properties of water, and the protein structure variability is represented by the ensemble of snapshots. For the present calculations, the ionic strength is taken as zero and the resulting Poisson equation is solved numerically (see Methods). The hydrophobic effect is taken into account by a surface-energy term proportional to the solvent-accessible area of the protein. These terms, together with the intra-protein terms of the molecular mechanics force-field, comprise our estimate of  $E$ , the free energy terms excluding chain entropy, for a given structural snapshot. Ensembles of structures representing the folded and unfolded states are taken from the appropriate time segments of the unfolding simulation, and calculation of energies corresponding to neutral or acidic pH is accomplished by setting ionizable side-chain charges to the appropriate values. The details of these calculations, as well as the methods of estimating chain entropy, are given in Methods; and average energies,  $\Delta E$ , for various parts of the trajectory are presented in Figure 5. As expected, the energy decreases in going from the unfolded to the native state, consistent with the idea that the latter corresponds to the global energy minimum. The overall energy change upon folding is higher at neutral

**Table 3.** Energetics of apomyoglobin folding

Conditions	$\Delta E_{\text{int}}$	$\Delta E_{\text{elec}}$	$\Delta E_{\text{vdw}}$	$\Delta E_{\text{surf}}$	$\Delta E$	$-T\Delta S_c$	$\Delta G$ (calc)	$\Delta G$ (exp)
Neutral pH	$\sim 0$	+71	-354	-28	-311	+294	-17	-13
Extreme acidic pH	$\sim 0$	+132	-354	-28	-250	+294	+44	?

Energy, entropy, and free energy  $\Delta G = \Delta E - T\Delta S_c$  components at neutral and strongly acidic conditions at zero ionic strength. All energies are in kcal/mol and refer to unfolded to folded transition.  $\Delta S_c$  is the change in a protein's conformational (backbone and side-chain) entropy. Experimental value of  $\Delta G$  of folding at neutral pH is from Ref. 56.

pH:  $\Delta E_{\text{neut}} = -311$  kcal/mol *versus*  $\Delta E_{\text{acid}} = -250$  kcal/mol at acidic conditions. The large favorable  $\Delta E$  of folding is offset by almost equal and opposite entropy component which favors unfolding. We estimate protein's change in conformational entropy in going from the folded to the completely unfolded state using per-residue parameters proposed by D'Aquino *et al.* and Doig & Sternberg<sup>38</sup> (see Methods). The key results are summarized in Table 3, showing that the predicted free energy of folding at neutral pH is in agreement with experiment, and that (as expected) folding should be very unfavorable at low pH.

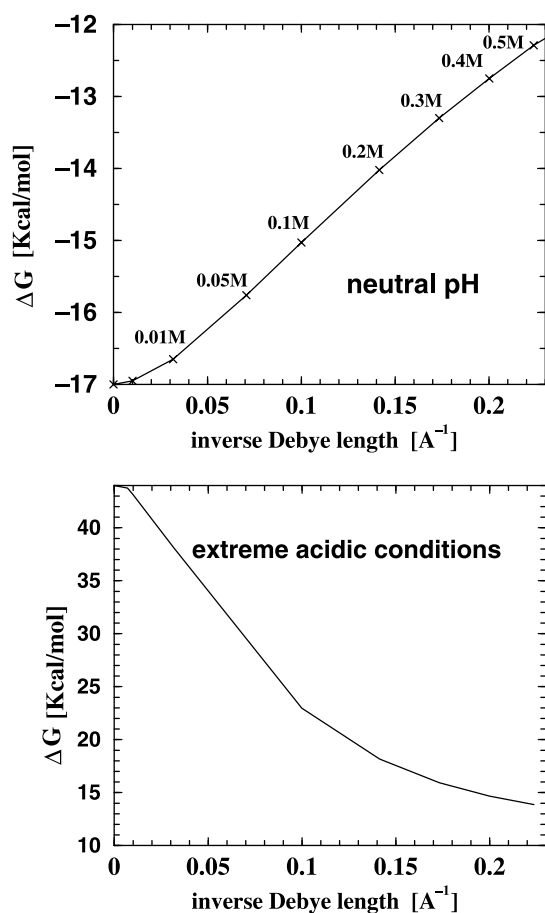
To gain further insights into the folding process we break down the overall energy contribution into components according to:  $E = E_{\text{int}} + E_{\text{elec}} + E_{\text{vdw}} + E_{\text{surf}}$ , where  $E_{\text{int}}$  represents the energy of internal degrees of freedom such as bond stretching,  $E_{\text{elec}}$  is the total electrostatic (free) energy including solvation,  $E_{\text{vdw}}$  corresponds to Van der Waals interactions between protein atoms, and  $E_{\text{surf}}$  mimics the hydrophobic effect. The results for apomyoglobin are summarized in Table 3. Note that internal degrees of freedom such as bond stretching have essentially no effect on folding,  $\Delta E_{\text{int}} \sim 0$ , but among other contributions to the total  $\Delta E$  of folding at neutral pH listed in Table 3, the dominant one is the Van der Waals interaction between protein atoms which favors tightly packed structures. We find the overall electrostatic contribution to disfavor folding of apomyoglobin at neutral pH, in agreement with earlier theoretical predictions using a similar energetic model.<sup>39</sup> In the neutral pH calculations, large  $\Delta E$  terms favoring folding are balanced against a large conformational entropy term favoring unfolding to produce a net folding free energy near zero, in agreement with the general experimental finding that overall protein stability is about 13 kcal/mol. This result is quite satisfying given the large size of the individual terms, and the general difficulty of calculating large protein conformational changes. However, given the uncertainty associated with the chain entropy estimate, the close agreement with experiment ( $-17$  kcal/mol calculated *versus*  $-13$  kcal/mol, experimental) is probably only fortuitous. On the other hand, the calculated difference between the  $\Delta G$  of folding at acidic *versus* neutral pH is independent of the chain entropy estimate, and can therefore be used, together with the experimental  $\Delta G$  of folding at neutral pH, to obtain an independent estimate of the  $\Delta G$  of folding under acid

conditions. From Table 3 the calculated difference,  $\Delta G_{\text{neut}} - \Delta G_{\text{acid}}$  is  $+61$  kcal/mol. Adding this to the experimental value of  $-13$  kcal/mol for  $\Delta G_{\text{neut}}$  gives an estimate of  $+48$  kcal/mol for  $\Delta G_{\text{acid}}$  in the limit when all titratable groups are protonated regardless of protein conformation, corresponding to the conditions of extremely low pH. Knowledge of the structural details of the acid unfolded state appears to be necessary for an accurate estimate of this quantity: Yang & Honig<sup>40</sup> used a simplified model of the unfolded state and obtained  $\Delta G_{\text{acid}} \sim +80$  kcal/mol, which is significantly larger than our estimate. It should be noted that our estimate corresponds to zero ionic strength, which can hardly be achieved under experimental conditions of very low pH. As we will see in the following section, the effect of counter-ions on protein stability, while being relatively small at neutral pH, becomes more significant at low pH (when all titratable groups become protonated) where it tends to lower  $\Delta G_{\text{acid}}$  substantially, in general agreement with experimental findings.<sup>41</sup>

### Salt effects

The predicted salt dependence of the  $\Delta G$  of folding was calculated within the continuum solvent framework, using a generalized Born-type approximation (see Methods) and is presented in Figure 6. In agreement with general experimental observations, salt destabilizes folded apomyoglobin at neutral pH (making the  $\Delta G$  of the unfolded-to-folded transition less negative), but has a very strong stabilizing effect at low pH (lowers  $\Delta G$ ). Although we know of no experimental data for apomyoglobin, the predicted decrease in stability of  $\sim 2$  kcal/mol in going from 0 M to 0.1 M NaCl at neutral pH is similar to that observed experimentally on cyanomyoglobin.<sup>42</sup> The inverse Debye length  $D^{-1} \approx 0.316\sqrt{[\text{NaCl}]} \text{ \AA}^{-1}$ , is an appropriate length-scale of the system, and so plotting  $\Delta G$  *versus*  $D^{-1}$  clarifies its salt dependence better than the more common  $\Delta G$  *versus* salt concentration plot. In particular, one does not expect any significant (non-specific) salt effects on folding free energy if the Debye length is considerably larger than the system size,  $D \gg R_g$ . For a medium-size protein this is equivalent to monovalent salt concentration being less than  $\sim 0.005$  M—notice that both lines in Figure 6 do not extrapolate linearly to the origin, but level-off for small  $D^{-1}$ . At very low pH, the effects of non-zero anion concentration must be taken into account<sup>41</sup> to obtain an estimate





**Figure 6.** Predicted salt dependence of the free energy of the unfolded-to-folded transition of apomyoglobin at neutral pH and acidic pH. At neutral pH salt reduces the stability of the native state relative to the unfolded state, while under strongly acidic conditions it has the opposite effect. The lines do not extrapolate linearly to the origin, but level, indicating cross-over to a regime where the Debye length  $\gg$  system size, and the folding free energy becomes nearly salt-independent.

of  $\Delta G_{\text{acid}}$  that can be directly compared with experiment. Assuming  $[\text{Cl}^{-1}] = 0.1 \text{ M}$ , we estimate  $\Delta G_{\text{acid}} \approx 22 \text{ kcal/mol}$  from Figure 6. Extrapolating using the model of Barrick & Baldwin,<sup>43</sup> which is based on their experimental data, gives  $\Delta G_{\text{acid}} \approx 18 \text{ kcal/mol}$  for apomyoglobin at pH 1.

It is commonly believed that, by virtue of decreasing charge-charge interactions, salt should disfavor folding if electrostatic interactions favor it, and *vice versa*. Without a more detailed analysis, this over-simplified picture may be misleading, as the following example demonstrates. From Figure 6, the effect of adding 0.1 M salt is predicted to disfavor folding by  $\approx 2 \text{ kcal/mol}$  at neutral pH, but to favor folding by  $\approx 20 \text{ kcal/mol}$  at low pH. However,  $\Delta E_{\text{elec}} > 0$  at all pH values (Table 3). The seeming paradox is resolved by breaking down  $E_{\text{elec}}$  into its two distinct components:  $E_{\text{elec}} = E^{\text{self}} + E^{\text{cross}}$ , where the first term represents the sum of solvation (Born) energies of individual charges, and the second corresponds to pairwise inter-

actions between them. Since it is energetically unfavorable to bury a charge inside the folded protein,  $E^{\text{self}}$  always disfavors folding. This term outweighs the stabilizing effects of  $E^{\text{cross}}$ , resulting in  $\Delta E_{\text{elec}} > 0$  at both neutral and acidic conditions. Increasing ionic strength tends to modestly increase  $\Delta E^{\text{self}}$ , regardless of pH, reflecting the fact that in the unfolded state the charges are more exposed to solvent than in the native state and  $E^{\text{self}}$  of the unfolded state is, therefore, more strongly affected (lowered) by the addition of salt. For apomyoglobin, an increase in monovalent salt concentration from 0.0 M to 0.1 M causes  $\Delta E^{\text{self}}$  to increase by +0.9 kcal/mol at neutral pH, and +0.8 kcal/mol at low pH. In contrast, the corresponding changes in  $\Delta E^{\text{cross}}$  are strongly pH-dependent:  $\Delta E^{\text{cross}}$  increases by +1.1 kcal/mol at neutral pH, while under strongly acidic conditions it decreases by  $-21.8 \text{ kcal/mol}$ , since in this case salt tends to decrease the unfavorable repulsion between extra positive charges. Hence, one cannot conclude, simply from the sign of the salt dependence of folding free energy, whether the net effect of electrostatic interactions is to favor or disfavor folding.

## Concluding remarks

We have carried out a molecular dynamics simulation of the acid-induced unfolding of apomyoglobin at room temperature, including all protein atoms and  $\sim 13,000$  water molecules, with no added biasing forces or other unusual conditions to induce unfolding. It is generally believed that the folding and acid unfolding of this protein follow similar pathways,<sup>19,20</sup> and thus, we have, in reverse sequence, a series of detailed structural models for states along the folding pathway.

During the initial stages of unfolding the protein globule swells slightly, disrupting a number of native tertiary contacts. However, its fold geometry remains practically unchanged, and water does not enter the protein interior, so the disruption of energetically favorable native contacts is not compensated by an increase in favorable protein-water contacts. This dry swollen globule is therefore likely to be an energetic barrier around the native state. Further along the simulated unfolding pathway we find a partially unfolded, intermediate stage in which helices A, B, G and H form a compact core, and rest of the protein is considerably less structured. This intermediate appears to be similar to the experimentally observed equilibrium molten globule I-state at pH 4. Our analysis of structural fluctuations in the core region shows that it has characteristics of both folded (fairly rigid) and the molten (relatively loose) states.

During the final stages of unfolding, the contacts between the N terminus and the C terminus of the H helix appear to be the last ones lost. The acid-unfolded state, which we observe towards the end of the simulation, is far from being a random coil.

In agreement with experiment, it retains about 20% of residual helical structure. We also find that although no off-diagonal native contacts remain, some non-native, albeit transient, tertiary contacts appear, in particular between the N-terminal region and C/D loop—D helix regions. We attribute this to the fairly hydrophobic nature of the N-terminal region (VLSEG), especially at low pH when glutamic acid is neutral. Clusters of hydrophobic residues involving distant regions of the sequence have also been recently observed experimentally in lysozyme under strongly denaturing conditions.<sup>44</sup>

By quantifying the amounts of native-like secondary structure and tertiary contacts at various stages of simulated unfolding, we find that loss of helical structure and loss of native tertiary contacts tend to go hand-in-hand during most of the unfolding trajectory, except in the very beginning when the protein is in the dry swollen globule state. These trends are those that would be expected from changes in the dielectric environment of H-bonding groups: electrostatic interactions responsible for helix stabilization generally become weaker as the protein unfolds and the groups become more exposed to water.

Finally, we have explored the energetics of apomyoglobin folding by using a continuum electrostatic model to represent solvent effects implicitly. As expected, the native state is considerably lower in energy than the unfolded state by  $\Delta E \sim 300$  kcal/mol. This favorable folding energy is nearly offset (at neutral pH) by the entropy component which favors unfolding, the latter being estimated from published per-residue parameters. Remarkably, the two contributions nearly cancel to yield a reasonably small ( $\sim 10$  kcal/mol) folding free energy. We find that electrostatic interactions disfavor folding at all pH and salt concentrations, and reconcile this finding with the experimentally observed destabilizing effect of monovalent salt on the native state at neutral pH.

## Methods

### MD simulations at neutral pH

We use version 5.0 of the AMBER suite of programs. An all-atom force field<sup>45</sup> is employed. The SHAKE method is used to restrain hydrogen-heavy atom bond distances. The integration time-step is 2 fs, with a 12 Å cut-off for long-range interactions. The average temperature of the system is maintained at 300 K by coupling to a heat bath with coupling constant of 2 ps. The starting structure is prepared by removing the heme group from the holo-Mb coordinate set (PDB-ID: 2mb5) obtained by neutron diffraction.<sup>46</sup> We keep all the hydrogen atoms found in the PDB set, except for a few histidine residues whose protonation state we change to make the total charge of the protein +5. The three most basic histidine residues 36, 81 and 116 are doubly protonated, and all the remaining histidine residues are in their uncharged form. (The charge of all Glu and Asp residues is  $-1$ , and that of all Arg and Lys is  $+1$ .) This state of the

protein corresponds roughly to neutral pH. We follow a droplet simulation protocol, in which the protein is solvated inside a large droplet of water. It was shown earlier<sup>47</sup> that around a thousand water molecules is enough to fully hydrate myoglobin in an MD simulation. In our simulations, we use a much larger number of water molecules to avoid possible edge effects by ensuring at least a 10 Å buffer between the protein and the surface of the droplet.

The protein is solvated by  $\approx 13,000$  TIP3P water molecules, forming a spherical droplet of radius  $\sim 55$  Å around the molecule's center of mass. Each simulation cycle begins with a 100 steps of steepest-descent minimization followed by 100 ps of equilibration during which the temperature is gradually raised to 300 K, while the protein atom coordinates remain fixed by harmonic restraints (force constant  $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ) to their positions at the end of the previous cycle (stage 1); for the first cycle, crystallographic positions are used. After the equilibration is completed the constraints are removed, and the simulation continues for another 100 ps at 300 K. Protein and solvent coordinates are saved every 10 ps (stage 2). At the end of this stage, the water molecules are removed, and the protein is re-solvated by  $\approx 13,000$  water molecules forming a spherical droplet of radius 50 Å around protein's center of mass (stage 3). The above three-stage cycle is then repeated 20 times, yielding a total of 2.0 ns of unconstrained trajectory. The re-solvation procedure ensures that the protein always stays well within the droplet during the simulation. Note that very little global unfolding has occurred over the 2 ns of the neutral-pH simulation (Table 1 and Figure 5); the radius of gyration has increased by only a few percent. Therefore, the use of droplet protocol does not, by itself, induce global unfolding on the time-scale of a few nanoseconds.

### Acid-unfolding MD simulations

To model the conditions of extremely low pH, all aspartate, glutamate, and histidine side-chains, and the C terminus are protonated, making the overall charge of the globule +36, in agreement with the experimentally observed value under these conditions. We then continue the simulation for 3 ns using the protocol described above, but switch to a very large (24 Å) cut-off for non-bonded interactions to better approximate long-range forces. No re-solvation is performed within first 200 ps of the acid-unfolding simulation. Protein and solvent coordinates are saved every 2 ps. The computation took approximately 60 days on 16 CPUs of R12000 SGI Origin 2800 machine.

To verify that the re-solvating procedure or edge effects do not introduce any extra bias for unfolding, two tests were performed. First, a larger droplet of  $\sim 60$  Å was used to carry out a 200 ps low pH simulation, again without re-solvation at 100 ps, starting from the same apomyoglobin model as above. During the simulation, the protein stayed well within the droplet, and in the end, its structural characteristics were almost identical with the ones obtained before. Note that 200 ps is enough to induce considerable unfolding of apomyoglobin at low pH, according to Figures 1 and 3. Second, we have repeated the first 200 ps of the unfolding protocol with re-solvation at 100 ps, but retained all water molecules within 5 Å of the protein upon re-solvation (and added more to make the full droplet). Again, the unfolding proceeded along a pathway virtually identical

with the one presented in Figure 1,  $t = 0$ –200 ps. To further verify that the conclusions we have made are not sensitive to the details of the initial structure or the MD protocol, we have performed a separate unfolding simulation of apomyoglobin (results not presented here) which is different from the above as follows: the size of the water droplet used to solvate the protein is increased to  $\sim 65$  Å, and the cut-off for non-bonded interactions is reduced to 12 Å to make computations feasible. We perform a 1.6 ns simulation at 300 K starting from the holo-Mb coordinate set (PDB ID 2Mb5) without the heme group. As before, all titratable groups in the protein are protonated to model the conditions of acid shock. We find that the unfolding proceeds through a set of intermediate states generally similar to the ones described here.

### Calculation of structural signatures

Relative helical content (helicity) is defined as the ratio of the number of 1–4 hydrogen bonds (computed using DSSP<sup>48</sup> program) in the current conformation relative to that in the native holo protein.

Two residues are considered to form a contact if a water molecule, for this purpose considered to be a sphere of radius 1.4 Å, cannot fit between them, i.e. if there exists a pair of atoms not on the same residue such that the distance between the atomic centers minus their radii is less than 2.8 Å. To identify solvent exposed residues, we analyze positions of all water molecules surrounding the protein during the simulation. A residue is considered exposed to solvent if there exists an atom on its side-chain whose distance to an atom of at least one of the surrounding water molecules is less than 1 Å. The Bondi radius set<sup>49</sup> is used for the computations. The set of residues which become solvent exposed in going from the dry swollen globule state to the more open I-state is identified as follows: it consists of residues which are not exposed to solvent in the time interval 0.05–0.15 ns, but are exposed during 0.2–0.3 ns interval of the acid-unfolding trajectory. Each interval is represented by 20 consecutive snapshots, a residue is considered solvent exposed during the given time interval if it is exposed in more than 14 (i.e.  $\sim 70\%$ ) of the snapshots.

### Continuum solvent calculations

The folded state is represented by 25 conformations taken from the native-like portion of the acid-unfolding trajectory ( $t = 0.002$ – $0.052$  ns) with charges set appropriately to neutral pH. The unfolded state is represented by 500 consecutive conformations from  $t = 1.964$  ns to  $t = 2.964$  ns. The total energy of the solvated protein is calculated via  $E = E_{\text{int}} + E_{\text{elec}} + E_{\text{vdw}} + E_{\text{surf}}$  (terms defined in the main text).  $E_{\text{elec}}$  is computed as  $E_{\text{elec}} = E_{\text{vac}} + E_{\text{solv}}$ , where  $E_{\text{vac}}$  is the protein's Coulomb energy in vacuum and  $E_{\text{solv}}$  is the electrostatic component of the free energy of solvation.  $E_{\text{vac}}$ ,  $E_{\text{int}}$ ,  $E_{\text{elec}}$ ,  $E_{\text{vdw}}$ , are computed for each snapshot using the AMBER-5 force field parameters. Numerical solution of the Poisson equation to compute  $E_{\text{solv}}$ , is done using DELPHI-II<sup>50</sup> with a cubic box with 211 grid points in each direction. The dielectric constant of protein interior is 1 and the ionic strength is zero. We use (kcal/mol)  $E_{\text{surf}} = 0.005A$ , where  $A$  [Å<sup>2</sup>] is the calculated solvent accessible area of protein.<sup>36</sup> The change in a protein's conformational (backbone and side-chain) entropy upon folding is esti-

mated as:  $\Delta S_c = \Delta S_{\text{back}} + \Delta S_{\text{side}}$ , where for a completely unfolded protein  $\Delta S_{\text{back}} \sim 5$  cal mol<sup>-1</sup> K<sup>-1</sup> per residue<sup>38</sup> and  $\Delta S_{\text{side}} \sim 3$  cal mol<sup>-1</sup> K<sup>-1</sup> per residue.<sup>38</sup> To account for 20% residual helical structure in the acid-unfolded state of apomyoglobin we reduce the above value of  $\Delta S_c$  by 20%, and use  $\Delta S_c \sim 6.4$  mol/K per residue. ( $E_{\text{elec}}$  and  $E_{\text{surf}}$  contain other entropic effects implicitly).

The salt dependence of  $\Delta G$  of folding is calculated as  $\Delta G(\text{salt}) = \Delta G(\text{salt} = 0) + [\Delta E_{\text{elec}}(\text{salt}) - \Delta E_{\text{elec}}(\text{salt} = 0)]$ . To reduce computational time, we employ (for  $\Delta G(\text{salt})$  calculations only, which are not expected to be very sensitive to the details of the electrostatic model) the Generalized Born (GB) approximation<sup>51</sup> to estimate electrostatic energies, and use 50 consecutive conformations from  $t = 2.866$  ns to  $t = 2.964$  ns to represent the unfolded state. The GB model we use was demonstrated to be a reasonable approximation to PB for myoglobin<sup>51</sup> and other systems,<sup>52</sup> and was shown to describe the electrostatic effects of monovalent salt adequately.<sup>53</sup>

### Acknowledgements

We thank Peter Wright, Jane Dyson, Charles Brooks, Paul Schimmel, Alexandra Dyuysekina, Brian Dominy and Victoria Lunyak for helpful discussions. The work was supported by NIH grant GM 57513.

### References

- Duan, Y. & Kollman, P. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, **282**, 740–744.
- Kolinski, A. & Skolnick, J. (1994). Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins: Struct. Funct. Genet.* **18**, 353–366.
- Onuchic, J. N., Nymeyer, H., Garcia, A. E., Chahine, J. & Succi, N. D. (2000). The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Advan. Protein Chem.* **53**, 87–152.
- Zhou, Y. & Karplus, M. (1999). Interpreting the folding kinetics of helical proteins. *Nature*, **401**, 400–403.
- Shakhnovich, E. I. (1997). Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* **7**, 29–40.
- Mayor, U., Johnson, C. M., Daggett, V. & Fersht, A. R. (2000). Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl Acad. Sci. USA*, **97**, 13518–13522.
- Tirado-Rives, J. & Jorgensen, W. L. (1993). Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry*, **32**, 4175–4184.
- Smith, L. J., Dobson, C. M. & van Gunsteren, W. F. (1999). Molecular dynamics simulations of human alpha-lactalbumin: changes to the structural and dynamical properties of the protein at low pH. *Proteins: Struct. Funct. Genet.* **36**, 77–86.
- Mao, Y., Ratner, M. A. & Jarrold, M. F. (1999). Molecular dynamics simulations of the charge-induced unfolding and refolding of unsolvated cytochrome C. *J. Phys. Chem.* **103**, 10017–10021.
- Paci, E., Smith, L. J., Dobson, C. M. & Karplus, M. (2001). Exploration of partially unfolded states of

- human alpha-lactalbumin by molecular dynamics simulation. *J. Mol. Biol.* **306**, 329–347.
11. Boczeko, E. M. & Brooks, C. L., III (1995). First-principles calculation of the folding free energy of a three-helix bundle protein. *Science*, **269**, 393–396.
  12. Day, R., Bennion, B. J., Ham, S. & Daggett, V. (2002). Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.* **306**, 329–347.
  13. Caflich, A. & Karplus, M. (1995). Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *J. Mol. Biol.* **252**, 672–708.
  14. Wright, P. E. & Baldwin, R. L. (2000). *Frontiers in Molecular Biology: Mechanisms of Protein Folding* (Pain, R., ed.), pp. 309–329, Oxford University Press, London.
  15. Lecomte, J. T. J., Sukits, S. F., Bhattacharya, S. & Falzone, C. J. (1999). Conformational properties of native sperm whale apomyoglobin in solution. *Protein Sci.* **8**, 1484–1491.
  16. Eliezer, D., Yao, J., Dyson, H. J. & Wright, P. E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nature Struct. Biol.* **5**, 148–155.
  17. Yao, J., Chung, J., Eliezer, D., Wright, P. & Dyson, J. (2001). NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry*, **40**, 3561–3571.
  18. Jamin, M., Yeh, S. R., Rousseau, D. L. & Baldwin, R. L. (1999). Submillisecond unfolding kinetics of apomyoglobin and its pH 4 intermediate. *J. Mol. Biol.* **292**, 731–740.
  19. Jamin, M. & Baldwin, R. L. (1998). Two forms of the pH 4 folding intermediate of apomyoglobin. *J. Mol. Biol.* **276**, 491–504.
  20. Tsui, V., Garcia, C., Cavangero, S., Sizudak, G., Dyson, H. J. & Wright, P. E. (1999). Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate. *Protein Sci.* **8**, 45–49.
  21. Tcherkasskaya, O. & Ptitsyn, O. B. (1999). Molten globule versus variety of intermediates: influence of anions on pH-denatured apomyoglobin. *FEBS Letters*, **455**, 325–331.
  22. Brooks, C. L., III (1992). Characterization of native apomyoglobin by molecular dynamics simulation. *J. Mol. Biol.* **227**, 375–380.
  23. Shakhnovich, E. & Finkelstein, A. (1989). Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers*, **28**, 1667–1680.
  24. Kiefhaber, T., Labhardt, A. M. & Baldwin, R. L. (1995). Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature*, **375**, 513–515.
  25. Hoeltzli, S. D. & Frieden, C. (1995). Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-<sup>19</sup>F-tryptophan-labeled *Escherichia coli* dihydrofolate reductase. *Proc. Natl Acad. Sci. USA*, **92**, 9318–9322.
  26. Pappenberger, G., Saudan, C., Becker, M., Merbach, A. E. & Kiefhaber, T. (2000). Denaturant-induced movements of the transition state of protein folding revealed by high-pressure stopped-flow measurements. *Proc. Natl Acad. Sci. USA*, **97**, 17–22.
  27. Feng, Z., Ha, J. & Loh, S. N. (1999). Identifying the site of initial tertiary structure disruption during apomyoglobin unfolding. *Biochemistry*, **38**, 14433–14439.
  28. Kiefhaber, T. & Baldwin, R. L. (1995). Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulse hydrogen exchange. *Proc. Natl Acad. Sci. USA*, **92**, 2657–2661.
  29. Daggett, V. & Levitt, M. (1992). A model of the molten globule state from molecular dynamics simulations. *Proc. Natl Acad. Sci. USA*, **89**, 5142–5146.
  30. Kay, M. S. & Baldwin, R. L. (1996). Packing interactions in the apomyoglobin folding intermediate. *Nature Struct. Biol.* **3**, 439–445.
  31. Ptitsyn, O. (1996). How molten is the molten globule? *Nature Struct. Biol.* **3**, 488–490.
  32. Ballew, R. M., Sabelko, J. & Gruebele, M. (1996). Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl Acad. Sci.* **93**, 5759–5764.
  33. Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133–7155.
  34. Ösapay, K., Young, W. S., Bashford, D., Brooks, C. L., III & Case, D. A. (1996). Dielectric continuum models for hydration effects on peptide conformational transitions. *J. Phys. Chem.* **100**, 2698–2705.
  35. Uversky, V., Narizheva, N., Kirschstein, S., Winter, S. & Löber, G. (1997). Conformational transitions provoked by organic solvents in beta-lactoglobulin: can a molten globule like intermediate be induced by the decrease in dielectric constant? *Fold. Des.* **2**, 163–172.
  36. Demchuk, E., Bashford, D., Gippert, G. P. & Case, D. A. (1997). Thermodynamics of a reverse turn motif. Solvent effects and side-chain packing. *J. Mol. Biol.* **270**, 305–317.
  37. Srinivasan, J., Cheatham, T. E., Kollman, P. & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA and phosphoramidate–DNA helices. *J. Am. Chem. Soc.* **120**, 9401–9409.
  38. Brady, G. P. & Sharp, K. A. (1997). Entropy in protein folding and in protein–protein interactions. *Curr. Opin. Struct. Biol.* **7**, 215–221.
  39. Yang, A. S. & Honig, B. (1993). On the pH dependence of protein stability. *J. Mol. Biol.* **231**, 459–474.
  40. Yang, A. S. & Honig, B. (1994). Structural origins of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin. *J. Mol. Biol.* **237**, 602–614.
  41. Goto, Y., Calciano, L. J. & Fink, A. L. (1990). Acid-induced folding of proteins. *Proc. Natl Acad. Sci. USA*, **87**, 573–577.
  42. Ramos, C. H. I., Kay, M. S. & Baldwin, R. L. (1999). Putative inter-helix ion pairs involved in the stability of myoglobin. *Biochemistry*, **38**, 9783–9790.
  43. Barrick, D. & Baldwin, R. L. (1993). Three-state analysis of sperm whale apomyoglobin folding. *Biochemistry*, **32**, 3790–3796.
  44. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T. *et al.* (2002). Long-range interactions within a nonnative protein. *Science*, **295**, 1719–1722.
  45. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr, Ferguson, D. M. *et al.* (1995). A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **117**, 5179–5197.
  46. Cheng, X. & Schoenborn, B. P. (1990). Hydration in protein crystals. A neutron diffraction analysis of carbonmonoxymyoglobin. *Acta Crystallog., sect. B*, **46**, 195–208.



47. Steinbach, P. J. & Brooks, B. R. (1993). Protein hydration elucidated by molecular dynamics simulations. *Proc. Natl Acad. Sci. USA*, **90**, 9135–9139.
48. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
49. Bondi, A. (1964). van der Waals volumes and radii. *J. Phys. Chem.* **64**, 441–451.
50. Nicholls, A. & Honig, B. (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *J. Comp. Chem.* **12**, 435–445.
51. Onufriev, A., Bashford, D. & Case, D. A. (2000). Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B*, **104**, 3712–3720.
52. Tsui, V. & Case, D. A. (2000). Molecular dynamics simulations of nucleic acids using a generalized Born solvation model. *J. Am. Chem. Soc.* **122**, 2489–2498.
53. Srinivasan, J., Trevathan, M. W., Beroza, P. & Case, D. A. (1999). Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Accts.* **101**, 426–434.
54. Eliezer, D., Jennings, P. A., Wright, P. E., Doniach, S., Hodgson, K. O. & Tsuruta, H. (1995). The radius of gyration of an apomyoglobin folding intermediate. *Science*, **270**, 487–488.
55. Huang, X. & Powers, R. (2001). Validity of using the radius of gyration as a restraint in NMR protein structure determination. *J. Am. Chem. Soc.* **213**, 3834–3835.
56. Tanford, C. (1968). Protein denaturation. *Advan. Protein Chem.* **23**, 121–282.

*Edited by M. Levitt*

(Received 10 April 2002; received in revised form 15 October 2002; accepted 17 October 2002)