PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Molecular dynamics simulation of the acidic compact state of apomyoglobin from yellowfin tuna

Ettore Bismuto,[1,2*] Emiddio Di Maggio,[1] Stefan Pleus,[3] Martin Sikor,[4] Carlheinz Röcker,[3] G. Ulrich Nienhaus,[3,5] and Don C. Lamb[4,5]

[1] Dipartimento di Biochimica e Biofisica, Seconda Università di Napoli, Napoli, Italy

[2] Istituto Nazionale Biostrutture e Biosistemi, Roma, Italy

[3] Institute of Biophysics, University of Ulm, D-89069 Ulm, Germany

[4] Department for Chemistry and Biochemistry and Center for Nanoscience (CeNS), Ludwig-Maximilians-Universität München, D-81377 Munich, Germany

[5] Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

ABSTRACT

A molecular model of the acidic compact state of apomyoglobin (A-state) from yellowfin tuna was obtained using molecular dynamics simulations (MD) by calculating multiple trajectories. To cause partial unfolding within a reasonable amount of CPU time, both an acidic environment (pH 3 and 0.15M NaCl) and a temperature jump to 500 K were needed. Twenty-five acidic structures of apomyoglobin were generated by MD, 10 of them can be clustered by RMSD in an average structure having a common hydrophobic core as was reported for acidic sperm whale apomyoglobin, with shortened helices A,G,E, and H (the helix A appears to be translated along the sequence). Prolonging the MD runs at 500 K did not cause further substantial unfolding, suggesting that the ensemble of generated structures is indicative of a region of the conformational space accessible to the apoprotein at acidic pH corresponding to a local energy minimum. The comparison of experimentally determined values of specific spectroscopic properties of the apomyoglobin in acidic salt conditions with the expected ones on the basis of the MD generated structures shows a reasonable agreement considering the characteristic uncertainties of both experimental and simulation techniques. We used frequency domain fluorometry, acrylamide fluorescence quenching, and fluorescence correlation spectroscopy together with far UV circular dichroism to estimate the helical content, the Stern–Volmer quenching constant and the radius of gyration of the protein. Tuna apomyoglobin is a single tryptophan protein and thus, interpretation of its intrinsic fluorescence is simpler than for other proteins. The high sensitivity of the applied fluorescence techniques enabled experiments to be performed under very dilute conditions, that is, at concentrations of subnanomolar for the FCS measurements and 6 μM for the other fluorescence measurements. As high concentrations of proteins can strongly affect the association equilibrium among partially unfolded states, fluorescence techniques can provide complementary information with respect to other techniques requiring higher sample concentrations, such as NMR. The analysis of exposed hydrophobic regions in each of the MD-generated acidic structures reveals potential candidates involved in the aggregation processes of apomyoglobin in the acidic compact state. Our investigation represents an effective model system for studying amyloid fibril formation found in important diseases that are believed to proceed via aggregation of protein in the molten globule state.

Proteins 2009; 74:273–290.

Key words: molecular unfolding simulation; frequency domain fluorometry; acrylamide fluorescence quenching; fluorescence correlation spectroscopy; acidic A-state; molten globule state.

INTRODUCTION

Much of the initial interest in non-native protein conformations was motivated by the aim to achieve a detailed understanding of the process of protein fold-

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: Regione Campania L-5-2005; Grant sponsor: Seconda Università di Napoli RA-2007; Grant sponsor: Deutsche Forschungsgemeinschaft; Grant number: NI 291/3; Grant sponsor: DFG; Grant number: SFB749; Grant sponsor: Fonds der Chemischen Industrie; Grant sponsor: LMUInnovative BioImaging Network (BIN); Grant sponsor: German Excellence Initiative "Nanosystems Initiative Munich (NIM)."
*Correspondence to: Ettore Bismuto, Department of Biochemistry and Biophysics of Second University of Naples, Via de Crecchio 7, 80138 Naples, Italy.
E-mail: ettore.bismuto@unina2.it
Received 23 July 2007; Revised 28 April 2008; Accepted 19 May 2008
Published online 10 July 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22149

ing.[1–7] Nowadays, it is increasingly recognized that non-native, partially folded or denatured states of proteins can yield important insights into a range of other issues, including protein translocation, misfolding, and aggregation.[8,9] Indeed, it has been proposed that partially folded states of proteins with exposed hydrophobic surfaces can be precursors for amyloid fibril formation. Various proteins, apparently unrelated to any human disease, form fibrils *in vitro* that are indistinguishable from fibrils of patients affected by sporadic and familial amyloidosis. For example, myoglobin, an α-globular protein that is well known as a model system for protein dynamics,[10,11] forms fibrils containing β-strands under experimental conditions that favor the formation of partially folded states.[12] Acid denaturation of apomyoglobin, as in the case of numerous other proteins, causes an accumulation of a partially folded state called the "acidic compact" or "A-state" under appropriate pH and ionic strength conditions. The A-state evidences several features of the "molten globule" state: (1) substantial secondary structure, often similar to what is present in the native state; (2) absence of much of the native-like tertiary structure; (3) a hydrodynamic radius closer to the native state than that of the unfolded state; (4) a hydrophobic core; and (5) the propensity to aggregate because of a substantial exposure of hydrophobic surfaces.[2,13,14]

The properties of compact intermediates from different proteins vary, and even compact intermediates from the same protein may differ, depending on the experimental conditions.[15] Compact intermediates consist of either a relatively native state-like topology with native state-like regions of secondary structure connected by disordered regions of the polypeptide chain, or a core of native-like structure, surrounded by totally or partly unfolded polypeptide chain. The latter is the case for acidic apomyoglobin, which has a native-like core formed by a subset of the A, G, and H-helices.[16–18]

The goal of this work was to characterize the acidic compact state of apomyoglobin from yellowfin tuna in molecular detail as a model of the "molten globule" state of the protein. To this end, an ensemble of non-native structures of apomyoglobin was generated by molecular dynamics (MD) simulation, which was considered as a representative sample of the distribution of conformers constituting the acidic compact state. For our MD investigations, myoglobin from yellowfin tuna offered particular advantages over the better studied myoglobin from the sperm whale: tuna myoglobin has a single tryptophan, which simplifies the interpretation of the intrinsic fluorescence emission and its decay rate,[19–21] and it has a lower unfolding free energy when compared with sperm whale myoglobin as deduced from guanidinium chloride and/or acidic unfolding studies[22] (9.2 and 12.5 kJ mol^{-1}, respectively). Therefore, a smaller computational effort is necessary to obtain MD unfolding trajectories.

MD simulations are a powerful tool to explore the dynamic properties of partially folded states of proteins.[23] The all-atom simulation of protein folding by realistic models, such as GROMACS,[24] AMBER,[25] or CHARMM[26] is still computationally out of reach. The major problem with MD simulations is the conformational sampling efficiency as the number of possible conformations is astronomically large.[27] Three different techniques have been proposed to overcome, at least partially, this limit: the use of elevated temperatures in the simulations, the addition of extra terms in the MD forcefield and changes to the initial protein-solvent system.[28] Another difficulty in MD simulations of protein unfolding is that, often, only one pathway is studied at a time. In contrast, classical denaturation experiments performed on bulk samples monitor a large ensemble of molecules passing from the native to unfolded states. Therefore, it is necessary to perform multiple unfolding simulations to compare the results of the simulations with experiments.[29] We note that protein folding experiments have also been performed at the single-molecule level,[30,31] but many molecules and folding trajectories need to be observed to capture the statistical aspects of the process.

In this article, we have explored the conformational space of tuna apomyoglobin, using MD simulations at higher temperature (500 K) and acidic conditions (pH 3.0) in the presence of 0.15*M* NaCl. Under these conditions, apomyoglobin is known to exist mostly in the acidic compact state. The high temperature MD simulation was followed by short, re-equilibrium steps at intermediate temperatures until 300 K was reached. The overall procedure was sufficient to partially disorganize the native structure. Protonation of the amino acid residues at pH 3 was assessed by structure-based theoretical methods,[32–34] which predict the pK_a of ionizable groups in macromolecules and are very sensitive to details of the local protein conformation. Twenty-five different MD simulations were performed to generate 25 acidic structures for apomyoglobin. The analysis of the surfaces of the MD-generated acidic protein structures provides evidence of hydrophobic regions that are potential sites for oligomerization of apomyoglobin molecules in the A-state.[10,35]

To test the results of the MD simulations, frequency domain fluorometry,[36] fluorescence correlation spectroscopy (FCS),[37–40] and acrylamide fluorescence quenching[41,42] experiments together with far UV circular dichroism were performed. These measurements provide experimental values that can be compared with those determined from the acidic apomyoglobin structures obtained by MD simulations. The spectroscopic properties obtained by fluorescence and far UV circular dichroism are shown to be in reasonable agreement with the structure-based values.

RESULTS

MD simulations

To generate an ensemble of non-native structures at pH 3.0 and in the presence of 0.15*M* NaCl, a preliminary

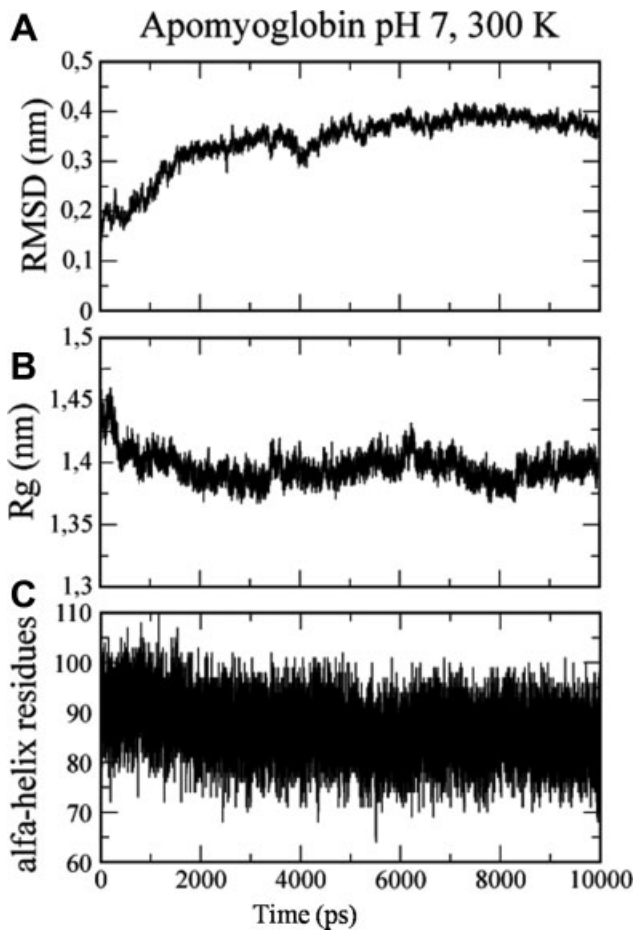

Figure 1

Molecular dynamics simulations of apomyoglobin in water and 0.15*M* NaCl at pH 7. Starting coordinates were taken from the Protein Data Bank file 1MYT. **A**: Root-mean-square deviation of all atoms as a function of time showing a structural drift from starting coordinates during the first 2 ns. **B**: The calculated radius of gyration. **C**: The molar ellipticity at 222 nm calculated along the whole trajectory and expressed as number of helix residues. The temperature during simulation was kept at 300 K.

MD simulation of neutral apomyoglobin was necessary. The apomyoglobin structure was obtained by starting from the myoglobin crystallographic structure (PDB entry: 1MYT),[43] removing the coordinates of the prosthetic heme group and allowing the protein to equilibrate for 10 ns at neutral pH in the presence of 150 m*M* NaCl. The root-mean-square standard deviations (RMSD) of the apomyoglobin atoms with respect to the crystal structure were calculated by a least-square fitting of the atom positions to the crystallographic X-ray structure and subsequently calculating the RMSD.[44] As can be seen in Figure 1(A), after the first nanosecond, the overall deviation from the initial structure of myoglobin is less than 0.45 nm, indicating that the structure is already fluctuating around a stable conformation. Figure 1(B) reports the radius of gyration calculated using:

$$R_g = (\Sigma |r_i|^2 m_i / \Sigma m_i)^{1/2}. \quad (1)$$

The output structure of the 10 ns MD simulation of apomyoglobin at neutral pH and 150 m*M* NaCl was considered as the starting structure of the acidic unfolding using MD simulations. Figure 1(C) shows the time dependence of the expected molar ellipticity at 222 nm expressed in terms of the average number of residues with an α-helical conformation in apomyoglobin at neutral pH as calculated from the MD trajectories according to Hirst and Brooks.[45] The major difference between our results and a previously MD study of apomyoglobin from sperm whale (performed with shorter MD runs)[46] occurs in the sequence region from residues 45 to 50 (helix D), which appears to be unfolded as in the crystal structure of tuna myoglobin.[43] This observation confirms the idea that the lacking or instable D helix in myoglobin favors the intake of O_2 molecules at the heme pocket where the O_2 binds.[47] The first step was the protonation of specific residues as expected at pH 3 and at 150 m*M* NaCl

The addition of protons was performed on the basis of p*K*a values as estimated by the freely available webserver http://biophysics.cs.vt.edu/H++[33] that generates the properly protonated protein starting from the three-dimensional structure. The H++ package is based on standard continuum solvent methodology using the generalized Born or Poisson–Boltzmann (PB) models and was originally established by Basford and Karplus[32]; we used the more accurate PB procedure. The residues that are additionally protonated at pH 3.0 were: the N- and C-termini, D2, D49, H88, K108, and E136. It is not possible to follow the acidic unfolding by MD simulations in a reasonable amount of CPU time. As shown in Figure 2, a RMSD change smaller than 0.25 nm during the MD run. More specifically, the apomyoglobin structure changes only slightly during the first 2.0 ns of the MD simulation, specifically the F and G-helices make a slight rotation (Fig. 2, inset).

Therefore, we explored a more efficient strategy that involves a temperature jump of the protein from 300 to 500 K followed by successive re-equilibration steps at intermediate temperatures before returning to 300 K. Specifically, the protonated structure of acidic apomyoglobin was instantaneously heated by generating random velocities for the protein atoms and solvent molecules that follow the Maxwell–Boltzmann distribution at 500 K. A 1.0 ns MD simulation run was then performed. This "heat-driven" procedure has been carried out 25 times using the same starting structure, generating an ensemble of 25 conformations for the acidic compact state of apomyoglobin. A MD run of 1.0 ns for the heating step was considered long enough since the calculated molar ellip-

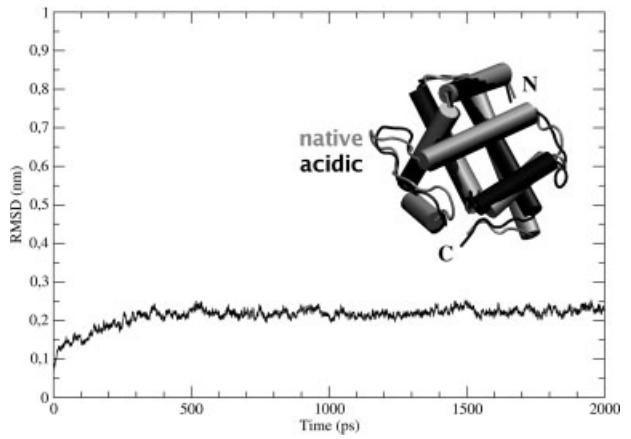

Figure 2

Molecular dynamics simulation of apomyoglobin in water and 0.15*M* NaCl at pH 3. Starting coordinates were taken from the output coordinates of the 10 ns MD (Fig. 1); temperature was kept constant at 300 K. Structural drift from starting coordinates is shown as root-mean-square deviation of all atoms. The insert depicts the superposition of the output apomyoglobin structure at pH 7 (grey) and pH 3 (black).

ticity value at 222 nm was reduced by an average of ~50%, indicating a residual core structure even though large segments of the protein were unfolded and corresponds to what was experimentally observed for the acidic compact of apomyoglobin. The temperature of each acidic unfolded structure of apomyoglobin was re-equilibrated at 300 K by performing three further 200 ps long MD steps at the intermediate temperatures of 450, 400, 350, and 300 K, respectively. A final 1.0 ns MD run was then made at 300 K (data not shown). Prior to performing each step of the MD simulations, the protonation state of apomyoglobin residues was adjusted on the basis of p*K*a values estimated by H^{++} package considering the specific unfolded three-dimensional structure. Table I reports the predicted pK_a values of the amino acidic residues that have a different protonation state among the 25 generated acidic-compact and the native apomyoglobin structures. In particular, D117, D134, E34, and H77 are mostly protonated, while E31, H59, H88, and H92 show high variability. Figure 3 illustrates a typical MD run with the described T-jumps. Figure 3(A) shows the time course of the RMSD: during the first nanosecond, in which the "heating" takes place, the RMSD increases continuously. The subsequent steps consist of MD runs of 200 ps each at 450, 400, 350, and 300 K successively. A small decrease in RMSD occurs indicating the rearrangement of the acidic structure at each specified temperature with smooth oscillations around more stable conformations. The "heating" process, as expected, occurs with a loss of ordered structure as evidenced by the course of the molar ellipticity at 222 nm as well as

Table I

Calculated pK_a's and Consequently Assigned Protonation States of Titratable Residues Along the 25 Apomyoglobin Acidic Structures After a 1 ns MD Simulation at 500 K

	D4	E14	E31	E34	H59	H77	H88	H92	E112	D117	D134
01	3,071	2,178	3,946	−0,47	6,371	5,345	5,779	0,701	−0,04	−5,12	0,791
02	0,698	2,382	3,329	−2,02	3,853	1,593	0,035	2,726	2,02	2,33	17,57
03	−2,1	−1,07	1,999	−2,3	3,772	7,846	−5,01	1,088	1,22	1,763	0,874
04	2,237	1,819	0,838	0,353	5,34	6,769	5,431	1,005	2,887	1,934	1,909
05	1,746	4,16	2,638	−6,84	1,026	6,737	−0	5,575	3,566	−0,17	2,633
06	3,816	1,296	1,598	−1,52	4,69	12,69	−4,05	0,822	0,566	0,109	−1,39
07	3,669	3,682	3,55	2,298	1,676	7,279	4,505	−5,15	3,274	−1,95	0,191
08	−2,65	2,795	0,656	3,662	3,505	18,01	1,423	3,564	3,167	−5,59	1,575
09	3,327	1,126	1,143	3,598	3,783	11,29	6,181	3,282	0,351	3,671	−3,98
10	1,919	2,546	3,287	−1,82	−1,49	6,389	6,515	5,113	1,357	−2,71	0,805
11	6,414	1,76	3,854	−0,28	1,959	10,78	4,069	3,763	−0,06	0,572	−1,57
12	1,618	3,527	3,999	1,977	4,97	16,15	5,635	5,711	2,611	1,561	1,41
13	3,246	2,619	4,099	−1,31	4,578	6,752	0,547	−1,36	1,026	−1,31	2,749
14	2,528	1,242	1,24	2,847	−2,21	6,107	4,401	4,756	1,837	−9,49	2,256
15	3,086	2,95	4,833	0,914	6,101	6,806	7,068	4,94	2,321	3,802	1,772
16	2,284	3,602	4,11	0,312	5,626	11,37	3,704	6,635	−1,48	−0,39	−0,4
17	1,199	1,216	3,77	−1,84	−0,97	6,89	5,842	2,04	3,935	2,1	0,809
18	−1,18	4,328	4,096	−3,08	2,901	15,91	−3,34	1,906	2,723	1,248	−0,16
19	3,098	3,738	3,092	−0,79	1,866	7,033	4,014	5,277	1,876	−0,11	1,751
20	−1,6	3,349	0,803	1,815	5,869	6,099	6,444	−1,11	1,35	2,516	3,621
21	−1,32	0,818	5,105	0,058	5,464	10,83	2,496	3,177	2,105	1,694	3,07
22	4,412	4,51	3,399	0,345	4,822	8,897	0,583	2,565	0,849	0,134	−9,3
23	2,762	1,747	0,963	1,464	−3,55	6,882	−0,33	0,408	0,544	3,581	−10,3
24	4,169	3,991	3,823	−0,27	8,042	16,75	4,86	4,058	1,256	1,446	1,932
25	−2,89	3,978	4,493	2,265	1,427	6,556	4,895	4,854	2,89	2,122	1,364

Sites which carry one net positive charge have a dark grey background, while sites carrying one net negative charge have a light grey background. Only the sites which show variability are represented.

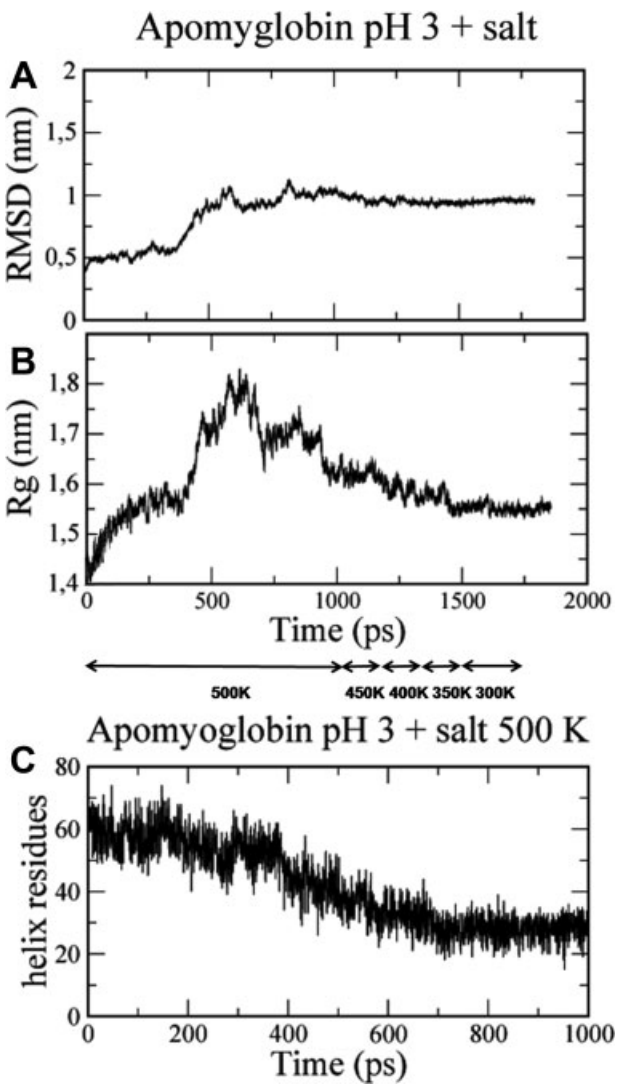

Figure 3

Molecular dynamics simulations of apomyoglobin in water and 0.15*M* NaCl at pH 3. The starting coordinates were taken from the output of the 10 ns MD on apomyoglobin at neutral pH (Fig. 1) and a temperature jump to 500 K was performed. The temperature was kept constant at 500 K during the first nanosecond and then returned stepwise back to 300 K with steps of 0.2 ns at the intermediate temperatures of 450, 400, and 350 K as indicated by the below arrows. After returning to 300 K, the simulation was run for another nanosecond. Only 0.2 ns of the trajectory at 300 K is shown in the figure. The dependence relative to the other 0.8 ns of the trajectory remains quite the same. **A**: Shows the root-mean-square deviation of all atoms signifying the structural drift from starting coordinates. **B**: Shows the calculated radius of gyration as a function of time. **C**: Shows molar ellipticity at 222 nm as number of helix residues in the structure calculated along the first ns of trajectory during the simulation at 500 K. The simulation shown here corresponds to structure 1 in Figure 4.

the radius of gyration as shown in Figure 3(B,C). The gradual annealing of the structure down to 300 K causes a partial decrease of these values suggesting a small contraction of the acidic structure. The MD simulation is ended with a further MD run of 1.0 ns at 300 K, in which further substantial changes are not observed (data not shown).

Figure 4 shows the 25 MD-generated three-dimensional structures of apomyoglobin at pH 3 and 0.15*M* NaCl. Each color represents a segment of the amino acid chain that was formerly organized in one of the α-helices of the native protein as indicated in the lower portion of the figure. The α-helix content of the acidic structures ranges from 18 to 47% with an average value of 33%. Short β-sheet segments appear in 10 of the acidic structures. To investigate whether the ensemble of MD generated acidic structures represent a reasonable representation of the acidic compact state of apomyoglobin, we calculated some predictable properties on the basis of the individual A-state structures and compared the average values with those determined experimentally. Specifically, we have calculated the secondary structure content by dssp algorithm[48] with emphasis on α-helical content that is better determined by CD analysis predictive methods,[49] the acrylamide accessibility to the single tryptophanyl residue and the radius of gyration of tuna apomyoglobin, using the average structure from the last 300 ps of each trajectory at 300 K (Table II). The collisional fluorescence quenching rate for each acidic structure was determined by considering the ratio, σ_i, of the corresponding acrylamide accessibility to tryptophan residue (reported in the second column of the Table II) to the accessibility of the acrylamide to fluorophore free in solution (*N*-acetyl tryptophanylamide, NATA). Acrylamide was treated as a sphere of radius 1.7 Å as estimated from the energy minimized molecular model. The NATA accessibility of 2.7 nm^2 was calculated from a 1 ns MD trajectory of NATA in water at 300 K (data not shown). Individual values for each acidic structure of apomyoglobin was then obtained by multiplying the ratio σ_i with k_q relative to NATA in solution (7.1×10^9 L M^{-1} s^{-1}).[50]

From the calculated values of k_q for each MD generated structure, the Stern–Volmer dependence expected for an equilmolar mixture of all the MD generated acidic structures can be determined using the expression[50]:

$$F/F_0 = \sum_i 1/\{(1 + \sigma_i k_q \tau[Q])\exp(V_q[Q])\}, \qquad (2)$$

where, F and F_0 are the fluorescence intensity for a given acrylamide concentration and in the absence of acrylamide, respectively, V_q is the volume of the sphere of effective quenching and τ is the tryptophan lifetime in the absence of acrylamide quencher. Figure 5 shows the calculated and experimentally determined plot of acrylamide fluorescence quenching of apomyoglobin at pH 3 in the presence of salt. The results of a non-linear regression to both the experimental and structure-based data are reported in Table III. The K_{SV} values are rather similar. Table IV summarizes the comparison of the average expected value from the MD simulations for the α-helical content, collisional quenching constants, and radius of gyration along with the experimentally determined values.

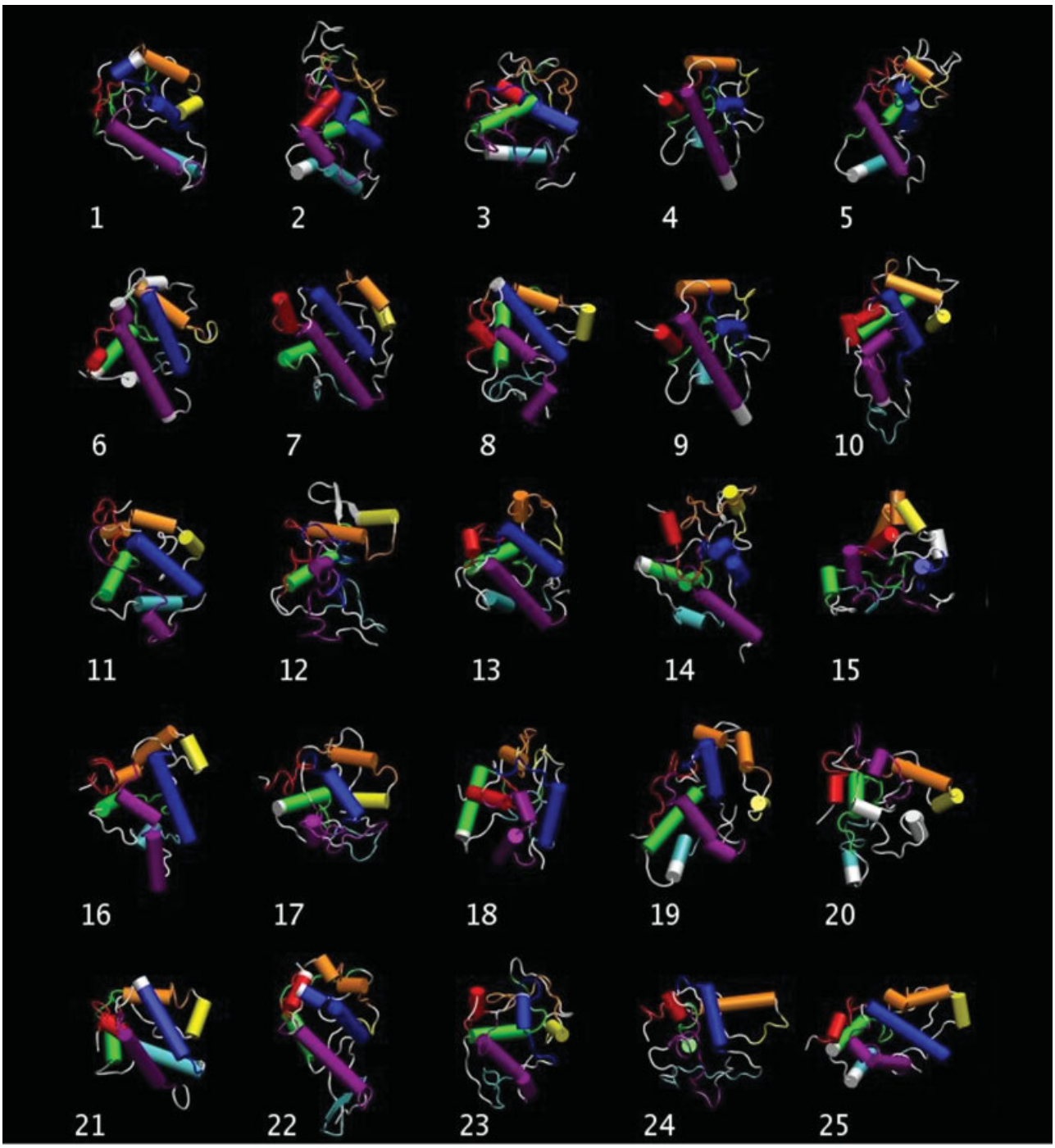

Figure 4

Output structures of the 25 apomyoglobin molecular dynamics simulations at pH 3, represented in "cylinder style." The former α-helix segments of the native structure, as indicated in the lower part of the figure along with the amino acid sequence, are depicted in different colors.

Table II
Structural and Spectroscopic Properties Derived from the MD Simulations of Native Apomyoglobin (*N*) and 25 Acidic Forms

Structure	R_g (nm)	Acrylamide accessibility to tryptophan (nm^2)	%α-Helix residues
Native	1.40	0.39	67
1	1.55	1.36	26
2	1.62	0.7	33
3	1.56	0.4	27
4	1.56	0.57	24
5	1.65	1.24	35
6	1.56	0.23	40
7	1.56	0.84	39
8	1.6	1.98	40
9	1.62	1.74	39
10	1.76	0.37	24
11	1.58	0.88	32
12	1.68	0.9	30
13	1.63	0.78	27
14	1.6	0.37	29
15	1.56	0.38	32
16	1.62	0.43	45
17	1.5	0.38	32
18	1.53	0.37	32
19	1.66	1.29	38
20	1.57	0.63	36
21	1.55	0.28	32
22	1.65	1.11	34
23	1.6	0.65	34
24	1.64	1.7	35
25	1.61	0.48	45
Average and St. Dev.	1.60 ± 0.05	0.80 ± 0.49	33 ± 6
Standard error[a]	0.01	0.098	1.2

The given values are the mean quantities calculated from the last 300 ps of the isothermal 300 K trajectory.

Ten of the 25 acidic structures can be grouped together into a single cluster with a RMSD of less than 0.8 nm as shown in Figure 6. The average structure of this cluster, shown in the inset of Figure 6, displays the more conserved structural regions in the acidic compact state with shorter helices A,G,E, and H (the helix indicated by as "A" appears to be translated along the sequence with respect to the position occurring in the native apomyoglobin). To investigate the stability of the acidic structures generated by our MD trajectories, the heating step at 500 K for each of 25 non-native structures was continued for an additional 1 ns followed by a re-equilibration run at 300 K. The RMSD remained almost constant during the MD simulations and no further significant unfolding was observed. This is illustrated in Table V, which compares the average lengths of helical segments calculated for the 25 structures before and after the additional 1.0 ns MD run at 500 K.

Aggregation of acidic apomyoglobin is attributed to interactions among solvent-exposed hydrophobic segments of the amino acid chain. To speculate over which regions of the protein and which acidic structures could be potential candidates for nucleation sites of the aggregation process, we have explored the solvent accessibility surface of the MD generated acidic structures. Specifically, we have attributed an index of hydrophobicity to the surrounding of the main hydrophobic residues as identified by the Kyte–Doolittle's plot[51] for each acidic structure. The results are summarized in Table VI. Each

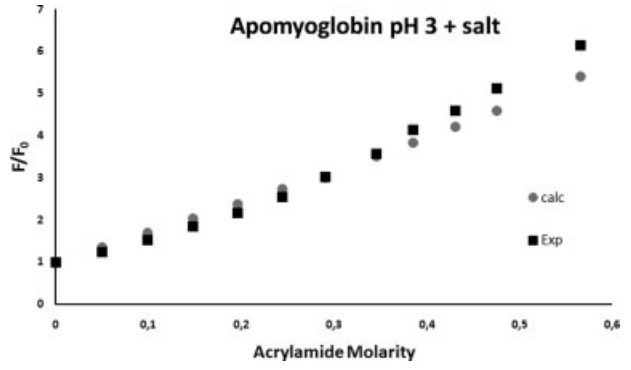

Figure 5
Simulated acrylamide fluorescence quenching measurements using an equimolar mixture of the 25 MD obtained acidic structures of apomyoglobin at 27°C. The results of acrylamide fluorescence quenching experiments at pH 3 are shown for comparison.

Table III
Comparison between the Non-Linear Regression Analysis of the Experimental and MD Calculated Data of Figure 5 for Acrylamide Fluorescence Quenching of Apomyoglobin at pH 3 and 0.15*M* NaCl

K_{SV} (M^{-1})	V_q (M^{-1})	Correlation coeff.
Experimental		
4.9 ± 0.32	1.1 ± 0.11	0.990
Calculated		
4.2 ± 1.0	0.65 ± 0.35	0.980

Table IV
Comparison of Structure-Based Properties for the Mean from the 25 MD Generated Acidic Structures at pH 3 and for Native Apomyoglobin at pH 7 with Values Determined Experimentally

	R_g[a] (nm)	Collisional quenching constant (M^{-1} ns^{-1})	%α-Helix residues
Calculated pH 3	1.6	1.5	33
Experimental pH 3	2.5	1.3	35
Calculated pH 7	1.4	0.6	67
Experimental pH 7	1.8	0.9	63

[a]The measured hydrodynamic radius, from which the radius of gyration was approximated, includes the hydration shell about the protein. The hydration shell was not included in the calculation of the radius of gyration from the MD simulations.

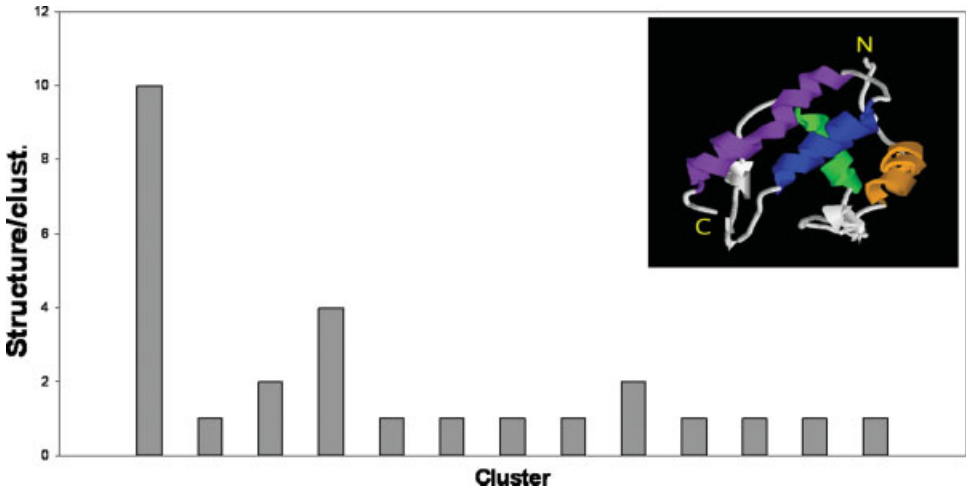

Figure 6

Cluster analysis of the acidic apomyoglobin ensemble. Structures have been clustered by RMSD; two structures have been considered as belonging to the same cluster if their displacements differ by less than 0.8 nm. The average structure is shown in the inset. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

value in the table has been calculated according to the formula:

$$\text{hydrophobicity} = \Sigma h_{KD}{}^{*}\text{SAS}_r/\text{SAS}_a \qquad (3)$$

where the sum is extended to the residues within a sphere of radius 0.8 nm centered at the residue indicated in the title row. For each residue included in the sphere, h_{KD} is its Kyter–Doolittle's hydropathy index, SAS_r is the solvent accessible surface of the residue, and SAS_a is the solvent accessible surface of that residue if completely exposed to solvent. To determine the SAS_a, a 1 ns MD simulation for each type of residue has been run, in which the single residue X was inserted into a 3-peptide chain of general sequence GXG (see supplementary material). The larger and more positive the value shown in the column of Table VI, the higher is the hydrophobicity in that specific region of the corresponding acidic structure. The last row of Table VI shows the average values indicating that the larger hydrophobicity are in average those surrounding leucine 25 and Alanine 54.

Table V
Average Lengths of Helical Segments through 1 and 2 ns MD of Acidic Apomyoglobin at 500 K

Helical segment	1 ns MD (T = 500 K)	2 ns MD (T = 500 K)
A	—	—
B	17–29	17–31
C	33–35	33–35
D	—	—
E	61–68	62–70
F	—	—
G	98–110	99–109
H	120–138	120–136

Experimental spectroscopic measurements

Experimental techniques, such as circular dichroism, steady-state, and time resolved fluorescence, and FCS have been utilized to provide observable quantities to

Table VI
Hydrophobicity of Selected Sites for the 25 Acidic Apomyoglobin Structures

Structure	L25	A54	P83	V107	L135
1	−0.234	2.9842	1.5813	−0.399	0.0093
2	−1.998	−2.086	−2.319	4.2648	3.2892
3	3.5514	2.8137	−0.51	2.4724	−0.53
4	1.8161	1.1409	0.4132	−5.483	2.636
5	0.4749	5.3323	−3.427	−4.021	−2.252
6	0.205	3.3457	−0.555	−6.973	−1.668
7	3.6373	0.7102	−0.676	−5.508	−2.779
8	1.5443	5.4212	−0.728	−7.187	0.4188
9	4.437	6.2934	−5.376	−3.549	0.5366
10	−1.299	4.4861	−3.241	−0.431	−0.339
11	4.9011	5.3132	−0.484	−4.67	−0.455
12	2.9432	1.496	0.8419	−0.618	−4.954
13	1.7601	−4.985	2.0863	−6.89	2.202
14	3.1088	2.661	−3.586	1.8483	−5.849
15	−0.813	−1.13	−4.791	−8.002	0.9338
16	3.356	−1.477	−1.477	0.4919	2.6598
17	0.3976	0.3976	−3.152	−3.152	3.8749
18	6.8936	5.7444	−0.733	2.5777	0.1021
19	4.8958	1.9587	−3.895	0.2284	−0.416
20	−0.713	−3.862	0.9156	2.8584	−3.571
21	−3.248	5.6927	−2.989	−4.682	1.2333
22	−1.525	1.8274	−0.771	−1.767	0.3491
23	2.228	1.1708	3.5462	2.8343	1.4626
24	4.0709	2.5264	0.6105	−2.08	3.6829
25	2.6899	0.1251	−5.125	−3.172	0.6771

Hydrophobicities are calculated as sum of the residues within 0.8 nm to the residues indicated in columns. See text for a more detailed discussion.

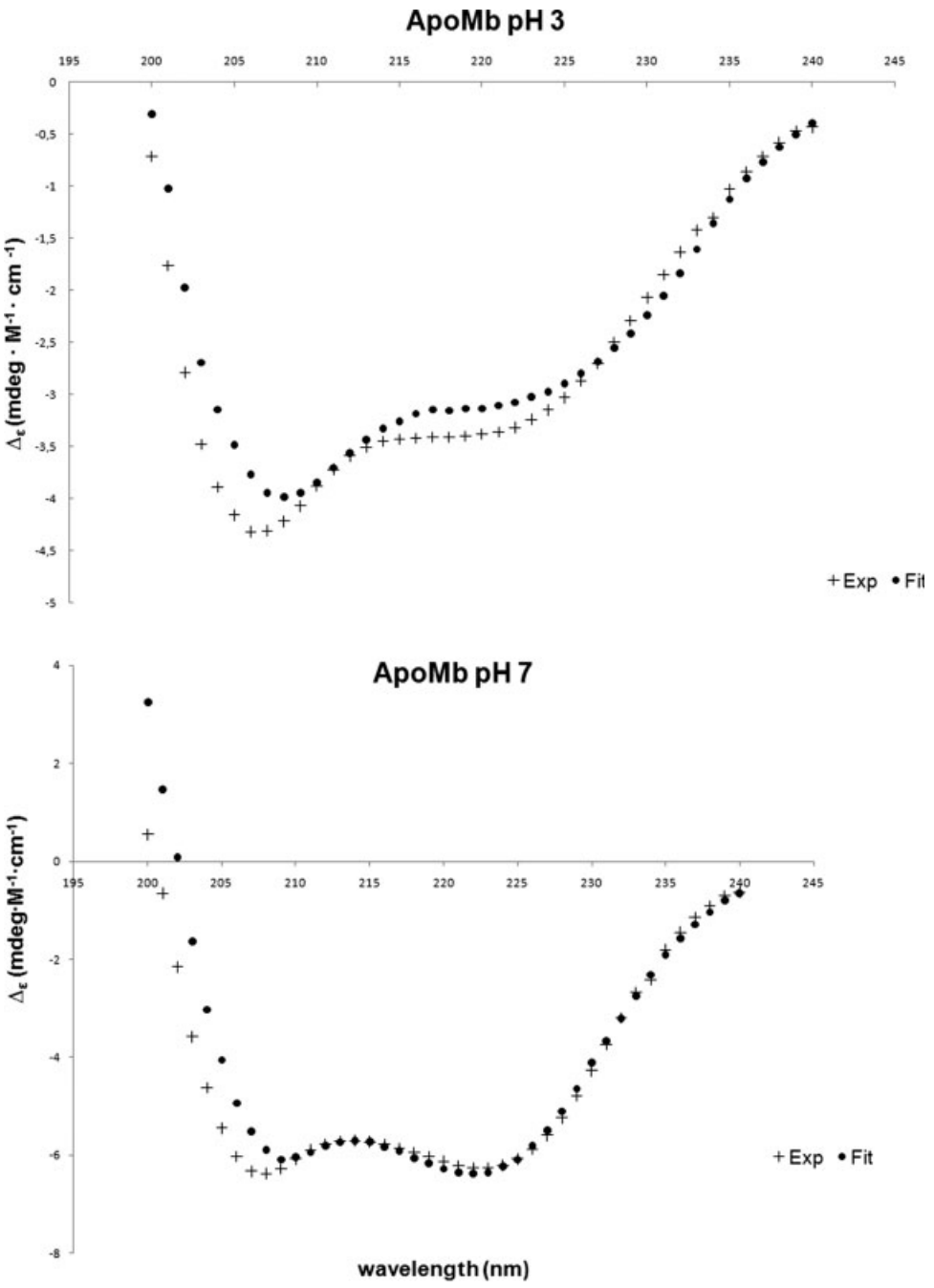

Figure 7

Molar ellipticity per residue of apomyoglobin in 0.01*M* phosphate, 0.15*M* NaCl, at pH 3 (upper panel), and pH 7 (lower panel). Solid and dash lines refer to experimental and fitted data respectively. Measurements were taken at 27°C.

compare with those expected on the basis of the partial unfolded structures of acidic apomyoglobin (pH 3.0, 0.15*M* NaCl) generated with MD simulations. Figure 7 shows the far-UV CD spectra of apomyoglobin in the presence of 150 m*M* NaCl both at neutral pH and at pH 3.0. The pronounced minima of the molar ellipticity at 222 nm essentially comes from the α-helical content of the protein. The analysis of CD data by the program k2d, based on a neural-network algorithm[52] and available on the webserver DICHROWEB (http://www.cryst.bbk.ac.uk/cdweb),[53] evidences a content of 63 and 35% of α-helical content for neutral and acidic apomyoglobin, respectively; the unordered structure was 37 and 45% for native and acidic apomyoglobin. Moreover,

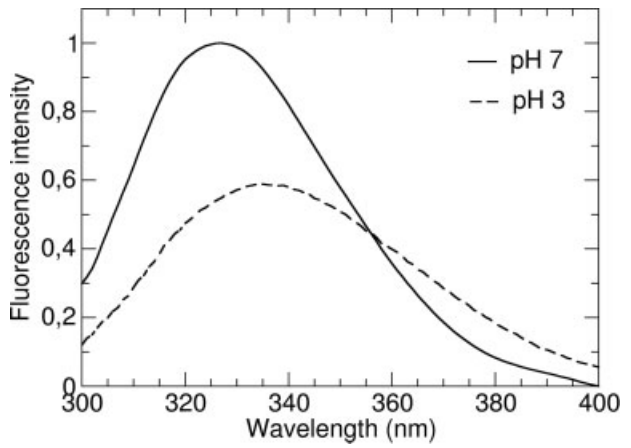

Figure 8

Fluorescence emission spectra with 290 nm excitation of apomyoglobin in 0.01*M* phosphate, 0.15*M* NaCl, at pH 7 (solid line) and pH 3 (dash line). Measurements were taken at 27°C.

20% β-structure was evidenced for the case of acidic apomyoglobin. Similar results were observed for the case of the CD spectrum of horse apomyoglobin at acidic pH.[13] The substantial amount of β-structure observed by circular dichoism corroborates the presence of short β-segments in 10 of the acidic structures obtained by MD simulations. However, the reliability of Andrade's algorithm[52] to find beta structure is questionable since the explored interval ranges from 200 to 240 nm because of the noise intrinsically present in the acidic solution of apomyoglobin. For example, Lees *et al.*[49] reports the case of jacalin, for which the β-sheet content was predicted to be 15% instead of 47%.

Figure 8, shows the single tryptophan fluorescence spectra of both native and acidic apomyoglobin. A marked shift in the maximum emission wavelength from 325 nm for the native form to 335 nm for the apomyoglobin in the A-state is observed, suggesting that the tryptophanyl residue is deeply buried in the native state and more exposed to the solvent in the acidic compact state. However, the emission maximum is still far from its value of 350 nm in the fully unfolded state.

Decay of the intrinsic fluorescence emission of a protein can be complex even in single-tryptophan proteins, such as tuna apomyoglobin.[21] The observed decay is not a single exponential and it is often found to be best fitted by a quasi continuous distribution of fluorescence lifetimes, indicative of the variety of different microenvironments explored by the tryptophanyl residue in its excited state.[21,54] A very powerful fluorescence technique for investigating complex emission decay rates is frequency domain fluorometry.[36] In these measurements, the continuous excitation light is modulated in intensity at a selected frequency. The emission is also modulated at the same frequency but phase-delayed and demodulated; both phase delay and demodulation are directly related to the fluorescence lifetime. The possibility to excite the protein sample with several different modulation frequencies allows resolution of complex decay behaviors. Figure 9, left panel shows the phase shift and demodulation data for apomyoglobin in 150 m*M* NaCl at neutral pH and at pH 3.0. The data were fitted with different models and the best fits were obtained by using Lorentzian distribution, shown on the right panel of the same figure, centered at 2.64 and 3.29 ns for the native and acidic apomyoglobin, respectively. The full-width at half maximum were 0.16 ns (native) and 1.38 ns (acidic). The marked widening of the distribution at pH 3 is indicative of a much larger variety of tryptophanyl microenvironments existing in the A-state when compared with the native state. This finding supports the idea that the acidic compact state is formed by an ensemble of partial unfolded structures. In both distributions shown in Figure 9, a small discrete component at ~200 ps was needed to correct for instrumental contributions.[55]

Acrylamide fluorescence quenching

Quenching of tryptophanyl fluorescence in proteins by addition of acrylamide is described by the classical Stern–Volmer plot showing the dependence of F_0/F ratio of fluorescence intensity over the fluorescence intensity in the absence of quencher as a function of acrylamide concentration.[41] The Stern–Volmer plot for tuna apomyoglobin in the presence of 50 m*M* sodium phosphate and 150 m*M* NaCl at neutral pH and pH 3 are shown in Figure 10. Both plots are nonlinear with a more pronounced upward curvature for acidic apomyoglobin. An upward curvature in a Stern–Volmer plot is indicative of simultaneous static and dynamic fluorescence quenching described by the relation[41]:

$$F_0/F = (1 + K_q\tau[Q])\exp(V_q[Q]), \qquad (4)$$

where, K_q is the collisional quenching constant, and V_q and τ is the tryptophan lifetime in the absence of acrylamide quencher are defined as in Eq. (2). The nonlinear fit of data by Eq. (4) gives K_q values of $(1.3 \pm 0.4)\cdot 10^9$ L M^{-1} s^{-1} and $(6.4 \pm 0.7)\cdot 10^8$ L M^{-1} s^{-1} for acidic and native apomyoglobin, respectively. The corresponding values of active V_q are 1.15 ± 0.1 M^{-1} and 1.52 ± 0.09 M^{-1} for acidic and neutral apomyoglobin, respectively. The K_q values are indicative of a larger accessibility to tryptophan of acrylamide molecules in the acidic compact state of tuna apomyoglobin compared with that of the native form, although less than the value of 7.1×10^9 L M^{-1} s^{-1} for measurements with NATA.[42] The values of the static quenching constants V_q also provide information about the exposure of tryptophan residues but are less reproducible.[41]

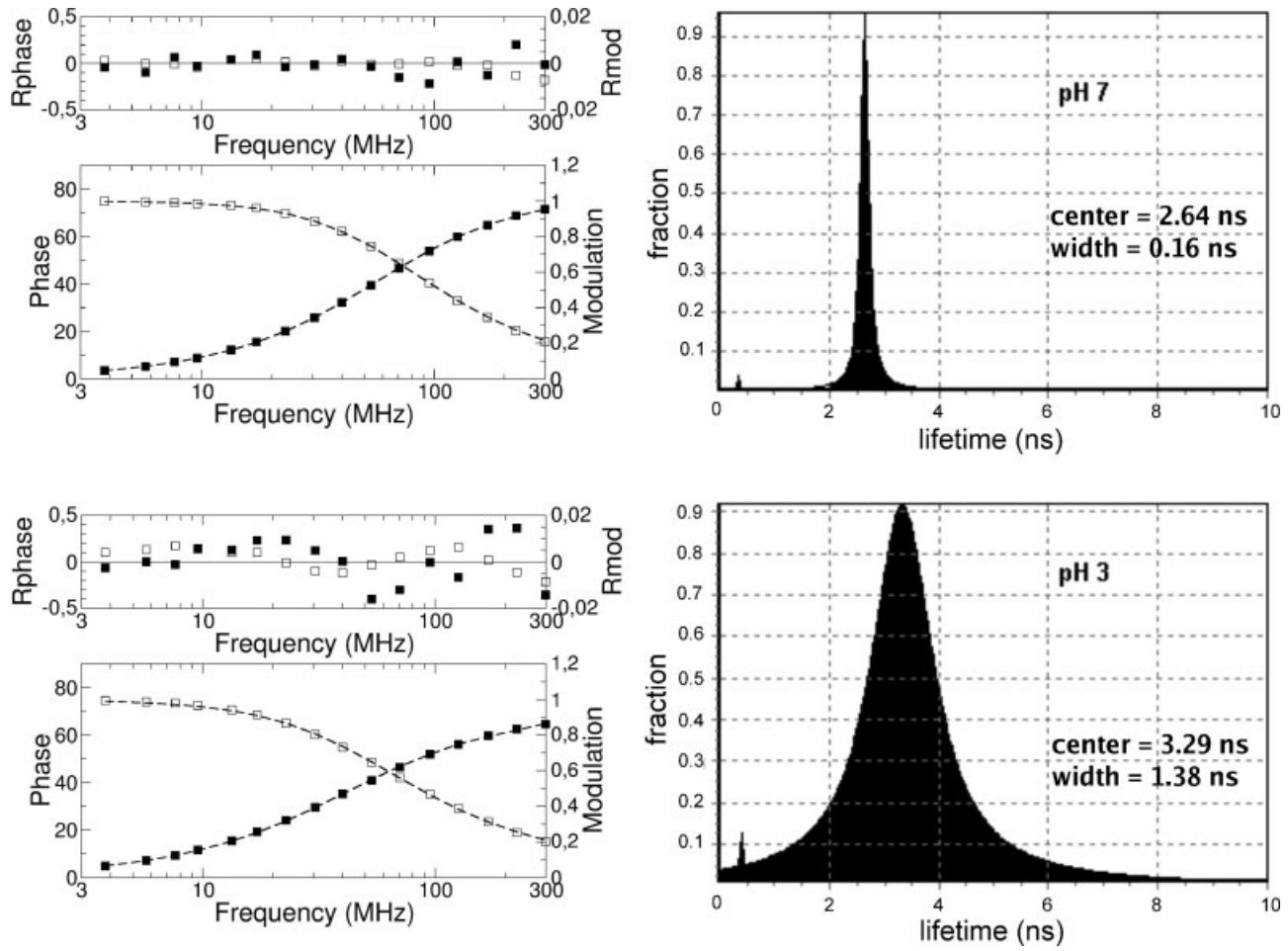

Figure 9

Triptophanyl fluorescence decay analysis at pH 7 (upper graphs) and pH 3 (lower graphs). The left side of the figure shows the frequency dependence of the phase shifts (solid symbols) and demodulation factors (open symbols) along with the respective residuals. On the right side, the Lorentzian shaped distribution functions obtained by fitting the emission decay data are plotted. The measurements were performed with 295 nm excitation in a solution containing 0.01*M* phosphate and 0.15*M* NaCl at 27°C.

FCS measurements

To compare the approximate radius of hydration in apoMb at neutral pH and in the acid compact state, FCS was applied. As FCS is typically performed at dilute concentrations (subnanomolar concentrations), it is very advantageous for measurements with denatured proteins as aggregation is not a significant problem at these concentrations. In FCS, the fluorescence coming from the small (<1 fL) focus of a confocal microscope (probe volume) is detected and fluctuations in intensity are analyzed using a correlation analysis.[37–40] Phenomenon leading to fluctuations in fluorescence intensity, such as translational and rotational diffusion, triplet-state dynamics, and reaction rates, can be measured with this method.

The translation diffusion coefficient of proteins can be measured using FCS by labeling them with fluorescent markers. For freely diffusing non-interacting proteins at low concentration, the number of proteins in the probe volume varies with time due to thermodynamics fluctuations. From the decay time of autocorrelation function (ACF), also referred to as the diffusion time, the translational diffusion coefficient, D, can be determined, which is related to the hydrodynamic radius (R_H) of the protein through the Stokes–Einstein relation:

$$R_H = \frac{k_B T}{6\pi\eta D} \tag{5}$$

where, k_B is the Boltzmann constant, T is the temperature, and η is the viscosity of the buffer. Measurement of the diffusion time is sensitive to the shape of the probe volume as well as on the excitation power used, because of the saturation effects and photobleaching of the fluo-

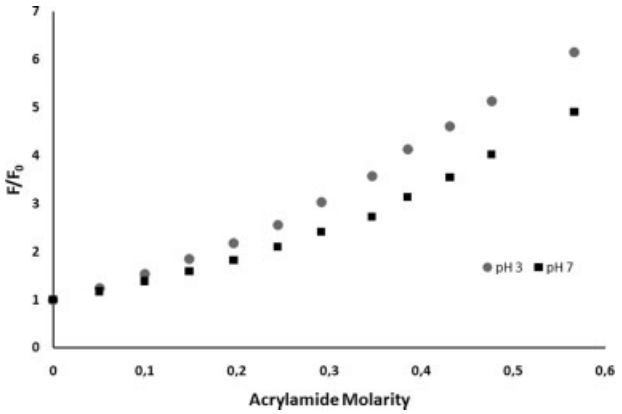

Figure 10

Acrylamide fluorescence quenching measurements (290 nm excitation, 335 nm emission) of 6 μ*M* apomyoglobin in 0.15*M* NaCl at pH 7 and 3. Measurements were performed at 27°C. Fluorescence emission at each acrylamide concentration was corrected for the absorption at 290 nm as indicated in the text.

rophore while in the probe volume.[56,57] While the shape of the probe volume can be calibrated using a fluorescent dye with known diffusion coefficient, such as Rhodamine 6G or Atto655, correction for the excitation power dependences is more difficult as the photoproperties of fluorophores often change when attached to biomolecules and the labeled molecules are illuminated for longer times because of their low mobility. To circumvent these shortcomings of single focus FCS, Enderlein and coworkers have introduced the method of two-focus FCS (2fFCS).[58] In 2fFCS, two offset but overlapping volumes are excited alternately and the photons generated from the two volumes are cross-correlated. As the volumes are offset, an absolute distance enters into the expression for the cross-correlation function (CCF), allowing an absolute determination of the diffusion coefficient independent of the form of the excitation beam. The distance between the two foci can be measured with high accuracy independent of the FCS measurement. The ACF of both excitation volumes and the CCF between the excitation volumes are fitted globally to ensure the best determination of the measured diffusion coefficient. The probe volume was expressed using the following empirical expression:

$$W(\mathrm{r}) = \frac{\kappa(z)}{w^2(z)} \exp\left[\frac{-2(x^2+y^2)}{w^2(z)}\right] \tag{6}$$

where

$$w(z) = w_0\left[1+\left(\frac{\lambda_{ex} z}{\pi w_0^2 n}\right)^2\right]^{1/2},$$

$$\kappa(z) = 1 - \exp\left(-\frac{2a^2}{R^2(z)}\right) \tag{7}$$

and

$$R(z) = R_0\left[1+\left(\frac{\lambda_{em} z}{\pi R_0^2 n}\right)^2\right]^{1/2}$$

w_0 is the lateral beam waist at the focus of the laser, λ_{ex} and λ_{em} are the wavelengths of excitation and emission, respectively, n is the index of refraction of the immersion medium, a is the radius of the confocal pinhole divided by the magnification, and R_0 is a model parameter. The CCF for two volumes separated by a distance δ is given by:

$$G(t,\delta) = \frac{1}{c}\sqrt{\frac{16}{\pi^3 Dt}}\left[\int_{-\infty}^{\infty} dz_1 \int_{-\infty}^{\infty} dz_2 \frac{\kappa(z_1)\kappa(z_2)}{8Dt + w^2(z_1) + w^2(z_2)} \times \exp\left[-\frac{(z_2-z_1)^2}{4Dt} - \frac{2\delta^2}{8Dt + w^2(z_1) + w^2(z_2)}\right]\right] \Big/ \left[\int_{-\infty}^{\infty} k(z)dz\right]^2 \tag{8}$$

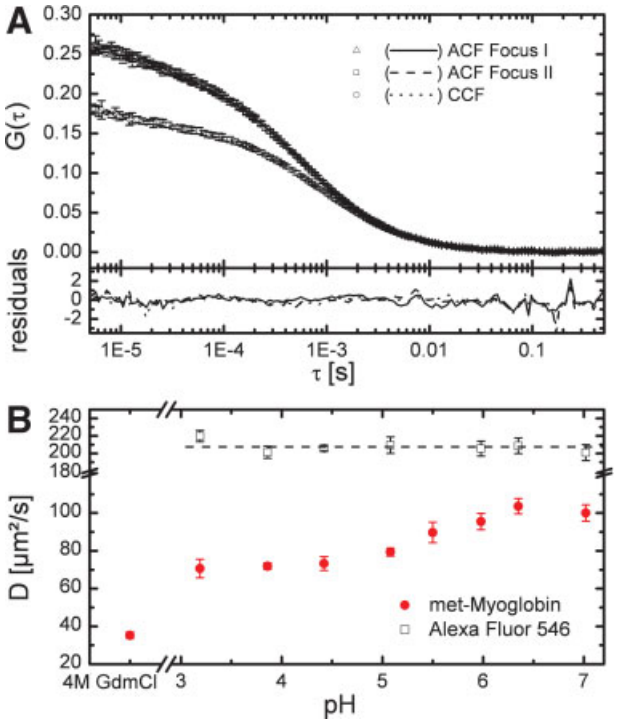

Figure 11

The diffusion coefficient of Tuna metmyoglobin. **A**: The ACF and CCF of the two foci for a 2fFCS experiment on fluorescently labeled Tuna myoglobin at pH 3.19 and 150 m*M* NaCl. The error bars represent the standard deviation of multiple measurements. The fit and reduced residuals are shown as lines in the appropriate graphs. **B**: The diffusion coefficient of Alexa 546 and Tuna metmyoglobin as a function of pH determined using 2fFCS. To the left, a measurement of the diffusion coefficient of fully denatured myoglobin in 4*M* GHCL at pH 5.0 is shown for comparison. The error bars represent the standard deviation of at least 4 measurements. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

A representative 2fFCS measurement of freely diffusing tuna metmyoglobin labeled with Alexa546 in 10 m*M* phosphate 150 m*M* NaCl buffer (pH 3.2) and the corresponding global fit are shown in Figure 11(A). From the data, a diffusion coefficient of 71 ± 5 $\mu m^2\ s^{-1}$ was determined. The diffusion coefficient of tuna Mb and Alexa 546 as a function of pH is given in Figure 11(B). The diffusion coefficient of Alexa 546 (207 ± 6 $\mu m^2\ s^{-1}$) is independent of pH, verifying that there is no significant change in the optical or viscosity properties of the buffer and demonstrating the reproducibility of the apparatus over the course of the measurement. While the free dye could be fit using only a single diffusion component, the ACF and CCF from tuna Mb at intermediate pH values contain additional fast dynamics on the time scale of ~15–25 μs. It was necessary to include an exponential relaxation term in the fit function to obtain reasonable results. We attribute the additional relaxation term to dynamic quenching of the heme group in the more flexible metmyoglobin structures at intermediate pH values. The presence of the additional relaxation term does not impact the determination of the diffusion coefficients. The diffusion coefficient of tuna Mb undergoes a change of ~30% from 100 ± 4 $\mu m^2\ s^{-1}$ at pH 7.0 to 71 ± 5 $\mu m^2\ s^{-1}$ at pH 3.2. Assuming the buffer has the same viscosity as water, the hydrodynamic radius can be calculated from Eq. (6). This corresponds to a hydrodynamic radius (which includes the hydration shell) of ~2.3 ± 0.1 nm at pH 7.0 and 3.3 ± 0.2 nm at pH 3.2. For comparison between the 2fFCS measurements and the MD simulations, we estimated the radius of gyration by assuming a spherical protein:

$$R_g = \sqrt{(3/5)}R_H = 0.775R_H \qquad (9)$$

The experimentally determined radius of gyration changes from 1.8 to 2.5 nm between pH 7.0 and 3.2. To compare the hydrodynamic radius of the acid compact state with that of the fully denatured protein, we have measured the diffusion coefficient of tuna metmyoglobin in 4*M* guanidinium hydrochloride (GHCL) at pH 5. The viscosity of the 4*M* GHCL buffer was calibrating using a fluorescent dye with known diffusion coefficient. The hydrodynamic radius changes from 2.3 nm at neutral pH to 3.3 nm at pH 3.2 to 4.2 ± 0.3 nm in the fully unfolded state. Hence, as also observed in the MD simulations, the size of the acid compact state is between that of the native state than to the fully unfolded state. The absolute value of the radius of gyration from MD is 1.4 and 1.6 at pH 7 and 3, respectively. As the radius of gyration calculated from the MD simulations does not include the hydration shell, the two measurements correspond very well. The agreement at pH 3.2 is not as good, as at pH 7, but considering the many factors that can lead to inaccuracies, both in the experimental as well as simulated data, the absolute agreement between the measured and calculated values are acceptable.

DISCUSSION

Folding intermediates comprise a variety of collapsed conformations that are more compact than the unfolded state and have exposed hydrophobic surfaces, which lead to binding of hydrophobic dyes and a propensity to aggregate. Stable compact intermediates have been frequently observed at low pH and moderate salt concentrations (A states).[2] The acidic compact state of apomyoglobin from sperm whale is a standard for investigating such A-state intermediates. It features a well-ordered hydrophobic core in which the A, G, and H helices are interconnected, as deduced from an increasing number of spectroscopic techniques, in particular NMR.[14,59] However, bulk spectroscopy measurements provide macroscopic observables that are related to average properties of the plethora of structures that populate a partially unfolded state. To understand the characteristics of such non-native protein states in molecular detail by MD simulations, it is important to understand which affects the accumulation of such partially folded state can have on the aggregation processes that are alleged to play a fundamental role in neurodegenerative and amyloid diseases.[8] In this work, we utilize apomyoglobin from yellowfin tuna as a convenient model system for the investigation of the acidic compact state instead of the more commonly studied sperm whale protein. The presence of a single tryptophan in tuna apomyoglobin allows an easier interpretation of its intrinsic fluorescence. The broadening of the fluorescence lifetime distribution (Fig. 9) suggests that the population of apomyoglobin molecules is spread over a large variety of non-native conformations at low pH in the presence of moderate salt concentrations. The high sensitivity of fluorescence-based experimental techniques has the advantage that the protein concentration can be kept very low so that protein aggregation among partially unfolded structures of proteins, and apomyoglobin in particular, can be minimized. In contrast, NMR experiments need higher concentrations, increasing the risk of aggregation. This problem is even more apparent if we consider that the FCS data, obtained at sub-nanomolar protein concentration, still reveal some propensity of acid-denatured apomyoglobin molecules to aggregate.

Previous MD simulations, both on native and acidic apomyoglobin from sperm whale, were based on single MD trajectories.[60,61] In contrast, experimental denaturation studies measure a large ensemble of molecules passing from the native to an unfolded state. Therefore, we consider it more realistic to perform a multiple trajectory study with short MD runs rather than to calculate a long, single MD trajectory.[29] Thus, we have performed 25 different MD simulations of protonated apomyoglobin at pH 3 in the presence of 150 m*M* NaCl, each for a total of 3.0 ns. A heating step using a temperature jump was needed because protonation of the apomyoglobin structure alone, calculated for the appropriate low pH condi-

tions, was unable to substantially unfold the native conformation in a 2.0-ns MD run (data not shown). This finding is different from that observed by Onufriev *et al.*[61] where protonated sperm whale apomyoglobin was unfolded at 300 K. This different result may be due to the fact that, in their study at pH 2.6 and low ionic strength, they considered basic amino acid residues to be totally protonated. By contrast, we calculated the protonation by considering the specific microenvironment of each basic residue in context of the three-dimensional structure. We also included the effect of 0.15*M* salt. Moreover, the sequences of tuna and sperm whale myoglobin are quite different; only 35% of the amino acid residues are identical and the two proteins also differ in length by seven residues. A further difference between the two studies concerns the different MD simulation procedures. In fact, we used periodic boundary conditions and a dodecahedral box with 13,000 solvent molecules and ions to attenuate boundary effects, whereas in the earlier referenced paper, periodic conditions were not applied, but the apomyoglobin molecule was enclosed by a spherical shell of solvent containing a similar number of water molecules to what we used in our MD simulations.

The structure-based average values of α-helical content and the collisional quenching constant of the intrinsic fluorescence emission of our 25 MD-generated acidic conformers of apomyoglobin are in good agreement with the reported experimental data. Diffusion coefficients obtained by FCS measurements are compatible with the calculated ones, although somehow larger at pH 3.2 than expected from the MD simulations. However, these differences are not large if we take in account the intrinsic uncertainties and limitations that characterize both the experimental methodologies (e.g., affected by the propensity of apomyoglobin to aggregate in salt acidic conditions) and MD simulations (approximation of the force field parameters, cut off, periodic conditions, and a limited number of generated structures).

The differences among the experimental values and the average values calculated by the successive addition of MD generated structures does not changes significantly when considering the further generated structures. This finding corroborates the notion that the ensemble of the 25 acidic structures can be viewed as an appropriate sample of the distribution of conformers within the acidic compact state of tuna apomyoglobin. In particular, 10 of 25 structures can be located in a single cluster on the basis of the mean square deviations of their atomic coordinates (Fig. 6). The common structural themes observed in this cluster of acidic structures are represented essentially by the four α-helix segments that could be correspond to translated and shortened A, E, G, and H-helices of the native apomyoglobin structure. Prolonging the MD trajectory at 500 K and pH 3 for an addition nanosecond for each acidic structure did not cause any significant additional loss of structure of these MD generated acidic forms (Table V). After short MD simulations at intermediate temperatures to permit a better equilibration at 300 K, further extending of MD trajectory does not seem to introduce large changes in the acidic structure. On the basis of this observation, we consider the MD generated acidic structures quite stable. This stability allows the accumulation of non-native structures with exposed hydrophobic segments that could nucleate protein aggregation as experimentally observed.[10,35] In particular, short segments of β-sheet secondary structure are present in several of the acidic structures. This finding appears interesting since diseases in humans, such as Creutzfeld–Jakob, fatal familial insomnia, kuru, and so forth, are associated with misfolding of the cellular prion PrPC protein to PrPSc form. In such diseases, the mechanism of conversion to a pathogenic form requires a substantial change of conformation from an α-rich monomer to a β-sheet-rich amyloid structure.[62]

CONCLUSIONS

A procedure has been presented to explore partial unfolding of a protein by MD simulation by using multiple trajectories and comparing the expected and experimental values of specific properties related to the secondary structural organization of protein macromolecule, the local tertiary structure (as tryptophanyl accessibility to acrylamide quencher), and the overall size of the protein (radius of gyration). The comparison between experimental and theoretical results gives a reasonable agreement considering the intrinsic limitations of the simulation and spectroscopic methods. However, the results seems to corroborate the point of view that folding intermediates are not represented by a single conformation with specific structural properties but rather they constitute a plethora of conformations largely different sometime but having common energetic characteristics. Some of these conformations have larger hydrophobic, solvent-exposed segments, specifically in the region surrounding the residues L25 and A54 and represent potential candidates to trigger the protein aggregation processes.

METHODS

Myoglobin purification and apomyoglobin extraction

The main component of tuna myoglobin was prepared according to the method described by Bismuto *et al.*[20] The homogeneity of the preparations was controlled by sodium dodecyl sulfate/polyacrylamide gel electrophoresis with 15% gels and 5% stacking gels. The heme was removed from myoglobin by the 2-butanone extraction procedure.[63] The contamination of the apoprotein by myoglobin was assessed spectrophotometrically. In all cases no significant absorption was observed in the Soret

region. The myoglobin concentration was determined spectrophotometrically at 409 nm in 0.01*M* phosphate, 0.15*M* NaCl buffer (pH 7.0) by using 139,000 M^{-1} cm^{-1} as the extinction coefficient at 407 nm.[47] For apomyoglobin, the molar extinction at 280 nm was calculated from the tryptophan and tyrosine content by using molar extinction coefficient of 5500 and 1250 M^{-1} cm^{-1}, respectively.[64] All common chemicals were reagent grade and were purchased from Sigma-Aldrich.

Circular dichroism spectroscopy

Far-UV CD spectra were recorded on a JascoJ-715 spectropolarimeter equipped with a temperature controller Neslab RTE-110. Cuvettes of 1-mm path length were used in the range of wavelength 250–200 nm. Spectra were acquired at 0.2-nm intervals with a 4 s integration time and a bandwidth of 1.0 nm. Five scans were obtained for each spectrum. Photomultiplier absorbance did not exceed 600 V in the whole examined spectral region.

Steady-state and frequency domain fluorescence measurements

Steady-state measurements were done on a Greg PC fluorometer from ISS (Champaign, IL). The excitation wavelength was set at 295 nm to exclude the tyrosine contribution to the fluorescence emission and the absorbance of the samples at the excitation wavelength did not exceed 0.07. The temperature of the compartment was controlled by using an external circulator (Neslab Model LT50). Frequency domain techniques were used to measure the fluorescence decay of all samples in the range 1–200 MHz using a multifrequency phase shift and demodulation cross-correlation fluorometer[36] GREG 200 (ISS, Urbana, IL). The emission was observed using an optical filter combination of UV34 and U340 (from Oriel). The temperature was monitored continuously during the measurements by attaching a thermocouple to the sample cuvette. Readings of the thermocouple were monitored by an Omega Digicator (from Omega Engineering, Stamford, CT) with an accuracy of $\pm 0.1^{\circ}C$.

The absorbance of the protein solution did not exceed 0.1 at the exciting wavelength. The lifetime data analysis was performed with Globals Unlimited[65] (University of Illinois at Urbana, Champaign, IL).

In the acrylamide quenching experiments, the steady-state fluorescence intensity at 335 nm was corrected for the absorption of acrylamide (A_Q^{290}) at the excitation wavelength (290 nm). A correction factor, $-\log(A_Q^{290}/2)$, was applied.[66]

FCS determination

Protein labeling

Purified yellowfin tuna metmyoglobin was incubated with an equimolar (150 μ*M*) mixture of Alexa Fluor 546 succinimidyl ester (Invitrogen) in 25 m*M* HEPES 75 m*M* NaCl (pH 8.2) for 2 h at room temperature. The labeled protein was separated from the remaining free dye by three purification steps using Biorad Micro Bio-Spin 6 spin columns at 1000 rcf for 4 min. The protein was stored in 10 m*M* sodium phosphate buffer (with 150 m*M* NaCl) at pH 8 at micromolar concentration until directly before the measurements. The purified labeled protein was then diluted with buffer at the respective pH values (3.19, 3.86, 4.42, 5.08, 5.50, 5.98, 6.35, and 7.02) or buffer containing 4*M* GHCL, 10 m*M* sodium phosphate and 150 m*M* NaCl (pH 5.0) to concentrations appropriate for the FCS measurements.

Two-focus FCS setup

Two-focus FCS experiments[58] were performed on a custom designed confocal microscope[67,68] built around an inverted epi-fluorescence microscope (Axiovert 35, Carl Zeiss, Göttingen, Germany). For pulsed interleaved excitation (PIE) of two separated but overlapping volumes, excitation light pulses of alternating perpendicular polarizations were generated and passed through a Nomarski prism (DIC 1033-911, Carl Zeiss) before entering the microscope objective (UPLAPO 60x/1.2w, Olympus, Hamburg, Germany). A train of pulses at 10 MHz was generated from a single mode-locked frequency-doubled solid-state laser of 532 nm (GE-100, Time–Bandwidth Products, Zürich, Switzerland) and a pulse picker (Pulse select, APE, Berlin, Germany). Alternating polarizations were achieved using a half-wave plate (AHWP05M-630, Thorlabs, Dachau, Germany) followed by separation and recombination of the light pulses with a pair of polarizing beam splitters (48571, Edmund Optics, Barrington, NJ). The S-polarized pulses were delayed by 30 ns with respect to the P-polarization with an optical delay line. The recombined pulse train was fed into a polarization maintaining single mode fiber (QPMJ, OZ Optics, Ottawa, Canada) and reflected by a dichroic beam splitter (z532/633 xr, AHF, Tübingen, Germany) before entering the Nomarski prism. The emitted fluorescence photons were focused with an achromatic lens (f = 150 mm) onto a confocal pinhole (100 μm), passed through an emissions filter (HQ 582/50, AHF) and detected by an avalanche photodiode (SPCM-AQR-14, Perkin–Elmer, Fremont, CA). As in PIE,[69] the photon arrival time with respect to the laser trigger was measured with time-correlated single-photon counting (Timeharp 200, Pico-Quant, Berlin, Germany) and used to attribute each recorded photon to the corresponding excitation pulse and therefore to the respective detection volume. After calculation of the ACF and CCF, the data were analyzed by custom-written software kindly provided by Jörg Enderlein *et al.*[57] Prior to the determination of diffusion coefficients, the separation of the foci was determined as 237 nm, using the diffusion coefficient

of rhodamine 6G in water (2.85×10^{-6} cm^2 s^{-1}) given by the average of two independent investigations.[39,70] The two-focus measurements were performed with an excitation power of 35 μW per focus in a temperature-controlled laboratory at 22°C.

MD simulations

All MDs simulations have been performed with GROMACS software package v3.3 and united-atom force field.[30] The starting coordinates of apomyoglobin structure have been taken from the Protein Data Bank, entry 1MYT.[43] Prior to performing any calculation, crystallization water and other solvent molecules coordinates were removed (for apomyoglobin MD simulations, the heme coordinates were also removed). Moreover, the protonation state of each titratable residue was calculated by means of the H++ web server,[37] which is able to calculate the p*K*a of titratable residues of a protein at a given pH. Parameters adopted include: a value of four for the internal relative dielectric constant and 80 for the external; for the native state, pH 7.0 and a salinity of 0.15*M*, were used as indicated; electrostatics were treated by the PB method. According to this procedure, a net charge of +2 and +4 were found for met-myoglobin and apomyoglobin, respectively. In preparation for the MD simulations, the myoglobin molecule was placed in the center of a rhombic dodecahedron of approximate dimensions ($6.07 \times 6.07 \times 4.29$) nm^3. The minimum distance between the protein and the edge of the box was 0.8 nm. All water molecules with their oxygen atom closer than 0.23 nm to any non-hydrogen protein atom were removed. The box was full of equilibrated SPC rigid water molecules (Single Point Charge, i.e., with a partial charge on each of the three constituting atoms). To ensure electroneutrality in the box, an equal number of Cl^- counterions were added to the simulation ensemble by substituting randomly the required number of water molecules; the same procedure was followed in adding the proper number of Na^+ and Cl^- pairs to bring salinity up to 0.15*M*. In total, the simulated system was composed of protein atoms, ions, and 13,532 solvent (for apomyoglobin 1376 + 13,559) atoms. Initial atom velocities were taken from a Maxwell–Boltzmann distribution at 300 K; temperature and pressure were kept constant by weak coupling to an external reference, using a Berendsen thermostat and barostat[71] with coupling constants of 0.1 and 1.0 ps, respectively; all molecular groups (protein, solvent, and counterions) were coupled independently. Long-range electrostatic interactions were calculated using the Particle Mesh Ewald method[72] with a 0.9 nm cut-off. The LINear Constraints Solver (LINCS) algorithm[73] was used to restrain bond lengths. To relieve unfavorable non-bonded interactions with the added water molecules, the system was energy-minimized by a steepest descent algorithm followed by a short 100 ps simulation during which the protein and non-hydrogen atoms were harmonically restrained with a force constant of 1000 kJ mol^{-1} nm^{-2}. All restraints were then removed and the simulation was run for 10 ns, saving coordinates every 0.5 ps and velocities every 10 ps for subsequent analysis. To simulate the acidic state of apomyoglobin, the output structure of the 10 ns MD simulation was first protonated according to pH 3.0 with the aid of the H++ web server[33]; this produced a considerable increase in the net charge when compared to the neutral pH state. Afterwards, atomic velocities were again generated from a Maxwell–Boltzmann distribution, but at 500 K; then the simulation ran for 1 ns with thermal coupling to a bath at 500 K. Three further short MD runs (200 ps) were performed at intermediate temperatures of 450, 400, and 350 K. At the end of the 500-K MD run, and at the end of the short MD runs at the intermediate temperatures, the protonation state was calculated again with the H++ web server, and in these new conditions a further MD step at 300 K was carried out for 1 ns. This "heat-driven" unfolding procedure was performed 25 times with different starting velocities, thereby generating 25 different acidic forms. All structural and dynamical properties have been calculated with proper GROMACS routines.

ACKNOWLEDGMENTS

We thank Prof. Jörg Enderlein for helpful advice and for providing the analysis routine for the 2fFCS data and Dr. Fabrizio Mancinelli for his technical assistance.

REFERENCES

1. Baldwin RL. Protein folding. Matching speed and stability. Nature 1994;369:183–184.
2. Fink AL. Compact intermediate states in protein folding. Annu Rev Biophys Biomol Struct 1995;24:495–522.
3. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free energy surface from theory and experiment. Trensa Biochem Sci 2000;25:331–339.
4. Mayor U, Johnson CM, Daggett V, Fersht A. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. Proc Natl Acad Sci USA 2000;97:13519–13522.
5. Plotkin SS, Oniuchic JN. Understanding protein folding with energy landscape theory, Part I: basic concept. Quart Rev Biophys 2002;35:111–167.
6. Plotkin SS, Oniuchic JN. Understanding protein folding with energy landscape theory, Part II: quantitative aspects. Quart Rev Biophys 2002;35:205–286.
7. Muñoz V. Conformational dynamics and ensembles in protein folding. Annu Rev Biophys Biomol Struct 2007;36:395–412.
8. Chiti F, Dobson CM. Protein misfolding. Functional amyloid and human disease. Annu Rev Biochem 2006;75:333–366.
9. Morozova-Roche L, Marisauskas M. A false paradise-mixed blessings in the protein universe: amyloid as a new challenge in drug development. Curr Med Chem 2007;14:1221–1230.
10. Frauenfelder H, Nienhaus GU, Johnson JB. Rate processes in proteins. Ber Bunsenges Phys Chem 1991;95:272–278.
11. Parak FG, Nienhaus GU. Myoglobin, a paradigm in the study of protein dynamics. Chem Phys Chem 2002;3:249–254.

12. Sirangelo I, Malmo C, Iannuzzi C, Mezzogiorno A, Bianco MR, Papa M, Irace G. Fibrillogenesis and cytotoxic activity of the amyloid-forming apomyoglobin mutant W7FW14F. J Biol Chem 2004;279:13183–13189.
13. Irace G, Bismuto E, Savy F, Colonna G. Unfolding pathway of myoglobin: molecular properties of intermediate forms. Arch Biochem Biophys 1986;244:459–469.
14. Barrick D, Baldwin RL. The molten globule intermediate of apomyoglobin and the process of the protein unfolding. Protein Sci 1993;2:869–876.
15. Goto Y, Fink A. Phase diagram for acid conformational states of apomyoglobin. J Mol Biol 1990;214:803–805.
16. Hughson FM, Wright P, Baldwin RL. Structural characterization of a partly folded apomyoglobin intermediate. Science 1990;249:1544–1548.
17. Jennings PA, Wright PE. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science 1993;262:892–896.
18. Eliezer D, Yao J, Dyson HJ, Wright P. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. Nat Struct Biol 1998;5:148–155.
19. Alcalà R, Gratton E, Prendergast F. The interpretation of fluorescence decay in proteins using continuous lifetime distribution. Biophys J 1987;51:925–936.
20. Bismuto E, Gratton E, Irace G. Effect of unfolding on the tryptophanyl fluorescence lifetime distribution in apomyoglobin. Biochemistry 1988;27:2132–2136.
21. Bismuto E, Irace G. Unfolding pathway of apomyoglobin. Simultaneous characterization of acidic conformational states by frequency domain fluorometry. J Mol Biol 1994;241:103–109.
22. Bismuto E, Irace G, Servillo L, Giovane A, Colonna G. Conformational stability and basal metabolic rate: reexamination of the case of myoglobin. Experientia 1984;40:1400–1401.
23. Sheraga HA, Khalili M Liwo A. Protein-folding dynamics: overview of molecular simulation techniques. Annu Rev Phys Chem 2007;58:57–83.
24. Van Der Spoel D, Lindhal E, Hess B, Groenhof G, Mark AE, Berendsen JC. GROMACS: fast, flexible and free. J Comput Chem 2005;26:1701–1718.
25. Pearlman DA, Case DA, Calwell JW, Ross WS, Cheatham TE, III, DeBolt S, Ferguson D, Seibel G, Kollmann P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun 1995;91:1–41.
26. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan DJ, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 1983;4:187–217.
27. Levinthal C. How to fold graciously. In: DeBrunner P, Tsibris JCM, Munck E, editors. Mossbauer spectroscopy in biological systems. Urbana, IL: University of Illinois Press; 1969. pp 22–24.
28. Smith L. Computational methods for generating models of denatured and partially folded proteins. Methods 2004;3:144–150.
29. Kazmirski SL, Li A, Daggett V. Analysis methods for comarison of a multiple molecular dynamics trajectories: applications to protein unfolding pathways and denatured ensembles. J Mol Biol 1999;290:283–304.
30. Kuzmenkina EV, Heyes CD, Nienhaus GU. Single molecule förster resonance energy transfer study of protein dynamics under denaturing conditions. Proc Natl Acad Sci USA 2005;102:15471–15476.
31. Kuzmenkina EV, Heyes CD, Nienhaus GU. Single molecule FRET study of denaturant induced unfolding of RNase H. J Mol Biol 2006;357:313–324.
32. Bashford D, Karplus M. p*Ka*'s of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
33. Gordon JC, Meyers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: a server for estimating pK_as and adding missing hydrogens to macromolecules. Nucleic Acids Res 2005;33:W369–W371.
34. Myers J, Grothaus G, Narayanan S, Onufriev A. A simple clustering algorithm can be accurate enough for use in calculations of p*Ks* in macromolecules. Proteins 2006;63:928–938.
35. Mancinelli F, Caraglia M, Abbruzzese A, D'Ambrosio G, Massa R, Bismuto E. Non-thermal effects of electromagnetic fields at mobile phone frequency on the refolding of an intracellular protein: myoglobin. J Cell Biochem 2004;93:188–196.
36. Gratton E, Limkeman M. A continuously variable frequency cross-correlation phase fluorometer with picosecond resolution. Biophys J 1983;44:2714–2723.
37. Magde D, Elson EL, Webb WW. Thermodynamic fluctuations in a reacting system-measurement by fluorescence correlation spectroscopy. Phys Rev Lett 1972;29:705–708.
38. Elson EL, Magde D. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. Biopolymers 1974;13:1–27.
39. Magde D, Elson EL, Webb WW. Fluorescence correlation spectroscopy. II. An experimental realization. Biopolymers 1974;13:29–61.
40. Thompson NL. Fluorescence Correlation Spectroscopy. In: Lakowicz JR, editor. Topics in fluorescence spectroscopy, Vol. 1: Techniques. New York: Plenum Press; 1991. pp 337–378.
41. Eftink MR, Ghiron CA. Fluorescence quenching studies with proteins. Anal Biochem 1981;114:199–227.
42. Zelent B, Kusba J, Gryczynski I, Johnson ML, Lakowicz JR. Time-resolved and steady-state fluorescence quenching of *N*-acetyl-L-tryptophanamide by acrylamide and iodide. Biophys Chem 1998;73:53–75.
43. Birnbaum GI, Evans SV, Przybylska M, Rose DR. 1.70 Angstrom resolution structure of myoglobin from yellowfin tuna. An example of a myoglobin lacking the D helix. Acta Crystallogr D 1994;50:283–289.
44. Maiorov VN, Crippen GM. Size independent comparison of protein in three-dimensional structure. Proteins 1995;22:273.
45. Hirst JD, Brooks CL, III. Helicity, circular dichroism and molecular dynamics of proteins. J Mol Biol 1994;243:173–178.
46. Brooks CL, III. Characterization of "native" apomyoglobin by molecular dynamics simulation. J Mol Biol 1992;227:375–380.
47. Bismuto E, Colonna G, Savy F, Irace G. Myoglobin structure and regulation of solvent accessibility of heme pocket. Int J Pept Protein Res 1985;29:195–207.
48. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
49. Lees JG, Miles AJ, Wien F, Wallace BA. A reference database for circular dichroism spectroscopy covering fold and secondary structure space. Bioinformatics 2006;22:1955–1962.
50. Valeur B. Molecular fluorescence: principles and application. Weinheim: Wiley-VCH; 2001. pp 72–124.
51. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105.
52. Andrade MA, Chacon P, Merelo JJ, Moran F. Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. Prot Eng 1993;6:383–390.
53. Whitemore L, Wallace B. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acid Res 2004;32:W668–W673.
54. Frauenfelder H, Parak F, Young RD. Conformational substates in proteins. Annu Rev Biophys Chem 1988;17:451–479.
55. Barbieri B, De Piccoli F, Gratton E. Synthesizer's phase noise in frequency domain fluorometry. Rev Sci Instrum 1989;60:3201–3206.
56. Hess ST, Webb WW. Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. Biophys J 2002;83:2300–2317.
57. Enderlein J, Gregor I, Patra D, Fitter J. Art and artefacts of fluorescence correlation spectroscopy. Curr Pharm Biotechnol 2004;5:155–161.

58. Dertinger T, Pacheco V, Hocht Ivd, Hartmann R, Gregor I, Enderlein J. Two-focus fluorescence correlation spectroscopy: A new tool for accurate and absolute diffusion measurements. Chem Phys Chem 2007;8:433–443.
59. Cocco M, Lecomte JTJ. Characterization of hydrophobic cores in apomyoglobin: a proton NMR spectroscopy study. Biochemistry 1990;20:11067–11072.
60. Tirado-Rives J, Jorgensen WL. Molecular dynamics simulations of the unfolding of apomyoglobin in water. Biochemistry 1993;32: 4175–4184.
61. Onufriev A, Case D, Bashford D. Structural details, pathways, and energetics of unfolding apomyoglobin. J Mol Biol 2003;325:555–567.
62. Gerber R, Tahiri-Alaoui A, Hore PJ, James W. Oligomerization of the human prion protein proceeds via a molten globule intermediate. J Biol Chem 2007;282:6300–6306.
63. Teale FWJ. Cleavage of the heme-protein link by acid methylethylchetone. Biochim Biophys Acta 1959;35:543.
64. Wetlaufer DB. Ultraviolet spectra of proteins and aminoacids. Adv Protein Chem 1962;17:304–320.
65. Beechem JM, Gratton E, Amelot M, Knutson JR, Brand L. The global analysis of fluorescence intensity and anisotropy decay data: second generation theory and programs. In: Lakowicz JR, editor. Topics in fluorescence spectroscopy 2. New York: Plenum Press; 1991. pp 241–305.
66. Parker CA. Photoluminescence of solutions. New York: Elsevier; 1968. pp 220–226.
67. Schenk A, Ivanchenko S, Rocker C, Wiedenmann J, Nienhaus GU. Photodynamics of red fluorescent proteins studied by fluorescence correlation spectroscopy. Biophys J 2004;86:384–394.
68. Wiedenmann J, Ivanchenko S, Oswald F, Schmitt F, Rocker C, Salih A, Spindler KD, Nienhaus GU. EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. Proc Natl Acad Sci USA 2004;101: 15905–15910.
69. Müller BK, Zaychikov E, Bräuchle C, Lamb DC. Pulsed interleaved excitation. Biophys J 2005;89: 3508–3522.
70. Gell C, Brockwell DJ, Beddard GS, Radford SE, Kalverda AP, Smit DA. Accurate use of single molecule fluorescence correlation spectroscopy to determine molecular diffusion times. Single Mol 2001;2: 177–181.
71. Berendsen HJC, Postma WF, van Gunstener A, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Phys Chem 1984;81:3684–3690.
72. Luty BA, Tironi IG, van Gunsteren WF. Lattice-sum methods for calculating electrostatic interactions in molecular simulations. J Chem Phys 1995;103:3014–3021.
73. Hess B, Bekker H, Berendsen HJC, Fraaje JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997; 18:1463–1472.