Protein Science (1997), *6*:2134–2142. Cambridge University Press. Printed in the USA.
Copyright © 1997 The Protein Society

Structural heterogeneity of the various forms of apomyoglobin: Implications for protein folding

RUDOLF GILMANSHIN,[1] R. BRIAN DYER,[2] AND ROBERT H. CALLENDER[3,4]

[1]Department of Physics, The City College of the City University of New York, New York, New York 10031
[2]CST-4, Mail Stop J586, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
[3]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461

(RECEIVED January 27, 1997; ACCEPTED January 23, 1997)

Abstract

Temperature-induced denaturation transitions of different structural forms of apomyoglobin were studied monitoring intrinsic tryptophan fluorescence. It was found that the tryptophans are effectively screened from solvent both in native and acid forms throughout most of the temperature range tested. Thus, the tryptophans' surroundings do not show a considerable change in structure where major protein conformational transitions have been found in apomyoglobin using other techniques. At high temperatures and under strong destabilizing conditions, the tryptophans' fluorescence parameters show sigmoidal thermal denaturation. These results, combined with previous studies, show that the structure of this protein is heterogeneous, including native-like (tightly packed) and molten globule-like substructures that exhibit conformation (denaturation) transitions under different conditions of pH and temperature (and denaturants). The results suggest that the folding of this protein proceeds via two "nucleation" events whereby native-like contacts are formed. One of these events, which involves AGH "core" formation, appears to occur very early in the folding process, even before significant hydrophobic collapse in the rest of the protein molecule. From the current studies and other results, a rather detailed picture of the folding of myoglobin is presented, on the level of specific structures and their thermodynamical properties as well as formation kinetics.

Keywords: apomyoglobin; denaturation; protein folding; temperature melting; tryptophan fluorescence

The folding of apomyoglobin (apoMb, myoglobin without its heme group) has received intense scrutiny, in part because the protein is relatively simple. It may serve as an archetype for folding of proteins which are small, single-domain, and globular. Another attractive feature of apoMb is the ability to access different equilibrium forms of the protein having intermediate stability between the native and unfolded states by varying the pH and salt concentration. Furthermore, recent evidence suggests that these equilibrium intermediate states are structurally close to kinetic intermediates on the folding pathway (native, molten globule, and unfolded; Barrick & Baldwin, 1993). A striking structural similarity between the acid form of apoMb and an early folding intermediate has been demonstrated (Jennings & Wright, 1993; Eliezer et al., 1995).

While the folding kinetics of small, single-domain proteins like apoMb are usually heterogeneous, the unfolding kinetics of those proteins generally follow a single first-order reaction whose amplitude accounts for the entire change measured in the equilibrium transition (Schmid, 1992). This has been taken to imply that no partial unfolding precedes the rate-limiting step for complete unfolding and that the folding reaction is a highly cooperative "all or nothing" transition. Recent studies have shown that this is not the case for apoMb, where a fast pre-denaturation process occurs in both the acid (Jamin & Baldwin, 1996) and the native forms (Dyer et al., 1996; Gilmanshin et al., 1997). A "nucleation-limited sequential folding" model was advanced to explain this behavior in the acid form, in analogy to similar observations of a burst phase in the unfolding of oligonucleotides (Jamin & Baldwin, 1996). Structurally, the latter is ascribed to the rapid unzipping of base pairs at either end of the double helix in the oligonucleotides. The model for apoMb combines the properties of gradual and two-state unfolding but does not offer a structural basis for such unusual behavior.

Our recent infrared (IR) studies of the equilibrium and transient denaturation of native apoMb provide some insight into the structural basis of the multi-phase kinetics of melting of apoMb. IR absorbance measurements reveal an unexpected structural heterogeneity within the protein (Gilmanshin et al., 1997). Specifically, discrete substructures are observed, composed of either native-like, tightly packed helices or loose, solvated helices. These substructures exhibit kinetics and thermodynamics properties that are

Reprint requests to: Robert Callender, Department of Biochemistry, Albert Einstein College of Medicine, Bronx, NY 10461; e-mail: call@aecom.yu.edu.
[4]On leave from the Physics Department, The City College of the City University of New York, New York, New York 10031.

very different. We have postulated that one native-like substructure comprises a complex between the A, G, and H helices (AGH core, including at least fragments of the A, G, and H helices and their tightly packed side chains). These helices exist in the acid form (Hughson et al., 1990) and are formed first on the protein-folding pathway (Jennings & Wright, 1993).

The purpose of this work is to investigate the equilibrium denaturation of the AGH core within the various forms of apoMb, using fluorescence spectroscopy. Apomyoglobin contains two tryptophan residues (Evans & Brayer, 1988), both in the A helix and fortuitously close to the junctions of the A helix with the G and H helices. The dependence of the tryptophan fluorescence spectrum position and intensity on the surrounding structure thus provides a sensitive measure of the structure of the AGH core (Postnikova et al., 1991; Rischel et al., 1996). We have monitored the intrinsic tryptophan fluorescence to follow the denaturation transitions of apoMb under the same conditions that we previously used in our IR absorbance study (Gilmanshin et al., 1977). Comparison of the complementary information from fluorescence (the structure surrounding the tryptophans, hence the AGH core) and IR (secondary structure and solvation of helices) provides a more complete picture of the extent of structural heterogeneity of the different forms of apoMb. We have studied the horse myoglobin protein here. However, its acid/salt denaturation behavior is quite similar to that of sperm whale myoglobin (Fink et al., 1991; Hughson et al., 1991). Therefore, in what follows below, both these systems are discussed interchangeably.

We find that the structure of native apoMb is even more heterogeneous than revealed by the IR measurements alone. Its structure includes two experimentally distinguishable domains with native-like properties. One of them exhibits a transition between native-like and molten globule–like structures with a pH change. Another one includes the A helix and probably consists of an A, G, and H interhelical core (AGH core). This substructure remains unmelted in the acid form over the entire temperature range measured. It is shown, however, that the AGH core can be melted under more destabilizing conditions. The results provide a structural rationale for the burst phase in the unfolding measurements of the acid form of apoMb since the AGH core responds independently, or quasi-independently, to denaturation conditions on a slow time scale compared to the faster response of other non-native-like parts of the protein.

Results

The different forms of apoMb are accessible by varying the pH and salt concentration (Fink et al., 1991). For a low salt concentration (10–20 mM), the apoMb structure is highly pH dependent; at pH* 5.3, it adopts its most stable, native-like structure; at pH* 4.2, it adopts its compact acid form (the so-called **I** form); and it is in a very expanded state at pH* 3.0. An increase of salt concentration (to 0.25 M NaCl) at pH* 3.0 contracts the protein back into the **I** form.

All fluorescence measurements were made in D_2O for two reasons. First and foremost, this allows interpretation of the fluorescence study results together with that of the equilibrium and fast kinetics T-jump study of apoMb unfolding carried out previously by IR spectroscopy (Gilmanshin et al., 1997). In addition, D_2O decreases the fluorescence temperature dependence of the indole ring of tryptophan (Kirby & Steiner, 1970), which reduces the need for corrections of the raw data. D_2O substitution shifts the transition temperature slightly, if at all (the shift is less than 5 °C; Makhatadze et al., 1995). The shift in the pH dependence is correspondingly small (the shift is less than 0.2 pH units for its phase boundaries and occurs presumably at pH < 4; Makhatadze et al., 1995).

The samples were prepared at room temperature, cooled down to the lowest temperature, and then increased by 4–7 °C increments. The fluorescence spectra were measured at each temperature after equilibration. Subsequent to measurement at the highest temperature, the sample was cooled again to check for reversibility. The position of the tryptophan spectrum was reproducible within a precision of 1 nm. However, reproducibility of the fluorescence intensity depended upon conditions and the highest temperature reached. Below 50 °C, all measurements were completely reproducible. After cooling from 80 °C (the worst case), measurements at pH* 5.3, 4.2, and 3.0 at lower salt concentrations were reversible within 5–10%. At pH* 3.0 and higher salt concentration, reversibility was within 20%. Incomplete reversibility is probably the result of the tendency of apoMb to associate at high temperatures, which varies for the different protein forms (Griko et al., 1988; Nishii et al., 1995).

The temperature dependence of both the intensity of intrinsic tryptophan fluorescence of apoMb and the wavelength of the spectral maximum (λ_{max}) is influenced by different interactions between the indole ring and its surroundings and, therefore, reflects different levels of transformation of protein structure (Eftink, 1994). The position of the fluorescence spectrum depends on polarity of the fluorophore's surroundings. Screening of a tryptophan residue from solvent by a protein structure results in a blue-shifted fluorescence. Protein unfolding thus red-shifts the spectrum toward the position of the fluorescence spectrum of a solvated tryptophan residue. Therefore, the fluorescence λ_{max} is most dependent on large-scale changes of protein structure that permit tryptophan solvation. The fluorescence intensity, in contrast, is a balance of many different quenching factors. These can be primary (i.e., due to a temperature change) and secondary (due to a protein structure transformation). The latter can be specific (e.g., due to the spatial proximity of a tryptophan residue to specific protein side chains within its native structure) or non-specific (e.g., due to a screening from a dissolved quencher). Hence, fluorescence intensity is sensitive to both *global* and *local* changes in protein conformation.

Figure 1 shows changes with temperature of the fluorescence spectrum position for the different protein forms. The position of λ_{max} is presented in Figure 1A and a ratio of fluorescence intensities I_{315}/I_{365}, measured before and after the maximum, is given in Figure 1B. Neither of these parameters are temperature dependent for an isolated tryptophan fluorophore in water, and therefore neither needs any correction. The tryptophan fluorescence spectrum is rather broad, and this restricts the accuracy in determining λ_{max} (±0.5 nm in our studies). The ratio I_{315}/I_{365} is more sensitive to small displacements, although it cannot be interpreted directly in terms of λ_{max}.

For the native state of apoMb, the maximum of the fluorescence spectrum is 331–333 nm between 10 and 60 °C with only a small temperature dependence. This result is similar to previous results (Sirangelo et al., 1994) for the same protein state but with slightly different conditions (334–335 nm between 15 and 58 °C, 0.01 M sodium phosphate, H_2O, excitation at 295 nm). At temperatures higher than 60 °C, the fluorescence maximum exhibits a stronger shift to longer wavelengths, reaching 337 nm at 80 °C. This reveals a conformation change of native apoMb at temperatures clearly higher than 60 °C.

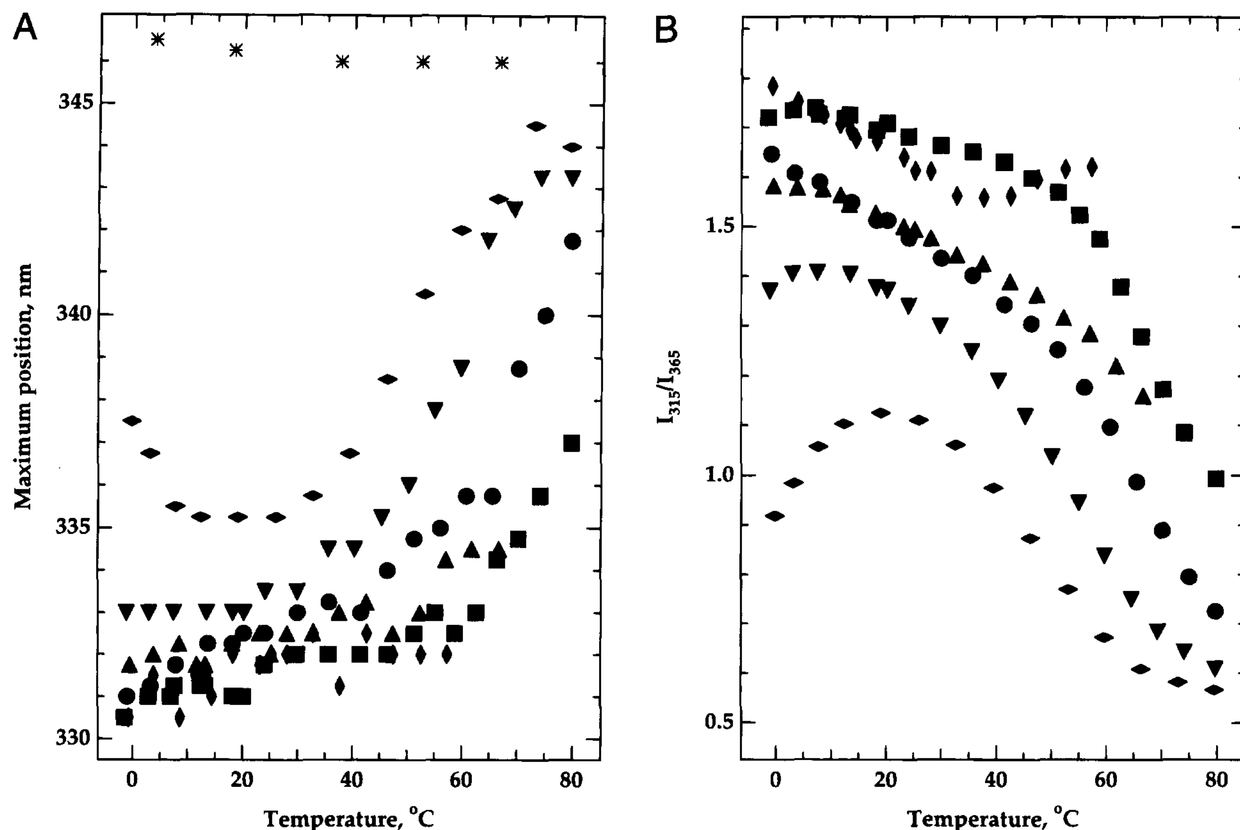

Fig. 1. (A) Maximum position of apomyoglobin's intrinsic fluorescence and (B) ratio of intensities at 315 and 365 nm, I_{315}/I_{365}, at different temperatures. (■), 0.01 M sodium cacodylate, pH* 5.3; (▲), 0.01 M sodium acetate, pH* 4.2; (♦), 0.25 M NaCl, pH* 3.0; (●), 0.15 M NaCl, pH* 3.0; (▼), 0.08 M NaCl, pH* 3.0; (◆), 0.02 M NaCl, pH*3.0; (*), L-tryptophan in D_2O.

Similar to the native state, both apoMb acid forms at pH* 4.2 and pH* 3.0 (high salt concentration) demonstrate only a weak temperature dependence of the maximum fluorescence position within the 331–334 nm interval. However, the temperature dependence of the I_{315}/I_{365} ratio reveals some changes for the **I** form at pH* 4.2. Its spectrum is slightly red shifted from the native form, has a steeper temperature dependence, and does not show a dependence break near 60 °C. The fluorescence λ_{max} of the **I** form at pH* 3.0 is the same as that of native state below 60 °C. Unfortunately, measurements of this form at higher temperatures were not possible because of increasing irreversibility.

Decreasing the salt concentration at pH* 3.0 results in a destabilization and expansion of the apoMb molecule (Griko et al., 1988; Nishii et al., 1995). As a result of this destabilization, a conformation transition shows up at a high temperature as can be readily observed in Figure 1. The transition temperature is correlated with salt concentration, being lower at lower salt concentration. For 0.02 M NaCl (the lowest salt concentration), the fluorescence λ_{max} nearly reaches the position typical of a completely solvated tryptophan side chain (346.5 nm in our measurements) at the highest temperatures. A protein destabilization not only lowers the midpoint of *heat* denaturation, it usually shifts the *cold* denaturation towards a higher temperature as well (Privalov et al., 1986). This is consistent with present results; the transition is observed at 0.02 M NaCl (pH* 3.0) below 20 °C. Traces of the cold denaturation transitions are also seen at 0.08 M NaCl and for the native state. In fact, both the heat- and cold-induced transitions strongly overlap each other at 0.02 M NaCl. Nevertheless, even with the transitions overlapping, it is clear that the tryptophan fluorescence λ_{max} at 10–30 °C is far below that of a completely solvated tryptophan residue. Therefore, for *all* the conditions studied herein, apoMb contains tryptophan residues that are considerably screened from solvent at room temperature.

The fluorescence intensity temperature dependence at 335 nm for the different apoMb forms is presented in Figure 2A. The results at other wavelengths are essentially similar and are not shown. The fluorescence intensity of the indole fluorophore in solution strongly changes with temperature. This dependence is generally found to be approximately linear for globular proteins (Eftink, 1994). To correct for this non-structural temperature dependence, we have interpolated the baseline temperature dependence by a straight line, and this was subtracted from the measured dependence (Fig. 2B). This simple procedure outlined the rather weak apoMb transitions that otherwise are masked considerably by the steep slope. The baseline regions for the low pH forms were determined from Figure 1. For pH* 3.0 they were 8–33 °C (0.02 M NaCl), 8–30 °C (0.08 M NaCl), −1–46 °C (0.15 M NaCl), −1–57 °C (0.25 M NaCl); for pH* 4.2: 9–47 °C (0.01 M Na acetate). For the native form (pH* 5.3, 0.01 M sodium cacodylate) the baseline was interpolated within the 18–30 °C range since this form of apoMb has another transition at a higher temperature that does not affect the tryptophan fluorescence λ_{max} (but does show up

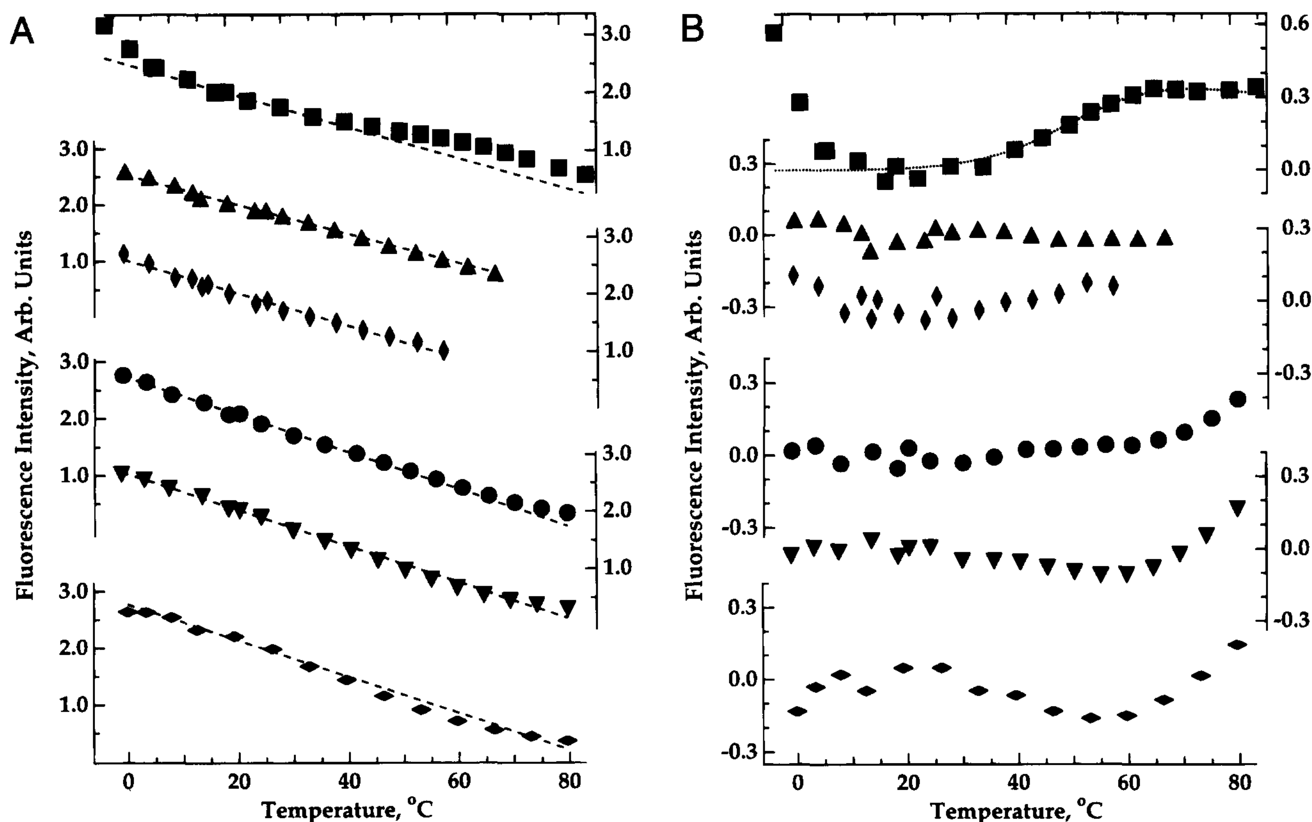

Fig. 2. Temperature dependence of apomyoglobin's intrinsic fluorescence intensity at 335 nm: **(A)** uncorrected data; **(B)** baseline subtracted. All data are scaled to the same intensity at 20 °C. Baselines are shown as dashed lines (see text for details). Dotted line shows a high temperature transition fit to the Equation (23) from Eftink (1994) with T_m = 55 °C. See Figure 1 legend for conditions and other symbol explanations.

in the IR measurements; Gilmanshin et al., 1997). A subtraction of a straight line will not introduce any sigmoidal pattern or sudden slope changes in the measured dependences.

The main difference between the temperature dependencies monitored by position of fluorescence λ_{max} and by intensity at a fixed wavelength is observed for the native form. While the fluorescence λ_{max} did not change its position until above 60 °C, the monitoring of fluorescence intensity at a fixed wavelength shows a cooperative transition with a midpoint at 56 ± 3 °C (averaged on fits of data at different wavelengths to the Equation (23) from Eftink, 1994) that is over at a temperature higher than 65 °C. This major denaturation transition, and at the same midpoint temperature, has also been found from circular dichroism (CD), calorimetry, intrinsic viscosity, and IR absorbance studies (Griko et al., 1988; Griko & Privalov, 1994; Nishii et al., 1994; Gilmanshin et al., 1997). The beginning of a cold denaturation (Griko et al., 1988; Nishii et al., 1994) for the native form is also obvious below 10 °C from the $I_{335}(T)$ dependence while it is hardly noticeable in the $I_{315}/I_{365}(T)$ dependence.

The same transitions are seen in both the temperature dependence for fluorescence λ_{max} and fluorescence intensity for the expanded form at pH* 3.0 (0.02 M NaCl). However, the midpoint temperature in this case cannot be determined accurately. The baseline of the $I_{335}(T)$ dependence is difficult to ascertain because of the strong overlapping of the cold and heat denaturation transitions. The amplitude of this high temperature transition is rather small, probably due to the partial compensation of the conformation transition effect by an opposite temperature dependence of fluorescence intensity of an exposed tryptophan. The salt concentration increase makes the transition amplitude even smaller. The dependence of $I_{335}(T)$ for each **I** form is well described by a straight line within experimental error.

Discussion

Terminology and other assumptions

The structural observations from many previous studies have suggested that apoMb may be stabilized in forms that closely resemble those on the folding pathway by varying chemical conditions (pH, salt, etc.). These forms have often been called "native" (pH 5), "molten globule" (pH 4, or lower pH at high salt), "expanded" or unfolded (low pH at low salt). The results of this study and previous work, as discussed below, clearly indicate that apoMb in all these conditions contains substructures or subdomains that behave independently in melting experiments of various sorts. These substructures are not evident in the crystallographic picture of the structure. That is, myoglobin is essentially a helical protein made up of eight separately identified helices (labeled A–H). The structure does not, for example, contain a structure like a semi-isolated "flap" that is connected to the rest of the protein by a hinge and that might be expected to exhibit thermodynamic behavior differ-

ent from that of the rest of the protein. Because myoglobin is a small protein and appears to have a single, homogeneous domain in crystal or NMR structures, many previous experimental investigations have been interpreted in just these terms. Our results demonstrate that some forms of apoMb are clearly heterogeneous, which thus requires a new way of viewing previous results and also attention to definitions. For example, while there are a number of definitions of "molten globule" structure, we take this to mean a generally collapsed state having the approximately correct secondary structure of the native protein. It is held together principally by hydrophobic forces, and melting of a local portion of this protein structure is possible. Molten globules generally do not show the cooperative melting behavior characteristic of closely packed native structures.

The characterization of each form of apoMb, in terms of conditions and some structural details, is summarized in the Table 1. We use the phrase "native state" (N) since, under the conditions that the protein is prepared in vitro, its structure is believed to be close to the species that is found in vivo. It exhibits all properties typical of densely packed "native-like" structure. The forms at lower pH in Table 1 are often referred to as molten globule-like. We avoid this so as not to suggest that this species is a homogeneous, hydrophobically-collapsed unit with no native-like substructures. Rather, we use the label **I** to refer to the pH 4 or **I′** to refer to the low pH, high salt species. It must be kept in mind that the **I** form likely is not an identical structure for various chemical conditions, although these structures appear to be very similar in many respects. Looking forward, it will be seen that the **I** form includes both molten-globule-like and native-like substructures. The label "**E**" is employed for the expanded species that exists at conditions (pH 3, low salt in our case) intermediate between the unfolded (**U**) and **I** forms. We emphasize here that this form is not unfolded but appears to be significantly more expanded than the **I** form(s). As it will be discussed below, it shares the same native-like substructure with the **I** form while the rest of molecule seems unordered. It therefore deserves its own designation. Finally, we use "**U**" for the unfolded state, which is reached at $1.5 < pH < 3$ in the absence of salt. This state can also be reached from either cold or heat denaturation of the **E** form. We shall use the word substructure to denote substantial portions of apoMb that behave apparently independently in melting behavior; the term subdomain is avoided because it is largely used in the literature to denote substructures found in large, multi-domain proteins.

In what follows, the current results are compared to previous studies of the various species. These studies include far- and near-UV CD, calorimetry, FT-IR, nanosecond kinetics studies, low-angle X-ray diffraction, and others. From all of this, a more complete picture of the structures of the various species emerges. The essential new conclusion is that all the folded or semi-folded species that we have studied contain an AGH core structure. The implications of the heterogeneous structure of apoMb on the folding pathway(s) are discussed. In view of this and other results, we ask the question whether or not other proteins of interest to the question of folding contain heterogeneous structures or if apoMb is anomalous.

The heterogeneous nature of the apomyoglobin I and I′ forms

Lowering the pH from neutral values converts apoMb to the **I** form near pH 4 (Griko et al., 1988). The **I** species is nearly as compact as the native state, has considerable secondary structure, and its tryptophan residues are screened from solvent (Griko et al., 1988; Hughson et al., 1990; Fink et al., 1991; Gast et al., 1994; Kataoka et al., 1995). **I′** exists at all lower pH values in the presence of an excess of salt (Fink et al., 1991). **I** at any salt concentration is very similar (by CD, fluorescence, and IR criteria) to **I′** existing at somewhat lower pH, (pH 3 in our measurements) and high salt concentration (Fink et al., 1991). The **I** and **I′** forms exhibit both

Table 1. *Comparison of parameters of different forms of apomyoglobin*[a]

ApoMb Form	Conditions	Compactness R_g (nm)	Relative Helicity Change[b] (from $[\Theta]_{222nm}$)	Fluorescence maximum λ_{max} (nm)
N	pH 6.5, 0.01 M Na Acetate	1.75[c]	1.00[e]	—
	pH 6.0, 0.01 M Hepes	1.97[d]		—
	pH* 5.3, 0.01 M Na Cacodylate	—		331.5
I	pH 4.0, 0.01 M Na Acetate	2.35[c]	0.65[e]	—
	pH* 4.2, 0.01 M Na Acetate	—	—	332.3
I′	pH 2.0, 0.25 M NaCl	—	0.44[f]	—
	pH* 3.0, 0.25 M NaCl	—	—	331.8
E	pH 2.0, 0.02 M NaCl	—	0.08[f]	—
	pH* 3.0, 0.02 M NaCl	—	—	335.3
U	pH 2.0 (no added salt)	≥4.67[c], 3.02[d]	0	346.5[g]

[a]Results obtained in H_2O, 20–25 °C.
[b]Change of helix content is relative to the total change in the **U** ⇔ **N** transition. The absolute helix content is about 55% for **N** and 5% for **U** from Nishii et al. (1995) as calculated from the size of the CD signal at 222 nm ($[\Theta]_{222nm}$).
[c]From small-angle X-ray scattering (Gast et al., 1994).
[d]From small-angle X-ray scattering (Kataoka et al., 1995).
[e]Estimated from ellipticity at 222 nm ($[\Theta]_{222nm}$) from Griko et al. (1988).
[f]Estimated from ellipticity at 222 nm ($[\Theta]_{222nm}$) from Nishii et al. (1995).
[g]Fluorescence maximum position of L-tryptophan in D_2O.

heat (>40 °C) and cold (<10 °C) denaturation transitions that result in substantial loss of secondary structure and swelling of the molecule (Griko et al., 1988; Gast et al., 1994; Nishii et al., 1995). These transitions are cooperative as determined by CD and scanning calorimetry measurements, although controversy exists about their order (Nishii et al., 1994, 1995; Griko & Privalov, 1994).

The structural differences between the **N** and **I** forms of apoMb do not manifest themselves in the current fluorescence measurements. The position of the fluorescence λ_{max} (332–333 nm) for the **I** and **I′** forms is practically the same as for **N** over the entire temperature range (Fig. 1). This means that the tryptophan residues in the **I** and **I′** forms are generally as well screened from solvent as in the **N** state. Screening of the tryptophans from solvent provides information on the structure of the A, G, and H helices because of their positions (Trp 7 and 14) close to the junctions of the A helix with the G and H helices (Evans & Brayer, 1988). The crystal structure of the holoprotein clearly shows the screening is provided by the backbone and side chains of all three helices. We therefore conclude that the A-helix remains in a native-like structure in I over an extended temperature range, even under conditions where the rest of molecule swells and loses substantial secondary structure. This conclusion is in agreement with a finding that the AGH helices form a core within the apoMb acid form (Hughson et al., 1990; Loh et al., 1995). Thus, the **I** forms of apoMb contain both a dense temperature-stable AGH core and loose helical segments that unwind at higher or lower temperatures from ~20–30 °C (compare our data with Gast et al, 1994; Sirangelo et al., 1994).

While no cooperative transitions were found within the measured temperature range for both **I** forms, the fluorescence λ_{max} of the **I** is a bit red shifted and shows some temperature dependence of its position compared to **I′**. This suggests slightly lower stability of this form in comparison with the salt-stabilized species at lower pH. The different stability of these two **I** forms could also explain why an excess heat absorption peak was found for a high salt-low pH form (Nishii et al., 1995) while no peak occurred for the pH 4 species (Griko & Privalov, 1994).

An important question is what characterizes the structure of the AGH core in **I**; is its structure rigid native-like or more flexible, molten globule-like. According to hydrogen-deuterium exchange, site-directed mutagenesis, and thermodynamic analysis of the denaturation transition, the AGH core has a very flexible molten globule-like structure, stabilized presumably by hydrophobic interactions with a low overall stability relative to extended structure (Hughson et al., 1990, 1991; Barrick & Baldwin, 1993; Loh et al., 1995). However, the observation that **I** contains substructures with different thermodynamic melting behavior means that the interpretations of previous data are difficult. The small size of the AGH core makes a majority of secondary structure hydrogen bonds exposed to solvent, which should strongly decrease protection factors. Moreover, the measurements are sometimes performed at subzero temperatures where **I** is considerably enlarged compared to the more stabilizing conditions at 30 °C (Gast et al., 1994). Finally, the CD in the far-UV region is often used to monitor the denaturation transitions, which is presumably sensitive to a secondary structure content. As we have seen, changes in secondary structure do not always correlate with the changes in tryptophan surroundings. Thus, the transition monitored with far-UV CD includes a transformation of a loose helical moiety within the **I** form, which masks any cooperative transition of the stable core (see also Gilmanshin et al., 1997).

We conclude that the **I** species AGH core has a native-like, rigid, stable structure for the following reasons. In the first place, the AGH core of the **I** form is stabilized by native-like tertiary interactions, at least partially (Kay & Baldwin, 1996). It can be denatured only by very strong agents such as several moles of guanidinium hydrochloride (Sirangelo et al., 1994) or urea (Jamin & Baldwin, 1996), or by high temperature in combination with low pH and small salt concentration (our data, see below). Also, **I** has considerable CD in the near UV, which suggests rigid, asymmetric surroundings of tryptophan residues (Gast et al., 1994). This near-UV CD exists both at 0 and 30 °C and decreases only at 80 °C for the pH 4 **I** species. For the **I′** form, a cooperative melting with an excess heat absorption peak was found by differential scanning calorimetry (Nishii et al., 1995) at temperatures corresponding to the transitions found by us with intrinsic fluorescence. These results are quite consistent with the existence of a native-like substructure. In this regard, the distance distribution function found in low angle X-ray scattering shows a substructure inside the pH 4 **I** species with a density close to native one (Gast et al., 1994). It should be noted, finally, that a native-like AGH core within the **I** forms provides a structural basis for the previously proposed combined equilibrium denaturation mechanism incorporating the two-state transition and a gradual structural change (Nishii et al., 1995), and also the biphasic kinetics found in the denaturation of an apoMb **I** form (Jamin & Baldwin, 1996).

The E state—Intermediate between I and U

At $1.5 < pH < 3$ in the absence of salt, apoMb is highly unfolded according to different parameters that are similar to those of the unfolded state in the presence of high concentrations of strong denaturants such as urea or guanidinium hydrochloride (Table 1). It has no signs of tertiary structure, only 5% of secondary structure, no cooperative melting, completely exposed tryptophan residues, and a volume similar to that of the completely unfolded state (Griko et al., 1988; Fink et al., 1991; Gast et al., 1994; Griko & Privalov, 1994; Nishii et al., 1994; Kataoka et al., 1995).

The lowest salt concentration (0.02 M NaCl) that we have used in our study at pH* 3 (the **E** state) is very close to the minimal possible concentration of Cl^- ions. At this pH and salt concentration, apoMb is considered to be close to, if not completely, the unfolded **U** state, like that found at pH 2. Its helicity is less than 20% of that of found in the **I′** form (Fink et al., 1991; Nishii et al., 1995). However, the maximum position of the fluorescence spectrum (Table 1) lies near 335 nm, which is much closer to that found in the native state (331 nm) than to that of unfolded one (346.5 nm for a solubilized tryptophan). We therefore conclude that the structure of the **E** species is much more unfolded than that found for the **I** state(s) but still must retain some folded structure, which includes the tryptophans of the A helix. It is for this reason that we have given this clearly extended, but still (partially) folded species a separate name.

An **E**-like state probably also exists at intermediate pH in the pH-induced **U** ⇔ **N** transition in the absence of additional salt because the transition curves monitored by far-UV CD and λ_{max} of tryptophan fluorescence (Fink et al., 1991) are not superimposable. The λ_{max}-monitored transition starts at pH 2.2 and is over at $pH < 4$, reaching the native value. On the contrary, the CD-monitored transition starts at pH 3 and reaches only 2/3 of its native value at pH 4.

The fluorescence λ_{max} results of Figure 1 also suggest that native-like contacts are present in the **E** state since the melting curve of the **E** species is clearly sigmoidal. In addition, sigmoidal transitions, but with higher T_ms, are evident at higher salt concentrations. These transitions have also been observed from differential scanning calorimetry measurements and are found at similar temperatures (Nishii et al., 1995). Unfortunately, we could not reach the melting temperature at 0.25 M NaCl (pH* 3.0) as further heating results in progressing irreversibility.

It seems clear from all the data that the **E** form shares at least some of the native-like substructure with the **I** form while the rest of the structure of **E** is essentially disordered.

The heterogeneous nature of native (N) apomyoglobin

Our current results, when combined with previous studies, show that **N** includes substructures and is heterogeneous. It has been shown previously that the native species of apoMb has two major denaturation transitions at low (below 10 °C) and high (with T_m at 55–60 °C) temperatures. These are highly cooperative, two-state transitions and have been observed by CD, differential scanning calorimetry, intrinsic viscosity, IR absorbance in the amide I band region (Griko et al., 1988; Griko & Privalov, 1994; Gilmanshin et al., 1997), and our current results on the intrinsic fluorescence intensity (see Fig. 2). On the other hand, there is *no sign* of these transitions from the dependence of the tryptophan fluorescence λ_{max} (Fig. 1). Thus, the surroundings of the two tryptophans within the A helix are nonpolar throughout the low (below 10 °C) and high (~55–60 °C) temperature denaturation transitions. The fluorescence λ_{max} only begins to show melting behavior at very high temperatures. Therefore, since our current results show that the AGH core is intact at 60 °C, other parts of apoMb are involved in this transition.

An additional native-like substructure, apart from the AGH core, also follows from results obtained by De Sanctis et al. (1994) who studied apo-minimyoglobin, a truncated version of apoMb with no AGH core since the A helix and a large part of the H helix have been cut off. The truncated protein exhibits equilibrium intermediates analogous to the **N**, **I**, and **U(E)** species of apoMb and pH-induced transitions at approximately the same conditions as the whole protein. The second substructure is, therefore, the one that undergoes a transition from molten globule to a native-like structure within a pH increase from 4 to neutral values. It also accounts for the additional 1/3 of apoMb helicity in this transition. In addition, native apoMb contains a portion of loose helices whose behavior is the same as that found in short helical peptides both thermodynamically and dynamically (Eliezer & Wright, 1996; Fontana et al., 1997; Gilmanshin et al., 1997). We thus conclude that the native form of apoMb contains three substructures: two native-like portions, which have cooperative melting transitions but with different stabilities, and a loose helical portion.

The "AGH core" may not include the complete A, G, and H helices and almost certainly differs for the **N**, **I**, and **E** forms. Minimally, the core probably includes the regions adjacent to the amino acid side chains that form the interhelical cluster of the A, G, and H helices and the GH hairpin turn as observed in the crystal structure of the holoprotein. This is supported by the fact that protection factors of the peptide protons are greatest for the peptide groups nearest to the interhelical cluster (Loh et al., 1995) and by the exceptional stability of the isolated GH turn (Shin et al., 1993). The native-like substructure other than AGH core probably includes the helices from the interface between the AGH core and the heme pocket. These are helices B, C, D, and E. The assignment is supported by the following. First, helices B, C, and E include peptide protons that are protected against hydrogen/deuteron exchange in native apoMb (Hughson et al., 1990; Johnson & Walsh, 1994). Second, helix B is stabilized by the addition of trichloroacetate even at low pH (Loh et al., 1995). Third, helices B–D have been observed by NMR in native apoMb (Cocco & Lecomte, 1994; Lecomte et al., 1996). It is reasonable to suppose that the solvated flexible helical portion of native apoMb are from the heme pocket surroundings; these have no major hydrophobic stabilization in the heme-free apoprotein. This includes the F helix and the ends of the E and H helices (Eliezer & Wright, 1996).

Implications for the folding pathway of apoMb

A number of experiments, particularly hydrogen exchange and CD kinetic experiments (cf., Jennings & Wright, 1993), point toward the formation of helices or nascent helical substructures and then the association of these helical structures, with the AGH core formed first followed by the association of the B, C, D, and E helices leading to the formation of the native protein. Important issues within this scheme have to do with the nature of the helical structures, how they are stabilized, and the timing of the events. NMR, CD, and hydrogen–deuteron exchange experiments suggest that portions of the protein can form in isolation with some population, particularly the H helix (Waltho et al., 1993) and GH turn (Shin et al., 1993). In the burst phase of these studies (less than a millisecond), formation of an early species is postulated to involve the association of an A-G-H helical core stabilized within a compact molten globule, which closely resembles the structure of the **I** form studied in equilibrium experiments.

The studies here and elsewhere modify and augment the picture of the folding pathway in two important aspects. In the first place, the structure of **I** is heterogeneous, with an AGH core containing native-like contacts, while the rest of the molecule consists of loosely associated but collapsed, solvated helices as well as an unstructured portion. In fact, the more destabilized **E** form appears to contain only a small native-like AGH core embedded within an otherwise unstructured molecule. In addition, there is a second substructure of apoMb that can be melted separately from the AGH core (De Sanctis et al., 1994; Gilmanshin et al., 1997; present results). Therefore, the folded protein, **N**, contains these two native-like substructures and a loosely associated helical substructure, which is collapsed and molten globule-like (Gilmanshin et al., 1997).

These considerations and previous results suggest that the folding pathway of apoMb may be described as the occurrence of two "nucleation" events. The first event puts together the two ends of apoMb of polypeptide chain and forms the very stable AGH "core," driven by the known propensity to form stable secondary structure (turns and helices) in isolation of parts of the polypeptide sequence that comprises this core (Shin et al., 1993; Waltho et al., 1993). Thus, the two ends of the chain are fixed very early in the folding process. This greatly reduces the conformational space available for further folding events. The second nucleation event is the formation of a second native-like substructure, likely involving portions of the B, C, D, and E helices as discussed above (the BCDE complex). We have found a helical structure forms on the time scale of tens of nanoseconds and is thus a very fast process compared to other folding events (Williams et al., 1996; Gilmanshin

et al., 1997). Thus, helical structure is present well in advance of any other association. The formation time of the AGH core can be predicted to occur between tens of microseconds (as would be deduced if simple diffusion controlled the formation time; Karplus & Weaver, 1994) and less than a few milliseconds (the AGH core is intact at 5 °C within the burst phase of stopped flow experiments; Jennings & Wright, 1993). It is possible that the 250 μs transient at 5 °C observed in recent T-jump relaxation experiments of apomyoglobin involves the formation of the AGH core (Ballew et al., 1996). The formation time of the BCDE complex is much slower, about 1–10 s at ~5 °C (Jennings & Wright, 1993) and about 280 μs at much higher temperature (60 °C; Gilmanshin et al., 1997) .

Is the folding pathway better characterized by the quasi-independent folding of apoMb substructures as opposed to a sequential picture as commonly assumed? It is very clear that the AGH core forms very early in the folding pathway and this association proceeds independently from the formation of the BCDE complex or any other structure apart from helix formation. Specifically, hydrophobic collapse of a substantial portion of apoMb is not necessary to the formation of the AGH core. However, it is not clear whether BCDE complex formation is dependent upon the presence of the AGH core. In general, it might be expected that tying the two ends of the protein together would change the helical propensities and the stabilities of substructures in between. However, the thermodynamic studies of De Sanctis et al. (1994) on apo-minimyoglobin, which showed transitions similar to those of apoMb despite the loss of the AGH core, suggest that BCDE complex forms largely independently from the AGH core. This important mechanistic issue requires further study.

Is apomyoglobin's heterogeneity unique?

An important question is the generality of the structural heterogeneity found in apoMb. Is apoMb a rare exception or are there other examples that might make this issue important for the folding problem? Of course, this issue requires specific study. Recently, Englander and his colleagues have inferred from hydrogen/deuterium exchange studies on cytochrome *c* that this protein contains several substructures (Bai & Englander, 1996), very similar in nature to the present results on myoglobin. In model systems, it has also been demonstrated that coexistence of different cooperative units does not necessarily require their spatial separation into lobes (Tiktopulo et al., 1995). Finally, many studies have found chemical conditions that destabilize the structure of a protein to fit the definition of a molten globule (increase of the protein's volume, loss of fixed tertiary contacts, etc.) while retaining enough ordered structure so that at least part of the protein's structure can be characterized by native-like NMR spectra of some part and/or exhibit a two-state denaturation transition. These species have been called "highly ordered" molten globules (see Ptitsyn (1995) for references). These highly ordered molten globules probably include a native-like core comprised of tightly packed side chains that unfold cooperatively (Shakhnovich & Finkelstein, 1989).

Materials and methods

ApoMb was prepared from horse heart myoglobin (Sigma) by a modified 2-butanone extraction (Rothgeb & Gurd, 1978; Teale, 1959), which was specially adjusted for the horse protein. After the heme extraction, the protein solution was dialyzed exhaustively against 0.01 M HCl and then against 0.01 M sodium phosphate buffer with 0.04 M NaCl, pH 6.0. The dialyzed protein solution was clarified by centrifugation and subjected to gel filtration on Sephadex G-75 (Pharmacia Biotech). Monomeric apoMb fractions were extensively dialyzed against deionized water and lyophilized. All these operations were performed at 4 °C. Final sample contained less than 0.3% of holomyoglobin. It was completely homogeneous on SDS polyacrylamide gel electrophoresis. Protein concentration was determined by absorption at 280 nm with extinction coefficient $E_{1\%} = 8.4\ cm^{-1}$ (Crumpton & Polson, 1965). Samples with concentrations of 0.12–0.06 mg/mL were used.

All solutions were prepared with D_2O. The presented pH* values are uncorrected pH-meter readings recorded at 20 °C. All chemicals were purchased from Sigma and were of the highest available purity.

Temperature was controlled with a Model 1167 bath circulator (VWR Scientific). It was measured directly in the cuvette using BAT-12 thermometer with a type IT-18 thermocouple microprobe (Physitemp Instruments, Inc.).

Fluorescence spectra were measured on FluoroMax-2 spectrofluorimeter (Instruments S. A., Inc.) with correction for a spectral dependence of registration response. Excitation wavelength was 290 nm. Band pass was 2–3 nm for excitation and 2 nm for emission. A 1-cm fused silica cuvette was used. The data were collected with DM-3000 software (Instruments S. A.) and processed with IgorPro (WaveMetrics, Inc.) software. Solvent background was measured at the same conditions and subtracted from every protein fluorescence spectrum.

Acknowledgments

We are indebted to Skip Williams and William H. Woodruff for stimulating and thoughtful discussions. The support of the National Science Foundation, grant MCB-9417892 to R.H.C., and the National Institutes of Health, grants GM53640 to R.B.D. and GM35183 to R.H.C., is gratefully acknowledged.

References

Bai Y, Englander SW. 1996. Future directions in folding: The multi-state nature of protein structure. *Proteins Struct Funct Genet 24*:145–151.

Barrick D, Baldwin RL. 1993. The molten globule intermediate of apomyoglobin and the process of protein folding. *Protein Sci 2*:869–876.

Ballew RM, Sabelko J, Gruebele M. 1996. Observation of distinct nanosecond and microsecond protein folding events. *Nature Struc Biol 3*:923–926.

Cocco MJ, Lecomte JTJ. 1994. The native state of apomyoglobin described by proton NMR spectroscopy: Interaction with the paramagnetic probe HyTEMPO and the fluorescent dye ANS. *Protein Sci 3*:267–281.

Crumpton MJ, Polson A. 1965. A comparison of the conformation of sperm whale metmyoglobin with that of apomyoglobin. *J Mol Biol 11*:722–729.

De Sanctis G, Ascoli F, Brunori M. 1994. Folding of apominimyoglobin. *Proc Natl Acad Sci USA 91*:11507–11511.

Dyer RB, Williams S, Woodruff WH, Gilmanshin R, Callender RH. 1996. Fast events in protein folding: Helix dynamics in proteins and model peptides. *Biophys J 70*:A177.

Eftink MR. 1994. The use of fluorescence methods to monitor unfolding transitions in proteins. *Biophys J 66*:482–501.

Eliezer D, Jennings PA, Wright PE, Doniach S, Hodgson KO, Tsuruta H. 1995. The radius of gyration of an apomyoglobin folding intermediate. *Science 270*:487–488.

Eliezer D, Wright PE. 1996. Is Apomyoglobin a molten glogule? Structural characterization by NMR. *J Mol Biol 263*:531–538.

Evans SV, Brayer GD. 1988. Horse heart metmyoglobin. A 2.8-Å resolution three-dimensional structure determination. *J Biol Chem 263*:4263–4268.

Fink AL, Calciano LJ, Goto Y, Palleros D. 1991. Conformation states in acid-denatured proteins. In: Nall BT, Dill KA, eds. *Conformations and forces in*

protein folding. Washington, D.C.: American Association for the Advancement of Science. pp 169–174.

Fontana A, Zambonin M, DeLaureto PP, DeFilippis V, Clementi A, Scaramella E. 1997. Probing the conformational state of apomyoglobin by limited proteolysis. *J Mol Biol 266*:223–230.

Gast K, Damaschun H, Misselwitz R, Müller-Frohne M, Zirwer D, Damaschun G. 1994. Compactness of protein molten globules: Temperature-induced structural changes of the apomyoglobin folding intermediate. *Eur Biophys J 23*:297–305.

Gilmanshin R, Williams S, Callender RH, Woodruff WH, Dyer RB. 1997. Fast events in protein folding: Relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc Natl Acad Sci USA 94*:3709–3713.

Griko YV, Privalov PL. 1994. Thermodynamic puzzle of apomyoglobin unfolding. *J Mol Biol 235*:1318–1325.

Griko YV, Privalov PL, Venyaminov SY, Kutyshenko VP. 1988. Thermodynamic study of the apomyoglobin structure. *J Mol Biol 202*:127–138.

Hughson FM, Barrick D, Baldwin RL. 1991. Probing the stability of a partly folded apomyoglobin intermediate by site-directed mutagenesis. *Biochemistry 30*:4113–4118.

Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science 249*:1544–1548.

Jamin M, Baldwin RL. 1996. Refolding and unfolding kinetics of the equilibrium folding intermediate of apomyoglobin. *Nature Struct Biol 3*:613–618.

Jennings PA, Wright PE. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science 262*:892–896.

Johnson RS, Walsh KA. 1994. Mass spectrometric measurement of protein amide hydrogen exchange rates of apo- and holo-myoglobin. *Protein Sci 3*:2411–2418.

Karplus M, Weaver DL. 1994. Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci 3*:650–668.

Kataoka M, Nishii I, Fujisawa T, Ueki T, Tokunaga F, Goto Y. 1995. Structural characterization of the molten globule and native states of apomyoglobin by solution X-ray scattering. *J Mol Biol 249*:215–228.

Kay MS, Baldwin RL. 1996. Packing interactions in the apomyoglobin folding intermediate. *Nature Struct Biol 3*:439–445.

Kirby EP, Steiner RF. 1970. The influence of solvent and temperature upon the fluorescence of indole derivatives. *J Phys Chem 74*:4480–4490.

Lecomte JTJ, Kao YH, Cocco MJ. 1996. The native state of apomyoglobin described by proton NMR spectroscopy: The A-B-G-H interface of wild-type sperm whale apomyoglobin. *Proteins Struct Funct Genet 25*:267–285.

Loh SN, Kay MS, Baldwin RL. 1995. Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc Natl Acad Sci USA 92*:5446–5450.

Makhatadze GI, Clore GM, Gronenborn AM. 1995. Solvent isotope effect and protein stability. *Nature Struct Biol 2*:852–855.

Nishii I, Kataoka M, Goto Y. 1995. Thermodynamic stability of the molten globule states of apomyoglobin. *J Mol Biol 250*:223–238.

Nishii I, Kataoka M, Tokunaga F, Goto Y. 1994. Cold denaturation of the molten globule states of apomyoglobin and a profile for protein folding. *Biochemistry 33*:4903–4909.

Postnikova GB, Komarov YE, Yumakova EM. 1991. Fluorescence study of the conformational properties of myoglobin structure 1. pH-dependent changes of tryptophanyl fluorescence in intact and chemically modified sperm whale myoglobins. *Eur J Biochem 198*:223–232.

Privalov PL, Griko YV, Venyaminov SY, Kutyshenko VP. 1986. Cold denaturation of myoglobin. *J Mol Biol 190*:487–498.

Ptitsyn OB. 1995. Structures of folding intermediates. *Curr Opinion Struct Biol 5*:74–78.

Rischel C, Thyberg P, Rigler R, Poulsen FM. 1996. Time-resolved fluorescence studies of the molten globule state of apomyoglobin. *J Mol Biol 257*:877–885.

Rothgeb TM, Gurd FRN. 1978. Physical methods for the study of myoglobin. *Methods Enzymol 52*:473–486.

Schmid FX. 1992. Kinetics of unfolding and refolding of single-domain proteins. In: Creighton TE, ed. *Protein folding*. New York: W. H. Freeman & Co. pp 197–241.

Shakhnovich EI, Finkelstein AV. 1989. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers 28*:1667–1680.

Shin H, Merutka G, Waltho JP, Tennant LL, Dyson HJ, Wright PE. 1993. Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. *Biochemistry 32*:6356–6364.

Sirangelo I, Bismuto E, Irace G. 1994. Solvent and thermal denaturation of the acidic compact state of apomyoglobin. *FEBS Lett 338*:11–15.

Teale FWJ. 1959. Cleavage of the haem–protein link by acid methylethylketone. *Biochim Biophys Acta 35*:543.

Tiktopulo EI, Uversky VN, Lushchik VB, Klenin SI, Bychkova VE, Ptitsyn OB. 1995. "Domain" coil-globule transition in homopolymers. *Macromolecules 28*:7519–7524.

Waltho JP, Feher VA, Merutka G, Dyson HJ, Wright PE. 1993. Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. *Biochemistry 32*:6337–6347.

Williams S, Causgrove TP, Gilmanshin R, Fang KS, Callender RH, Woodruff WH, Dyer RB. 1996. Fast events in protein folding: Helix melting and formation in a small peptide. *Biochemistry 35*:691–697.