# Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements

PAOLA PICOTTI, ANNA MARABOTTI,[1] ALESSANDRO NEGRO, VALERIA MUSI, BARBARA SPOLAORE, MARCELLO ZAMBONIN, AND ANGELO FONTANA

CRIBI Biotechnology Centre, University of Padua, Viale G. Colombo 3, I-35121 Padua, Italy

(RECEIVED January 19, 2004; FINAL REVISION March 8, 2004; ACCEPTED March 9, 2004)

**Abstract**

The conformational features of native and mutant forms of sperm-whale apomyoglobin (apoMb) at neutral pH were probed by limited proteolysis experiments utilizing up to eight proteases of different substrate specificities. It was shown that all proteases selectively cleave apoMb at the level of chain segment 82–94 (HEAELKPLAQSHA), encompassing helix F in the X-ray structure of the holo form of the native protein; for example, thermolysin cleaves the Pro 88–Leu 89 peptide bond. These results indicate that helix F is highly flexible or largely disrupted in apoMb. Because helix F contains the helix-breaking Pro 88 residue, we propose that helix F is kept in place in the native holo protein by a variety of helix–heme stabilizing interactions. To modulate the stability of helix F, the Pro88Ala and Pro88Gly mutants were prepared by site-directed mutagenesis, and their conformational properties investigated by both far-UV circular dichroism spectroscopy and limited proteolysis. The helix content of the Pro88Ala mutant was somewhat enhanced with respect to that of both native and Pro88Gly mutant, as expected from the fact that Ala is the strongest helix inducer among the 20 amino acid residues. The rate of limited proteolysis of the three apoMb variants by thermolysin and proteinase K was in the order native > Pro88Gly >> Pro88Ala, in agreement with the scale of helix propensity of Ala, Gly, and Pro. The possible role of the flexible/unfolded chain segment 82–94 for the function and fate of apoMb at the cellular level is discussed.

**Keywords:** apomyoglobin; circular dichroism; limited proteolysis; mass spectrometry; protein dynamics; protein engineering

**Supplemental material:** see www.proteinscience.org

Reprint requests to: Angelo Fontana, CRIBI Biotechnology Centre, University of Padua, Viale G. Colombo 3, I-35121 Padua, Italy; e-mail: angelo.fontana@unipd.it; fax: 39-049-8276159.

[1]Present address: Laboratory of Bioinformatics, Institute of Food Science, National Council of Research, Avellino, Italy.

*Abbreviations:* CD, circular dichroism; apoMb, apomyoglobin, that is, heme-free myoglobin; P88G and P88A, mutant apoMb with Pro substituted with Gly or Ala, respectively, at position 88 of the 153-residue chain of the protein; MG, molten globule; 3D, three-dimensional; EDTA, ethylenediaminetetraacetic acid; HPLC, high-performance liquid chromatography; NMR, nuclear magnetic resonance; TFA, trifluoroacetic acid; Tris, tris(hydroxymethyl)aminomethane; E:S, enzyme to substrate ratio; K, proteinase K; Su, subtilisin; Th, thermolysin; Ch, chymotrypsin; V8, protease V8 from *S. aureus*; T, trypsin; P, papain; E, elastase; [θ], mean residue ellipticity; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate; RP, reverse-phase; UV, ultraviolet; MS, mass spectrometry; ESI, electrospray-ionization.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.04635304.

The first protein structure determined by X-ray crystallography was that of sperm-whale myoglobin (Kendrew et al. 1958), shown to consist of eight α-helices (A–H; Takano 1977a,b; Kuriyan et al. 1986; see Fig. 1). The structure of the heme-free form of myoglobin (apomyoglobin, apoMb) is not yet determined by X-ray methods, and only recently, its structure was solved by two-dimensional NMR (Eliezer and Wright 1996; Lecomte et al. 1996, 1999; Eliezer et al. 1998; Cavagnero et al. 2000, 2001). Moreover, the molecular and conformational features of the partly folded or molten globule (MG) states adopted by apoMb at low pH have been analyzed by a plethora of biochemical and biophysical techniques, most notably by NMR (Goto and Fink 1990; Hughson et al. 1990; Barrick and Baldwin 1993; Kataoka et al. 1995; Fink et al. 1997; Eliezer et al. 1998; Cavagnero et

 

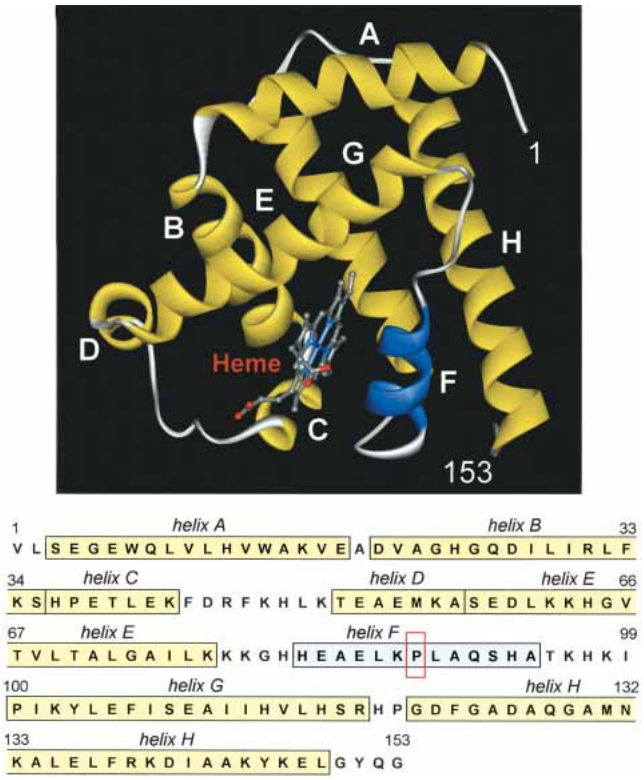


**Figure 1.** Three-dimensional structure (*top*) and amino acid sequence (one-letter code for amino acid residues) (*bottom*) of sperm-whale myoglobin. The colored boxes indicate the eight helical segments (*bottom*). Pro 88 residue is boxed in red. The model was constructed from the X-ray structure of myoglobin (PDB file 104M) taken from the Brookhaven Protein Data Bank (Bernstein et al. 1977) utilizing the program WebLab (Molecular Simulations Inc.).

al. 2000, 2001; Yao et al. 2001), making sperm-whale apoMb the most well-characterized protein in terms of MG state(s) (Arai and Kuwajima 2000; Gilmanshin et al. 2001; Tcherkasskaya and Uversky 2001) and folding mechanism (Jennings and Wright 1993; Eliezer et al. 1998; Nishimura et al. 2002).

It has been proposed that the dominant folding pathway of apoMb is U→AGH→ABGH →ABCDEGH→N, where U and N are the unfolded and native state, respectively, and A, B, C, etc., the helical segments of the native protein (Eliezer et al. 1998; see Nishimura et al. 2002 for a recent study). This folding scheme implies that helix F (residues 82–94)[2] is not formed in apoMb at neutral pH. Indeed, that helix F is disordered in apoMb was earlier proposed on the basis of molecular dynamics simulations conducted on sperm-whale apoMb (Brooks 1992; Tirado-Rives and Jorgensen 1993; Hirst and Brooks 1994, 1995) and subsequently confirmed by NMR measurements (Eliezer and Wright 1996). Actually, the conformational features of apoMb at neutral pH were considered like those predicted for the MG state (Lin et al. 1994a), that is, a dynamic state retaining the overall fold and the secondary structure of the native protein, but lacking its specific tertiary interactions (Ptitsyn 1995; Arai and Kuwajima 2000). However, NMR data provided evidence that apoMb is a folded globular protein much similar to the holo protein, with the major difference being that helix F is disordered (Eliezer and Wright 1996; Lecomte et al. 1996, 1999; Eliezer et al. 1998).

In a previous study, we analyzed the conformational features of horse apoMb by the limited proteolysis approach (Fontana et al. 1997a,b). The rationale of using proteolytic probes of protein structure and dynamics relies on the fact that the proteolytic event requires binding and adaptation of the polypeptide substrate to the specific stereochemistry of the protease's active site. It has been demonstrated that the sites of limited proteolysis in a globular protein are characterized by enhanced backbone flexibility, as there is a clear-cut correlation between sites of limited proteolysis and sites displaying high *B*-factors, these last determined by X-ray crystallography (Fontana et al. 1986). On this basis, it was proposed that limited proteolysis experiments can be used to identify, in a globular protein, the sites of local unfolding. When the simple biochemical approach of limited proteolysis was applied to horse apoMb, it was shown that proteases selectively cleave the protein at the level of the chain segment encompassing helix F of the native holo protein, thus demonstrating that this segment is highly flexible or largely disrupted in horse apoMb (Fontana et al. 1997a,b).

The aim of this study is to probe the molecular features of sperm-whale apoMb by limited proteolysis utilizing a large number of proteolytic enzymes of different substrate specificities. The horse and sperm-whale holo proteins show similarity of amino acid sequence (82%) and overall 3D-structure (Takano 1977a,b; Kuriyan et al. 1986; Evans and Brayer 1990). Nevertheless, it seemed to us relevant to perform proteolysis on the most studied sperm-whale protein species, in order to take full advantage of the wealth of information available for this protein (see above). Nowadays, it is recognized that the molecular features of homologous proteins are expected to be similar and, therefore, experimental results and their interpretations are often used interchangeably within a set of homologous proteins. However, this can be an oversimplification, as numerous protein engineering experiments have highlighted that even a single amino acid replacement can eventually cause significant perturbation of protein structure and function. In fact, recently it has been shown that two amino acid replacements (N132G and E136G) in sperm-whale apoMb lead to an

[2] In this study, helix F in the holo form of sperm-whale myoglobin is considered to comprise residues 82–94, at variance from chain segment 86–94 of other studies (e.g. Lecomte et al. 1996, 1999). However, the exact location of amino- and carboxy-terminal residues in helical segments in protein structures cannot be unequivocally defined. In the X-ray structure of the highly homologous horse myoglobin, helix F comprises residues 82–97 (Evans and Brayer 1990).

altered folding pathway of the protein (Cavagnero et al. 2001). In the present study, besides analyzing the conformational features of wild-type apoMb, we replaced (by recombinant methods) the helix-breaker Pro 88 residue, located at the chain segment encompassing helix F, with Gly or Ala residues. The P88G and P88A mutants were then studied by circular dichroism (CD) spectroscopy and limited proteolysis experiments. The expectation was to be able to modulate the conformational equilibria of helix F in apoMb and to monitor them by the proteolysis approach. The results of this study were positive, and indeed the rates of proteolysis of the apoMb variants were those expected from the scale of helix propensities of Pro, Gly, and Ala residues. It is concluded that the highly conserved Pro 88 residue is responsible for helix F unfolding in apoMb.[3]

## Results

### *Proteolysis of wild-type apoMb*

In this study, the approach of limited proteolysis for unraveling features of protein structure and dynamics was used with two goals. First, we aimed to analyze the conformational features of wild-type sperm-whale apoMb at neutral pH. Second, we were interested in modulating the structural integrity of helix F by introducing amino acid replacements at the chain segment encompassing helix F of the native holo protein. The idea was to possibly monitor the local conformational features of a chain segment embedded in an otherwise folded and rigid protein structure. To this aim, the two variants P88G and P88A were produced by recombinant methods and their apo forms obtained by removing the protein-bound heme moiety by 2-butanone extraction (Teale 1959).

First, we have demonstrated that the holo form of sperm-whale myoglobin is fully resistant to proteolysis with thermolysin when reacted under experimental conditions identical to those subsequently used for apoMb. The holo protein is resistant to thermolysin even after a 24-h reaction at room temperature (see Fig. S1, Supplemental Material). Proteolysis of wild-type apoMb was carried out utilizing an E/S ratio of 1/100 at 25°C with the proteolytic enzymes proteinase K, thermolysin, chymotrypsin, subtilisin, elastase, papain, V8-protease, and trypsin. The pattern of proteolysis was monitored by SDS-PAGE and RP-HPLC. The identity of protein fragments was established by electrospray ionization (ESI) mass spectrometry (MS) and amino-terminal sequencing. It was shown that all proteases produce quite clean protein fragmentation patterns, because few and relatively large fragments of apoMb are initially formed. Identification of these fragments enabled us to locate the sites of limited proteolysis along the 153-residue chain of apoMb.

The results of a typical limited proteolysis experiment conducted with proteinase K on apoMb are shown in Figure 2. The SDS-PAGE analysis shows that apoMb is rapidly cleaved by the voracious and unspecific proteinase K, and after a 1-min reaction, two principal electrophoretic bands are seen in the stained gel (Fig. 2A). The initial products of proteolysis are then further digested to low molecular weight peptides, poorly stained in the gel. The RP-HPLC

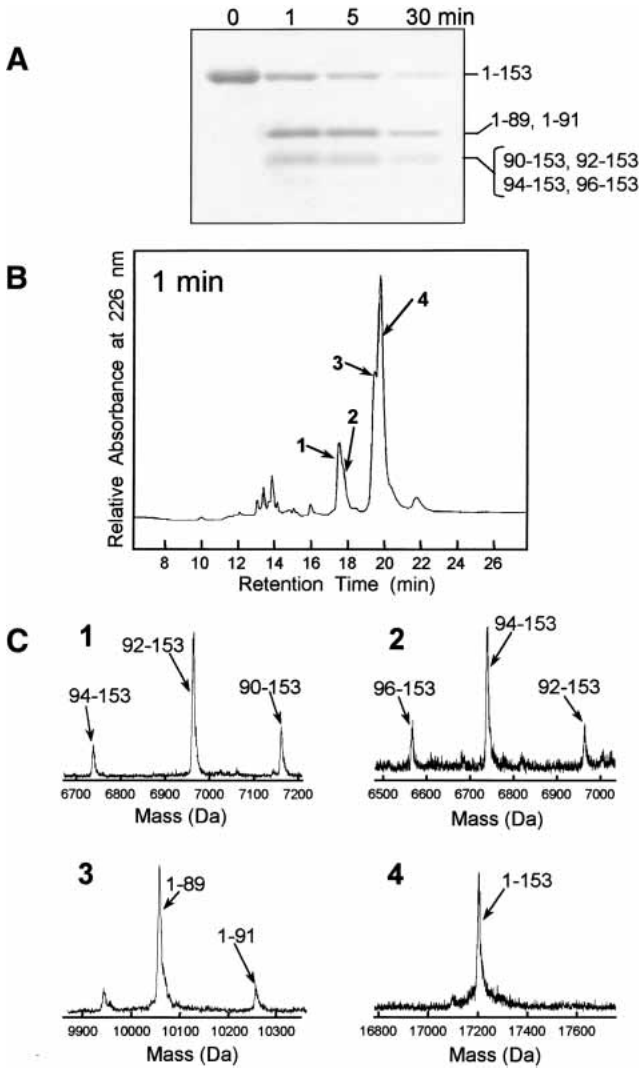


**Figure 2.** Limited proteolysis of sperm-whale apoMb with proteinase K. Proteolysis was conducted at 25°C (E/S 1:100, by weight) in 20 mM Tris·HCl, 0.15 M NaCl (pH 7.7). (*A*) Time course analysis by SDS-PAGE of the proteolysis mixture of apoMb with proteinase K after a 1-, 5-, and 30-min reaction. (*B*) RP-HPLC of an aliquot of apoMb reacted with proteinase K after a 1-min incubation. (*C*) Analysis by ESI–MS of peptide material contained in the RP-HPLC samples (1–4) of part *B*. The identification of apoMb and protein fragments is given in the deconvoluted ESI–MS spectra (1–4).

[3]This work was presented at the Fifth European Symposium of the Protein Society, Florence, Italy, March 29–April 2, 2003. (*Protein Sci.* **Suppl. 1** : Commun. 118.)

chromatogram of the proteolysis mixture, after a 1-min reaction, shows major peaks of peptide material that were sampled as indicated in Figure 2B and then analyzed by ESI–MS. It was shown that the isolated peptide samples corresponded to intact apoMb (fraction 4), a mixture of amino-terminal fragments 1–89 and 1–91 (fraction 3) and a mixture of the carboxy-terminal fragments 90–153, 92–153, 94–153, and 96–153 (fractions 1 and 2; see Table 1). Therefore, proteinase K initially cleaves apoMb at several peptide bonds, but all are located at a chain region encompassing helix F in the native holo protein (see Figs. 1, 3).

Similar results of site-specific proteolysis of apoMb were obtained with a variety of other proteases. The sites of cleavage of apoMb were identified in each case, and the results of these analyses are summarized in Table 1 and Figure 3. Some of these results, in the form of SDS-PAGE gels and RP-HPLC chromatograms, are also reported in the Supplemental Material. Briefly, it was shown that thermolysin selectively cleaves apoMb, after a 1-min reaction, the Pro 88–Leu 89 peptide bond; note that thermolysin cleaves at the amino terminus of mostly hydrophobic residues. Protease V8 cleaves apoMb at peptide bond Glu 85–Leu 86, chymotrypsin at Leu 89–Ala 90, whereas subtilisin and elastase cleave at Leu 89–Ala 90 and Gln 91–Ser92. Trypsin cleaves at Lys 96–His 97 and not at Lys 87–Pro 88, as the Pro residue hinders tryptic hydrolysis. In the proteolysis mixture of apoMb with papain, after a 1-h reaction, fragments 1–90, 91–153, 92–153, 97–153, and 90–153 were identified. In general, it is seen that the rather specific proteases (e.g., trypsin, V8-protease) cleave at one or very few peptide bonds, whereas the nonspecific proteases (e.g., subtilisin, proteinase K, papain) cleave at several peptide bonds, but all nearby to each other and essentially located at chain segment 82–94 encompassing helix F in the native holo protein (see Fig. 3 and Table 1). It may be noted here that the proteolysis by V8-protease and trypsin occurring at Glu 85 and Lys 96, respectively, is much slower than that observed with other less specific proteases (see Figs. S4 and S5, Supplemental material). Indeed, proteolysis occurs in hours or minutes with the specific or nonspecific proteases, respectively. We are inclined to explain this experimental observation by considering that Glu 85 and Lys 96 are at the boundaries of helix F (residues 82–94; see also footnote 3), and perhaps they are not as mobile as other residues occurring at mid-positions in chain segment 82–94 (see Fig. 3). Moreover, the Glu-specific V8-protease cleaves apoMb at Glu 85 and not at the nearby Glu 83, signifying that this last residue is sufficiently rigid to hamper proteolysis. Therefore, it seems that proteolysis experiments can be used to define the boundaries of the flexible/unfolded chain segment amenable to proteolytic attack.

**Table 1.** *Molecular masses of proteolytic fragments of wild-type apomyoglobin*

| | | Molecular mass (Da)[b] | |
|---|---|---|---|
| Protease | Fragment[a] | Found | Calculated |
| Thermolysin | 1–88 | 9943.9 | 9943.5 |
| | 89–153 | 7275.1 | 7274.4 |
| | 89–148 | 6756.1 | 6755.8 |
| Subtilisin | 1–89 | 10057.1 | 10056.7 |
| | 90–153 | 7162.2 | 7161.2 |
| | 92–153 | 6961.9 | 6962.0 |
| Chymotrypsin | 1–89 | 10057.6 | 10056.7 |
| | 90–153 | 7162.6 | 7161.2 |
| Trypsin | 1–96 | 10781.1 | 10780.5 |
| | 97–513 | 6437.8 | 6437.4 |
| Proteinase K | 1–89 | 10056.7 | 10056.7 |
| | 1–91 | 10255.1 | 10255.9 |
| | 90–153 | 7161.1 | 7161.2 |
| | 92–153 | 6961.5 | 6962.0 |
| | 94–153 | 6737.7 | 6737.8 |
| | 96–153 | 6565.5 | 6565.6 |
| V8 Protease | 1–85 | 9606.6 | 9605.1 |
| | 86–153 | 7613.2 | 7612.8 |
| Papain | 1–90 | 10128.1 | 10127.8 |
| | 91–153 | 7090.8 | 7090.1 |
| | 92–153 | 6962.3 | 6962.0 |
| | 90–153 | 7164.5 | 7161.2 |
| | 97–153 | 6437.5 | 6437.4 |
| Elastase | 1–89 | 10056.6 | 10056.7 |
| | 90–153 | 7161.5 | 7161.2 |
| | 92–153 | 6961.9 | 6962.0 |

[a] The protein fragments were obtained by limited proteolysis of apoMb as described in the text. Identification of the fragments was obtained by N-terminal sequencing (not shown) and ESI-MS data.
[b] Molecular masses were calculated from the known amino acid sequence of sperm-whale apoMb. Experimental molecular masses were determined by ESI–MS as described in Materials and Methods.

### *Proteolysis of P88A and P88G mutants*

The modulation of the integrity of helix F in the two apoMb mutants P88A and P88G was evaluated by conducting limited proteolysis experiments with thermolysin and proteinase K and comparing the results with those obtained with wild-type apoMb. Proteolysis of the apo forms of P88A and P88G mutants with thermolysin were conducted at 25°C and an E/S ratio of 1/100 (by weight). As shown by SDS-PAGE (Fig. 4, top), the P88A mutant is strongly resistant to the proteolytic attack by thermolysin and, after a 1-min reaction, the protein is still almost fully intact. After a 10-min reaction, faint bands appear in the gel, corresponding to small amounts of protein fragments. Upon prolonged incubation, the band corresponding to the intact protein progressively fades from the gel without producing relatively large fragments, indicating that the protein is degraded into a mixture of low molecular-weight peptides, not stained in the gel and early eluted from the RP-HPLC column (see Fig. 4, bottom). Therefore, the Pro→Ala substitution strongly hampers proteolysis and the proteolysis of the P88A mutant qualitatively parallels that of the native holo protein (see

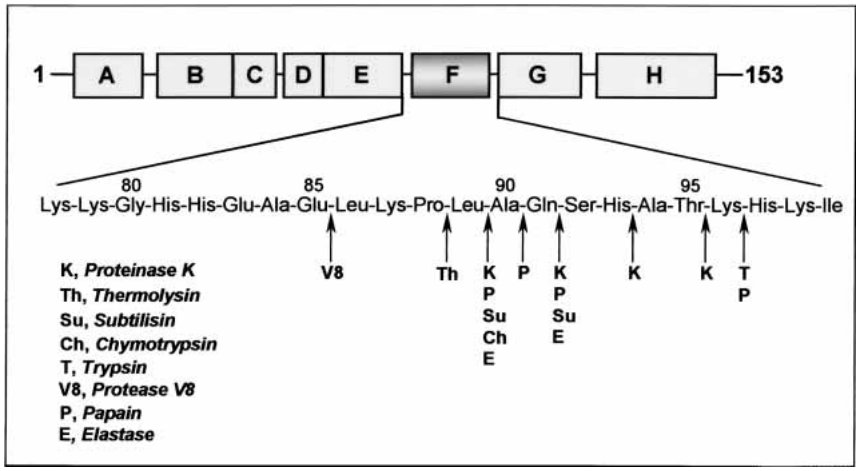


**Figure 3.** (*Top*) Schematic secondary structure of sperm whale myoglobin (Kendrew et al. 1958; Takano 1977a,b; Kuriyan et al. 1986). The height helices are shown as boxes. (*Bottom*) The amino acid sequence corresponding to loop EF, helix F, and loop FG (residues 78–99) is magnified, and the sites of initial proteolytic cleavage by proteinase K (K), thermolysin (Th), subtilisin (Su), chymotrypsin (Ch), trypsin (T), V8-protease (V8), papain (P), and elastase (E) are indicated by arrows.

Fig. S1, Supplemental Material), thus indicating a highly structured protein without flexible sites amenable to proteolytic attack.

Conversely, the proteolytic pattern of the P88G mutant with thermolysin closely resembles that obtained with wild-type apoMb, but the rate of proteolysis is lower with the mutant. The combined analysis of SDS-PAGE and RP-HPLC of the proteolysis mixture (Fig. 4) clearly indicates that two major fragments are initially formed. The identity of these fragments established that the initial thermolytic cleavage occurs at Gly 88–Leu 89 peptide bond, thus, at the same site as in native apoMb. An additional but minor cleavage occurs at the level of peptide bond Glu 85–Leu 86. Therefore, a Gly residue at position 88 of the apoMb chain allows proteolysis to occur, but at a reduced rate.

The different susceptibility to proteolysis of the three apoMb variants was further assayed by comparing the rates and patterns of fragmentation of these proteins obtained with proteinase K, a very active protease that does not show substrate specificity. Proteolysis of the apo forms of P88A and P88G mutants with proteinase K were performed at 25°C at an E/S ratio of 1/100 (by weight). The time-course analysis of the fragmentation reaction and the identification of the fragments produced were conducted as above (detailed data not shown here). The proteolysis of the P88A mutant occurs at a strongly reduced rate with respect to that observed with the wild-type protein. In fact, after a 10-min reaction, the P88A substrate was still ~90% present in the mixture together with small amounts of fragments. Despite the low quantity of the fragments obtained, it was possible to identify by ESI–MS the presence of fragments 1–87, 90–153, and 94–153 in the proteolysis mixture, thus implying that proteinase K initially cleaves mostly at peptide bonds Lys 87–Ala 88, Leu 89–Ala 90, and His 93–Ala 94. The proteolysis of the P88G mutant with proteinase K (detailed data not given here) displays an intermediate behavior between those evidenced for the wild-type protein and P88A mutant. After a 5-min reaction, protein fragments are produced and shown to consist of amino-terminal fragments 1–87 and 1–89 and carboxy-terminal fragments 90–153, 92–153, 94–153, and 96–153. Overall, proteolysis experiments conducted on the three variants of apoMb using proteinase K as a proteolytic probe parallel those obtained using thermolysin (Fig. 4). With both proteases, the rate of limited proteolysis is in the order wild-type > P88G >> P88A.

### *Circular dichroism measurements*

To possibly correlate the susceptibility to proteolysis of the apoMb variants with their conformational features, we measured the far-UV CD spectra of these proteins (Fig. 5). The three protein variants show far-UV CD spectra with minima of ellipticity near 208 and 222 nm, which are typical signatures of highly helical polypeptides (Chen et al. 1974; Woody 1995). The figure of negative ellipticity at 222 nm, $[\theta]_{222}$, reflecting the helical content of a polypeptide chain, is in the order P88A > P88G > wild type, indicating that the Pro→Ala substitution enhances the helical character of the protein, whereas the Pro→Gly exchange has a relatively minor effect. The fact that Pro→Ala replacement leads to an enhancement of helicity of the protein clearly results from the far-UV CD difference spectrum between that of the P88A mutant and that of the wild-type protein. The difference spectrum shows the minima of negative ellipticity near 208 and 222 nm, characteristic of helical polypeptides (see

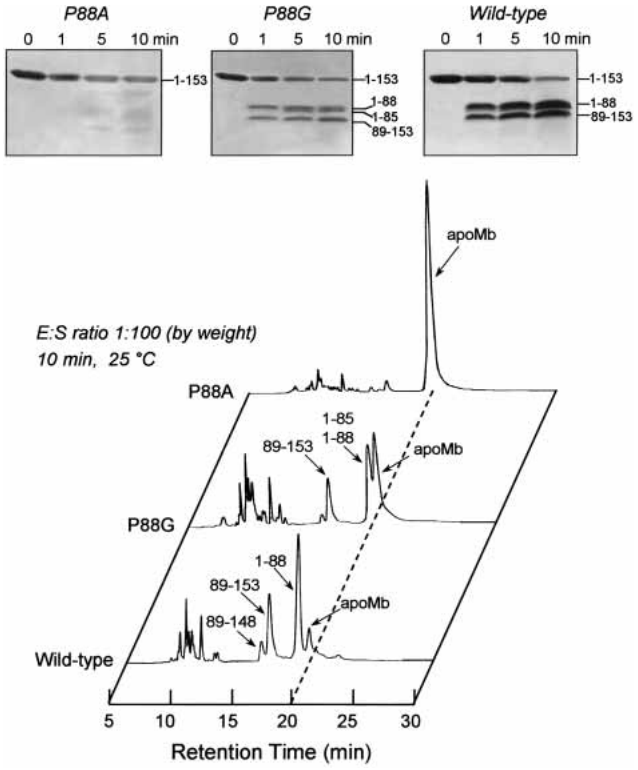


**Figure 4.** Limited proteolysis of P88A and P88G mutants, as well as of wild-type apoMb, with thermolysin. (*Top*) SDS-PAGE analysis of the proteolysis reactions of the variants of apoMb. Aliquots were taken from the reaction mixture after 1, 5, and 10 min of incubation. A sample of intact apoMb (lane *0*) was loaded in each gel. (*Bottom*) RP-HPLC analysis of an aliquot of the proteolysis mixture of the wild-type, P88A, and P88G variants of apoMb with thermolysin. The reaction time was 10 min at 25°C. Numbers near the chromatographic peaks refer to the identity of the apoMb fragments.

Fig. 5). Quantitation of the otherwise similar far-UV CD spectra of the three proteins (Fig. 5) in terms of their different helical contents should be considered within the experimental limitations in measuring exact protein concentrations that dictate the figures of ellipticity at 222 nm (Johnson 1990; Venyaminov and Yang 1996; Kelly and Price 2000). Nevertheless, the changes in the CD spectra with the three apoMb variants are qualitatively those expected from an enhancement of protein helicity by replacing Pro 88 with Gly and, in particular, with Ala (see Discussion).

### *Helical propensities of apoMb mutants*

The helical propensities of the 153-residue chain of wild-type apoMb, as well as of the P88G and P88A mutants, were calculated by using the semi-empirical algorithm AGADIR (Muñoz and Serrano 1997). As shown in Figure 6, differences in helical propensities of the three proteins are observed only at the level of chain segment 80–100, encompassing helix F in the holo protein. In this region, the wild-type protein is predicted to have an extremely low probability of forming a helix. The P88G mutant displays a weak helical tendency, whereas the P88A mutant shows a large helical propensity, about 30 times larger than that of wild-type apoMb and seven times larger than that of the P88G mutant. Therefore, the AGADIR predictions are consistent with the results obtained by limited proteolysis experiments (Fig. 4) and CD measurements (Fig. 5) and clearly show that an amino acid replacement of Pro 88 is sufficient to drastically change the conformational properties of the 82–94 chain segment.

## Discussion

### *Conformational features of apoMb*

The limited proteolysis technique (Mihalyi 1978; Price and Johnson 1990; Kriwacki and Siuzdak 1998) can be advantageously used to pinpoint in a globular protein site of flexibility or local unfolding (Fontana et al. 1986, 1993,

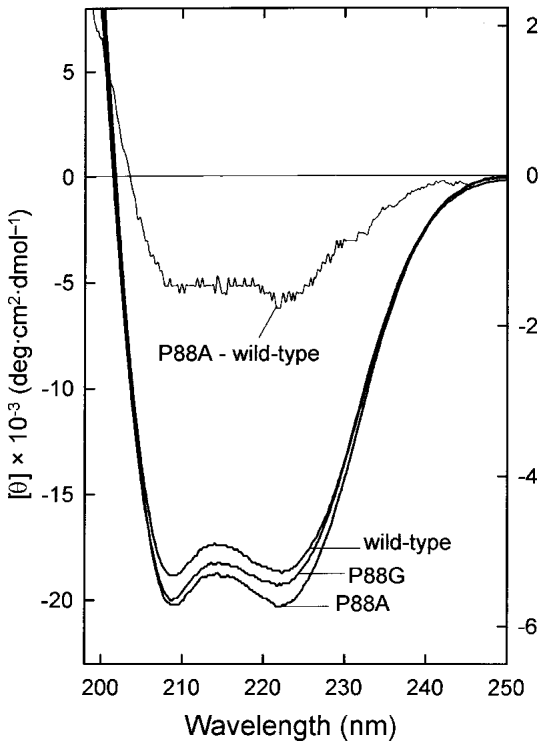


**Figure 5.** Far-UV CD spectra of the apo forms of wild-type sperm-whale myoglobin and its mutants P88A and P88G. The difference CD spectrum between that of the P88A mutant and that of the wild-type protein is also shown. Note that for this CD spectrum, the scale of ellipticity is at the *right* part of the diagram. Spectra were recorded at 25°C at a protein concentration of 0.11–0.15 mg/mL, in 20 mM phosphate buffer, 0.15 M NaCl (pH 7.0). Spectra were taken on protein samples gel filtered on a Superdex-75 equilibrated and eluted with the same phosphate buffer used for CD measurements.

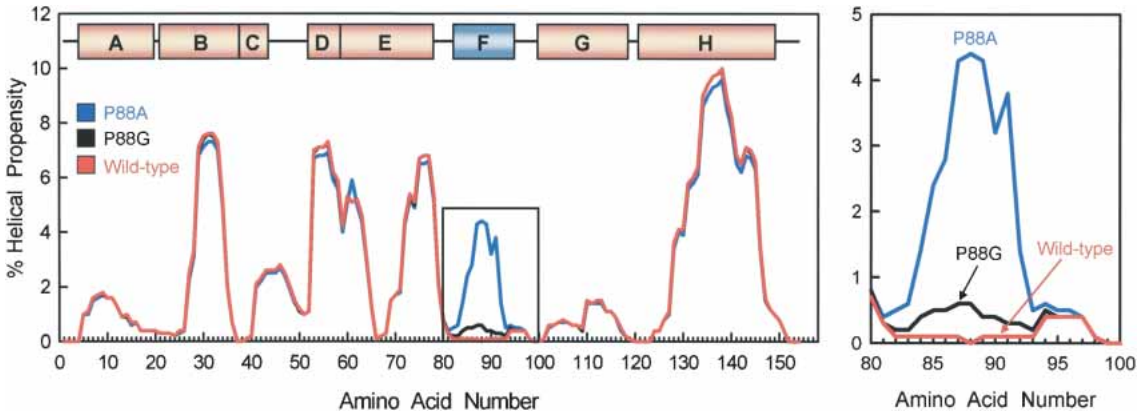


**Figure 6.** Helical propensity vs. amino acid number of wild-type apoMb (red), P88A mutant (blue), and P88G mutant (black). Helical propensities were calculated by using the AGADIR algorithm (Muñoz and Serrano 1997). The location of the eight helices (A–H) along the 153-residue chain is also indicated by boxes (*top*). In the *right* part of the figure, the helix propensities of the 80–100 chain segment of the apoMb variants are magnified.

1997a,b, 1999; Hubbard 1998). Here, the main conclusion derived from proteolysis experiments is that the chain segment 82–94 of sperm-whale apoMb, encompassing helix F in holo myoglobin, is unfolded or at least highly flexible, whereas the rest of the 153-residue chain of the protein is well packed, and thus resistant to the proteolytic attack (see Fig. 7A). Considering that the holo protein is fully stable to proteolysis, the observed patterns and rates of proteolysis can be related to the conformational and dynamic features of apoMb only. The selectivity of proteolytic cleavages for chain segment 82–94, summarized in Figure 3, are striking if one considers the variety of proteases used, including the very active and nonspecific proteinase K.

Upon examination of the amino acid sequence of chain segment 82–94 (see Fig. 1), it is obvious to note that this chain segment, otherwise helical in the holo protein, contains the helix-breaker Pro 88 residue. Therefore, it is reasonable to propose that there are several interactions between the heme and side chains of amino acid residues belonging to helix F that can stabilize this helix in the holo protein, and thus counterbalance the helix-breaking effect of Pro 88. As schematically shown in Figure 7B, helix F in holo myoglobin is strongly stabilized by the hydrophobic environment created by the nearby heme moiety and, in particular, by the covalent bond between the pyrrole nitrogen of His 93 and the heme iron (Liong et al. 2001). Moreover, hydrophobic interactions occur between the aliphatic side chain of Leu 89 and the heme and hydrogen bonding involves His 93, the hydroxyl oxygen of Ser 92 and the heme-7-propionate, as well as the main chain of Leu 89 (Smerdon et al. 1993). These interactions significantly restrict the conformational space available for chain segment 82–94 and constrain the heme pocket. Indeed when the hydrogen bonding is disrupted by replacing Ser 92 with aliphatic residues (Ala, Val, Leu), the rigid packing of the heme pocket is relaxed, and the heme easily dissociates from the protein (Smerdon et al. 1993; Shiro et al. 1994). Therefore, it is expected that, upon removal of the heme, apoMb acquires a more flexible state and that the chain

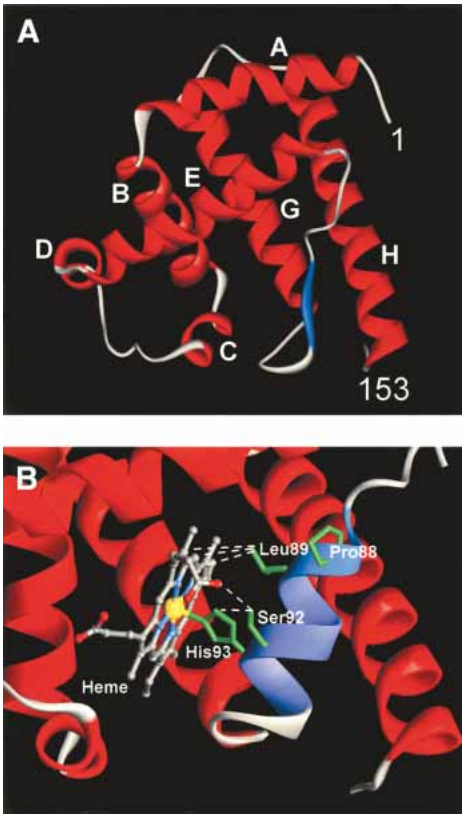


**Figure 7.** Schematic structure of apoMb (*A*) and of the heme environment in holo myoglobin (*B*) (Hargrove et al. 1996b). The Pro 88 residue and the key residues involved in heme–helix F interactions are shown in green. Electrostatic interactions, hydrogen bonds, and hydrophobic interactions are indicated with dashed white lines, whereas the imidazole-iron bond of His 93 is shown in green.

segment 82–94 adopts a conformation compatible with its intrinsic conformational propensity. This view is firmly supported by the profile of helix propensity along the 153-residue chain of wild-type apoMb, showing that the chain segment 82–94 tends to be disordered in the apoprotein (see Fig. 6).

The fact helix F is missing/disrupted in apoMb with respect to the holo protein is in agreement with the results of previous NMR studies (Eliezer and Wright 1996; Lecomte et al. 1996, 1999; Eliezer et al. 1998). Hydrogen exchange measurements by NMR fail to detect protection mostly at the chain segment encompassing helix F, whereas the rest of the hydrophobic helical core of apoMb shows structural and dynamic properties quite similar to those of the native holo protein. The enhanced flexibility of helix F in apoMb has been detected also by molecular dynamics simulations (Brooks 1992; Tirado-Rives and Jorgensen 1993; Hirst and Brooks 1994, 1995; Onufriev et al. 2003). Recently, it has been shown that the stepwise increase of pressure up to 3000 bar leads to enhanced populations of more and more unfolded species of apoMb and that, at low pressures (1–30 bar), the protein is predominantly native. NMR and hydrogen/deuterium exchange measurements, as well as rates of proteolysis with subtilisin, were used to monitor the pressure-induced conformational changes of apoMb (Tanaka et al. 2000; Kitahara et al. 2002). Of interest, NMR signals for native apoMb were missing in the chain region 82–102, thus indicating that helix F was disrupted (Kitahara et al. 2002). The fact that at least helix F is missing in apoMb results also from far-UV CD measurements, indicating that the holo protein is ~20% more helical than the apo form (Griko et al. 1988). Finally, we may mention here that the intrinsic propensity of the chain segment encompassing helix F to be disordered results also from studies of the conformational features of synthetic peptides of sperm-whale apoMb. Whereas all peptides covering the sequences of individual helices of myoglobin can adopt in aqueous trifluoroethanol a helical conformation to varying degrees, the helix F peptide, instead, remains fully disordered (Reymond et al. 1997). In conclusion, the results of proteolysis experiments, as well as of other previous studies, allow us to depict the 3D structure of apoMb as similar to that of the holo protein, with the difference that helix F is unfolded (see Fig. 7A).

### *Amino acid replacements at helix F*

Amino acid residue replacement is a most-used tool to investigate conformational features and transitions of polypeptides. Both Pro and Gly residues produce a large decrease in helix propensity in a polypeptide chain, but Pro is the strongest helix-breaker residue (Lyn et al. 1990; Merutka et al. 1990; O'Neil and De Grado 1990; Padmanabhan et al. 1990; Chakrabartty et al. 1991, 1994; Pace and Scholz 1998). As a matter of fact, in globular proteins, usually Pro residues are located at chain loops, at disordered regions, or at termini of helices, whereas Gly residues can be tolerated inside of helices. In the case of sperm-whale apoMb, Gly residues occur at chain segments encompassing helix A, B, E, and H. The four Pro residues are at the junction between helices B and C, at the junction of helices G and H, and at the amino terminus of helix G, whereas Pro 88 is located in the middle of the helical segment 82–94 (see Fig. 1). Therefore, it is expected that Pro 88 exerts a dramatic effect on the conformational features of this segment. Here, the Pro→Ala exchange was performed in order to remove the helix-breaking effect of Pro 88, as Ala is the best helix-forming residue in different scales for propensities of the 20 amino acid residues (Merutka et al. 1990; Chakrabartty et al. 1991) and occurs much more frequently than any other amino acid residue in the mid-helical positions of α-helices in globular proteins (Richardson and Richardson 1988). Therefore, a dramatic effect of the Pro→Ala exchange in favoring the helical secondary structure in the chain segment 82–94 is expected, whereas the Pro→Gly exchange appears to be a conservative replacement, and thus not influencing significantly the conformational features of segment 82–94.

The conformational features of sperm-whale apoMb (wild type, P88A, and P88G) have been analyzed by far-UV CD measurements. Within the limits of the computing methods, it is possible to evaluate the percent content of α-helix in a protein from far-UV CD spectra (Chen et al. 1974; Johnson 1990; Woody 1995; Kelly and Price 2000). The CD spectra for the three mutant species of apoMb indicate a ranking of the helical content as P88A > P88G > wild-type (Fig. 5). The enhancement of helicity for the P88A mutant, calculated from far-UV CD data (Chen et al. 1974), is ~5%, whereas it can be estimated that the presence or absence of helix F (segment 82–94) can contribute ~10% to the helical content of the protein bearing the eight helices (A–H). However, the contribution of helix F to the far-UV CD spectrum of apoMb, calculated by molecular dynamics simulations, was shown to be much less than those of the other helices of the protein (Hirst and Brooks 1994, 1995). Therefore, we may conclude that far-UV CD data are consistent with the existence of helix F in the P88A mutant.

Numerous mutants of sperm-whale apoMb have been prepared previously for studying features of folding, stability, and function of the protein (Hughson and Baldwin 1989; Barrick and Baldwin 1993; Lin et al. 1993, 1994a,b; Pinker et al. 1993; Low et al. 1995; Hargrove et al. 1996a,b; Scott et al. 2000; Cavagnero et al. 2001; Luo and Baldwin 2001). Frequent amino acid replacements were performed at the level of helices A, G, and H, in order to probe the structural features of the MG state of apoMb observed at low pH. Even the Pro→Ala mutation in position 88 of the 153-residue chain of the protein was earlier performed, but the reported figure of the CD ellipticity at 222 nm for the P88A

mutant was quite unusual and controversial, as it was reported that this mutant displayed a mean residue ellipticity at 222 nm of −8620 deg·cm$^2$·dmole$^{-1}$, whereas in that study, the corresponding figure for the wild-type protein was −17,500 deg·cm$^2$·dmole$^{-1}$ (Lin et al. 1994a). Of course, this large decrease in helicity (~50%) due to the single Pro→Ala exchange was fully unexpected. This prompted Eliezer and Wright (1996) to speculate on possible interpretations of the very unusual finding that Ala was a strong destabilizer of the helical conformation of apoMb. However, present data clarify the controversy, and show instead that the P88A mutant retains at least the helical content of the wild-type protein.

*Biological implications*

Here, we would like to comment on the possible biological significance of the high flexibility or unfolding of helix F in apoMb. Above, we have concluded that the Pro 88 residue is responsible for the disruption of the helix. The conformational features of apoMb herewith proposed for the sperm-whale protein likely are shared also by other homologous myoglobins, considering that Pro 88 is largely conserved (see Fig. S6, Supplemental Material). Therefore, a disordered chain segment 82–94 in apoMb may play a crucial role in the function and fate of myoglobin at the cellular level.

At room temperature, about half of the free energy change involved in the heme-apoprotein interaction is given by a hydrophobic partitioning of the heme moiety into the apoprotein, whereas the rest is due to specific interactions between side chains of residues of the heme pocket and formation of the $Fe^{3+}$–His 93 bond (Hargrove et al. 1996a,b). On this basis, it could be envisaged that the heme binding initially leads to a transient, compact hydrophobic cluster, and then the mutual heme-polypeptide interactions cause a local refolding of chain segment 82–94 of apoMb into a hydrogen-bonded, helical structure. Clearly, this binding process can be facilitated by a disordered chain segment, as given by the fact that the binding of heme to apoMb occurs in few seconds as a rapid bimolecular association reaction ($k_{on}$ of the order of $10^8$ $M^{-1}s^{-1}$), independently of protein primary structure (Hargrove et al. 1996a,b). The present case of heme–apoMb binding seems to hold generality, as it has been demonstrated that the ligand uptake by proteins requires some protein flexibility and that this process can involve an overall or local protein conformational change. Proteins also can adopt an open conformation to facilitate the binding of the ligand and a closed conformation to keep the ligand bound in a fixed stereochemistry (for a recent review, see Teague 2003). It may be important that in apoMb, the amino acid sequence of the chain segment roughly from residue 82 to 94 should be on one side flexible enough for facilitating the binding of heme to the apoprotein in an open conformation and, on the other, also able to provide a rigid scaffold in which to accommodate optimally the heme moiety in the holo protein in a closed conformation. Indeed, the well-packed holo protein loses the heme very slowly with a rate constant of $10^{-6}s^{-1}$, and a dissociation constant of the order of $10^{-14}$; moreover, the dissociation is strongly influenced by changes in the amino acid sequence of the protein (Hargrove et al. 1996a,b).

The protein region that allows the entry and exit of the ligand into a protein has been named "portal" (Sacchettini et al. 1989; Hodsdon and Cistola 1997; Jenkins et al. 2002). For example, a dynamic portal region was proposed to play a crucial role in the binding of retinol or fatty acids to lipocalins (for a recent study, see Franzoni et al. 2002). Perhaps we can propose the presence in myoglobin of a portal region, constituted mostly by chain region 82–94, which can undergo large movements to enable the opening and closing of the portal, as schematically shown in Figure 8. Once bound, heme would induce significant structural changes at that region from a highly flexible or unfolded chain segment to the helical secondary structure, which closes the entrance and stabilizes the complex. It can be predicted that the decrease in flexibility deriving from the heme binding is associated with an entropy loss, which is compensated by a favorable binding enthalpy resulting from novel interactions between the heme and amino acid side chains, such as those shown in Figure 7B. It would be of interest in future studies to examine, in kinetic and thermodynamic terms, the binding of heme to the apoMb mutants herewith investigated, in analogy to the studies already conducted on many mutants of sperm-whale apoMb (Hargrove et al. 1996a,b; Scott et al. 2000).

It is intriguing to understand why the holo and apo forms of myoglobin differ so much in their rate of proteolytic breakdown. First of all, the high stability of the holo protein to proteolysis reflects the need in mammals to prevent oxidative damage of muscle tissue. Consequently, we may understand why, in order to avoid protein degradation, hemin should be firmly bound to the protein (see above). Conversely, what are the peculiar molecular features of apoMb

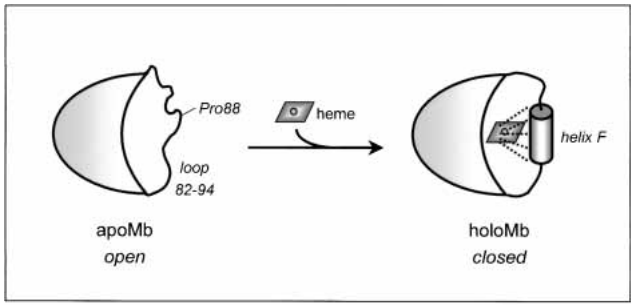


**Figure 8.** Schematic view of the portal hypothesis for apoMb. An open conformation facilitates the entry of the heme into the heme pocket, whereas the closed conformation, characterized by a newly formed helix F, closes the entrance of the portal and keeps the heme moiety buried and firmly bound to the protein.

that determine its extreme lability to proteolysis? In this context, we may mention that polypeptide sequences enriched in proline (P), glutamic acid (E), serine (S), and threonine (T) are favored sites of proteolysis and proteins containing a PEST region are rapidly degraded (Rogers et al. 1986; Rechsteiner and Rogers 1996). The PEST hypothesis implies that proteins may contain a hydrophobic stretch of 12 or more amino acid residues containing at least one of P, E, S, and T residues, flanked by lysine or histidine. Of note, the PEST regions of proteins of known X-ray structure are not resolved because of polypeptide backbone flexibility at that region (Rechsteiner and Rogers 1996). Interestingly, the chain region of apoMb, which is cleaved by a variety of proteases contains P88, E85, S92, and T95 (see Fig. 3). Therefore, it can be proposed that in apoMb, there is a PEST region that serves as a constitutive signal for rapid protein degradation, in analogy to other proteins (Rechsteiner and Rogers 1996). The PEST region is masked in holo myoglobin and becomes exposed and flexible upon heme dissociation from the protein, as given by the fact that the holo and apo forms are degraded in days (see Fig. S1, Supplemental Material) and minutes (see Fig. 2), respectively. In support of this hypothesis, we may mention that cAMP kinase, which contains a PEST region, shows a half-life of degradation by proteolysis of about 10 h, whereas the cofactor-free protein shows about 1 h (Rechsteiner 1990). Finally, we may add here that the amino acid residues constituting the PEST region are among the most frequently found in disordered regions of proteins (Radivojac et al. 2004).

The disordered chain region in apoMb herewith detected by limited proteolysis experiments does not appear to be unusual within globular proteins. In fact, numerous proteins contain largely disordered regions of up to 40 amino acid residues (Romero et al. 2001). Moreover, some proteins appear to be even fully unfolded under their normal conditions in the cell (Tompa 2000; Uversky et al. 2000; Dyson and Wright 2002; Uversky 2002). A number of experimental observations have highlighted that partly or fully unfolded states play a role in the functioning of proteins at the cellular level, including ligand binding (Wright and Dyson 1999; Teague 2003). Therefore, the protein structure-function paradigm appears to be challenged by these observations and prompts alternative views to the interpretation of function of proteins in terms of their fixed 3D structure (Dunker and Obradovic 2001; Dunker et al. 2002; Teague 2003). Therefore, the apoMb case should be viewed in this perspective, in the sense that the disordered chain segment, otherwise embedded in a well-packed protein moiety, is that required for both heme binding and protein turnover. In this respect, we may mention that the conformational features of apoMb parallel those of the heme-binding protein apocytochrome $b_5$. Also, in this 98-residue protein there is a disordered chain segment, roughly comprising residues 40–69, which appears to facilitate the binding of heme (Falzone et al. 1996).

### *Concluding remarks*

The results of this study further emphasize the utility of proteolytic probes of protein structure and dynamics (Mihalyi 1978; Fontana et al. 1986, 1993, 1997a,b, 1999; Price and Johnson 1990; Hubbard et al. 1994; Hubbard 1998). In previous studies, using a variety of spectroscopic and computational approaches, it has been proposed that helix F is mostly disrupted in native apoMb at neutral pH. Here, it is shown that, by using the simple biochemical approach of limited proteolysis, it is possible to detect the unfolding of helix F. Therefore, it is the mobility of the chain segment(s) of the protein substrate that dictates the limited proteolysis phenomenon. When the mobile chain region encompassing helix F is induced to adopt a quite rigid and hydrogen-bonded structure, as that resulting from a Pro→Ala replacement, the site of limited proteolysis is rather well protected against the proteolytic attack.

We wish to emphasize again that quite often in past and current literature, limited proteolysis events are wrongly interpreted in terms of exposure of the site(s) of cleavage (Novotny and Bruccoleri 1987). Of course, the notion of accessibility is a required property of the sites of cleavage in order that the bimolecular reaction between the protease and the protein substrate can take place, but not at all sufficient to explain the selective proteolysis of one single peptide bond among hundred(s) of bonds, as often observed in limited proteolysis experiments. There are plenty of exposed peptide bonds in a globular protein, but the one that is cleaved should be embedded in a highly flexible or unstructured chain region (Fontana et al. 1986, 1993, 1997a,b, 1999). Therefore, aiming to probe the "surface topography" of a protein by the limited proteolysis approach is simply unfounded. Instead, the approach is eminently suitable to pinpoint in a globular protein the sites of chain flexibility or local unfolding. The correlation between the sites of limited proteolysis and the mobile sites detected by X-ray or NMR methods (Fontana et al. 1986, 1997a,b), as well as by molecular dynamics simulations (Stella et al. 1999; Falconi et al. 2002), has been amply documented. In this respect, we may mention that in a recent study, disordered chain regions in protein structures were identified by both proteolysis experiments and predictions of their location along the polypeptide chain by the neural network program PONDR (Prediction of Natural Disordered Regions; Iakoucheva et al. 2001). Predictions nicely correlated with the results of limited proteolysis, thereby indicating that chain disorder or flexibility is the key parameter dictating limited proteolysis events in proteins.

## Materials and methods

### *Materials*

Sperm-whale myoglobin was purchased from Biozyme Laboratories. Thermolysin from *Bacillus thermoproteolyticus*, subtilisin

from *Bacillus subtilis*, trypsin from bovine pancreas, proteinase K from *Tritirachium album*, elastase from pig pancreas, and papain from *Carica papaya* were obtained from Sigma Chemical Co., whereas chymotrypsin and V8-protease (endoproteinase Glu-C) from *Staphylococcus aureus* from Fluka. Cyanogen bromide (BrCN), acetonitrile and trifluoroacetic acid (TFA) were purchased from Fluka. The materials used for sodium dodecyl sulfate (SDS-polyacrylamide gel electrophoresis PAGE) were from Bio-Rad, and those for amino-terminal sequence analysis from Applied Biosystems. The basic components for culture medium were purchased from Biokar Diagnostics. The DNA sequencing kit was from Amersham, and the DNA purification kit from Qiagen. All other chemicals were analytical grade and were obtained from Sigma or Fluka.

Apomyoglobin was obtained from holo myoglobin by removal of the heme using the 2-butanone extraction procedure (Teale 1959). The possible contamination of apoMb by myoglobin was assessed spectrophotometrically, and no significant absorption was observed in the Soret region.

### *Expression and purification of sperm-whale myoglobin and its mutants*

Myoglobin mutants containing the P88A and P88G replacements were constructed by site-directed mutagenesis from a pEMBL19+ vector containing the gene pEMbS-1 for wild-type myoglobin (Springer and Sligar 1987; Carver et al. 1992). The plasmid bearing the synthetic gene of sperm-whale myoglobin was a kind gift of Prof. Eileen W. Singleton (Department of Biochemistry and Cell Biology, Rice University). To introduce point mutations, inverse PCR was performed on plasmid pEMbS-1 as a template using the primers 5′-GCTGAGCTCAAAGCGCTTGCGCAATC GC-3′ (Pro1) and 5′-CTTTGAGCTCAGCTTCATGATGCCC-3′ (Pro3) for the P88A mutant and primers 5′-GCTGAGCTCAAA GGGCTTGCGCAATCGC-3′ (Pro2) and 5′-CTTTGAGCTCAG CTTCATGATGCCC-3′ (Pro3) for the P88G mutant. After amplification, DNA fragments were cut with SacI and themselves ligated. Plasmid constructions were checked by double-stranded dideoxynucleotide sequencing using a Perkin-Elmer DNA sequencer model 477R. Recombinant myoglobins were expressed in *Escherichia coli* (TB-I strain) and purified following essentially the procedure described by Springer and Sligar (1987) and Carver et al. (1992). Briefly, cells were resuspended in 50 mM Tris·HCl (pH 6.0), containing 1 mM EDTA, 0.5 mM dithiothreitol, and 1 mM *p*-toluenesulfonyl chloride. DNase I, RNase A, and lysozyme (purchased from Sigma) were added to 40 U/mL, 3 U/mL, and 2 mg/mL, respectively. Cells were stirred overnight at room temperature and centrifuged. The supernatant was fractionated by ammonium sulfate precipitation. The 50%–95% cut was redissolved in a minimum volume of 20 mM Tris·HCl, 1 mM EDTA (pH 8.0), dialyzed against the same buffer, and then loaded onto an anion-exchange DEAE-column (Whatman DE-52). The reddish-brown fractions containing myoglobin were pooled and, after lowering the pH to 6.0 with aqueous acetic acid, were loaded onto a cation-exchange CM-cellulose column (Whatman CM-52), eluted with a gradient from 20 mM sodium phosphate (pH 6.0), to 50 mM sodium phosphate (pH 9.0). This procedure yielded 2–6 mgs of protein per liter of cell culture. Protein purity was assessed by SDS-PAGE, RP-HPLC, and ESI–MS molecular mass determination. The recombinant sperm-whale myoglobin and its mutants P88G and P88A contain an extra amino-terminal Met, and this residue was numbered −1 in order to keep the same numbering of the 153-residue chain of the wild-type protein.

### *Proteolysis experiments*

Proteolyses of sperm-whale apoMb and mutants were conducted at 25°C with a variety of proteases (proteinase K, thermolysin, subtilisin, papain, elastase, chymotrypsin, V8-protease, and trypsin). Limited proteolysis of apoMb was performed with the protein dissolved (0.4–0.6 mg/mL) in 20 mM of Tris·HCl, 0.15 M NaCl, using an enzyme/substrate (E/S) ratio of 1:100 (by weight). The pH of the proteolysis mixture was 8.0 when trypsin, V8-protease, and elastase were used as proteolytic enzymes, 7.7 in proteolysis with proteinase K, and 7.5 in all other cases. The thermolysin proteolysis mixture contained also 1 mM $CaCl_2$, and with papain, 1 mM cysteine. At intervals, aliquots were taken from the reaction mixtures and the proteolysis was stopped by acidification of the solutions by adding TFA (final concentration 0.1%). The proteolysis mixtures were then separated by RP-HPLC utilizing a $C_{18}$ Vydac column (4.6 × 250 mm; The Separations Group) eluted with a linear gradient of water/acetonitrile, both containing 0.05% (v/v) TFA, from 5% to 40% in 5 min, and from 40% to 60% in 25 min, at a flow rate of 0.8 mL/min. The effluent was monitored by absorption measurements at 226 nm, and the fractions containing the protein fragments were pooled and then concentrated in a Speed-Vac system of Savant. The identity of the fragments was established by ESI–MS, as well as amino-terminal sequencing (actual data not shown). The time-course of the reaction was monitored by SDS-PAGE analysis using the Tricine buffer system (Schägger and von Jagow 1987) under reducing conditions in a vertical gel-slab apparatus (Miniprotean-II, Bio-Rad). A sample of partial BrCN-digest (Fontana and Gross 1986) of sperm-whale apoMb at the level of the two Met residues in positions 55 and 131 of the 153-residue chain was used as standard for molecular masses. The resulting BrCN fragments are made up of residues 1–131, 56–153, 56–131, 1–55, and 132–153, in order of decreasing molecular mass. SDS–polyacrylamide gels were stained with Coomassie Brilliant Blue R-250.

### *Mass spectrometry*

ESI–MS of native or mutant apoMb and their fragments was carried out with a Mariner System 5220 instrument (Applied Biosystems) equipped with a time-of-flight (TOF) analyzer. Spectra were deconvoluted by the software Data Explorer 4.0.0.1 provided by Applied Biosystems.

### *Amino-terminal analysis*

Automated Edman degradation was performed with an Applied Biosystems model 477A protein sequencer equipped with an on-line analyzer (model 120A) of phenylthiohydantoin derivatives of amino acids.

### *Circular dichroism measurements*

CD spectra were recorded at 25°C with a Jasco J-710 spectropolarimeter equipped with a thermostatically controlled cell holder. The instrument was calibrated with *d*-(+)10-camphorsulfonic acid. CD measurements in the far-UV region were performed in 20 mM phosphate buffer, 0.15 M NaCl (pH 7.0), at a protein concentration of 0.11–0.16 mg/mL, utilizing a 1-mm pathlength quartz cell. The results were expressed as mean residue ellipticity $[\theta]$ $(\text{deg}\cdot\text{cm}^2\cdot\text{dmole}^{-1})$ calculated from the formula $[\theta] = (\theta_{obs}/10)(MRW/\text{lc})$, in which $\theta_{obs}$ is the observed ellipticity at a given wavelength, *MRW* is the mean residue molecular weight (protein

molecular mass divided by the number of amino acid residues), l the optical path length in cm, and c the protein concentration in grams/milliliter. Measurements of CD spectra were made on protein samples gel filtered on a Superdex-75 column (type HR 10/30, 1 × 30 cm, Pharmacia), eluted with 20 mM phosphate, 0.15 M NaCl buffer (pH 7.0).

### *Structural calculations*

Helical propensities of the polypeptide chains of wild-type apoMb and its P88A and P88G mutants were calculated by the AGADIR algorithm (Muñoz and Serrano 1997) available at the Web site http://www.embl-heidelberg.de/services/serrano/agadir/agadir-start.html. The parameters of pH 7.0, ionic strength 0.15 M, and 25°C were used. The results were plotted along the 153-residue chain of the protein with the program SigmaPlot.

## Electronic supplemental material

The supplementary figures included in the electronic appendix show the SDS-PAGE gels and RP-HPLC chromatograms of aliquots of the proteolysis mixtures of myoglobin: Figure S1, proteolysis of holo myoglobin with thermolysin; Figure S2, limited proteolysis of apoMb with chymotrypsin; Figure S3, limited proteolysis of apoMb with subtilisin; Figure S4, limited proteolysis of apoMb with V8-protease; and Figure S5, limited proteolysis of apoMb with trypsin. The similarity of the amino acid sequences of apoMbs from different sources at the level of helix F is shown in Figure S6.

## Acknowledgments

We thank Drs. Patrizia Polverino de Laureto and Vincenzo De Filippis for insightful discussions, Dr. Daniele Dalzoppo for assistance with mass spectrometry analysis, and Prof. Eileen W. Singleton for the kind gift of the plasmid bearing the synthetic gene of sperm-whale myoglobin. The expert typing of the manuscript by Barbara Sicoli is also gratefully acknowledged. This work has been supported by the Italian Ministry of University and Research (FIRB Project on Protein Folding and PRIN-2003).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## References

Arai, M. and Kuwajima, K. 2000. The role of the molten globule state in protein folding. *Adv. Protein Chem.* **53:** 209–282.

Barrick, D. and Baldwin, R.L. 1993. Three-state analysis of sperm whale apomyoglobin folding. *Biochemistry* **32:** 3790–3796.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.K., Kennard, O., Shimanouchi, T., and Tasumi, M. 1977. The protein data bank: A computer-based archive for macromolecular structures. *J. Mol. Biol.* **112:** 535–542.

Brooks, C.L. 1992. Characterization of "native" apomyoglobin by molecular dynamics simulation. *J. Mol. Biol.* **227:** 375–380.

Carver, T.E., Brantley Jr., R.E., Singleton, E.W., Arduini, R.M., Quilin, M.L., Phillips Jr., G.N., and Olson, J.S. 1992. A novel site-directed mutant of myoglobin with an unusually high oxygen affinity and low autoxidation rate. *J. Biol. Chem.* **267:** 14443–14450.

Cavagnero, S., Thériault, Y., Narula, S.S., Dyson, H.J., and Wright, P.E. 2000. Amide proton hydrogen exchange rates for sperm whale myoglobin obtained from $^{15}$N-$^{1}$H NMR spectra. *Protein Sci.* **9:** 186–193.

Cavagnero, S., Nishimura, C., Schwarzinger, S., Dyson, H.J., and Wright, P.E. 2001. Conformational and dynamic characterization of the molten globule state of an apomyoglobin mutant with an altered folding pathway. *Biochemistry* **40:** 14459–14467.

Chakrabartty, A., Schellman, J.A., and Baldwin, R.L. 1991. Large differences in the helix propensities of alanine and glycine. *Nature* **351:** 586–588.

Chakrabartty, A., Kortemme, T., and Baldwin, R.L. 1994. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* **4:** 1305–1314.

Chen, Y.H., Yang, J.T., and Chau, K.H. 1974. Determination of the helix and β-form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13:** 3350–3359.

Dunker, A.K. and Obradovic, Z. 2001. The protein trinity: Linking function and disorder. *Nat. Biotechnol.* **19:** 805–806.

Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradovic, Z. 2002. Intrinsic disorder and protein function. *Biochemistry* **41:** 6573–6580.

Dyson, H.J. and Wright, P.E. 2002. Coupling folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12:** 54–60.

Eliezer, D. and Wright, P.E. 1996. Is apomyoglobin a molten globule? Structural characterization by NMR. *J. Mol. Biol.* **263:** 531–538.

Eliezer, D., Yao, J., Dyson, H.J., and Wright, P.E. 1998. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **5:** 148–155.

Evans, S.V. and Brayer, G.D. 1990. High-resolution study of the three-dimensional structure of horse heart metmyoglobin. *J. Mol. Biol.* **213:** 885–897.

Falconi, M., Parrilli, L., Battistoni, A., and Desideri, A. 2002. Flexibility in monomeric Cu,Zn superoxide dismutase detected by limited proteolysis and molecular dynamics simulation. *Proteins* **47:** 513–520.

Falzone, C.J., Mayer, M.R., Whiteman, E.L, Moore, C.D., and Lecomte, J.T.J. 1996. Design challenges for hemoproteins: The solution structure of apocytochrome $b_5$. *Biochemistry* **35:** 6519–6526.

Fink, A.L., Oberg, K.A., and Seshadry, S. 1997. Discrete intermediates versus molten globule models for protein folding: Characterization of partially folded intermediates of apomyoglobin. *Folding Des.* **3:** 19–25.

Fontana, A. and Gross, E. 1986. Fragmentation of polypeptides by chemical methods. In *Practical protein chemistry: A handbook* (ed. A. Darbre), pp. 67–120, J. Wiley and Sons, London-New York.

Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., and Zambonin, M. 1986. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **25:** 1847–1851.

Fontana, A., Polverino de Laureto, P., and De Filippis, V. 1993. Molecular aspects of proteolysis of globular proteins. In *Protein stability and stabilization* (eds. W. van den Tweel et al.), pp. 101–110, Elsevier Science Publ., Amsterdam.

Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. 1997a. Probing the partly folded states of proteins by limited proteolysis. *Folding Des.* **2:** R17–R26.

Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., and Scaramella, E. 1997b. Probing the conformational state of apomyoglobin by limited proteolysis. *J. Mol. Biol.* **266:** 223–230.

Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. 1999. Limited proteolysis in the study of protein conformation. In *Proteolytic enzymes: Tools and targets* (eds. E.E. Sterchi and W. Stöcker), pp. 257–284, Springer Verlag, Heidelberg, Germany.

Franzoni, L., Lücke, C., Révez, C., Cavazzini, D., Rademacher, M., Ludwig, C., Spisni, A., Rossi G.L., and Rüterjans, H. 2002. Structure and backbone dynamics of apo- and holo-cellular retinol-binding protein in solution. *J. Biol. Chem.* **277:** 21983–21997.

Gilmanshin, R., Gulotta, M., Dyer, R.B., and Callender, R.K. 2001. Structures of apomyoglobin's various acid-destabilized forms. *Biochemistry* **40:** 5127–5136.

Goto, Y. and Fink, A.L. 1990. Phase diagram for acidic conformational states of apomyoglobin. *J. Mol. Biol.* **214:** 803–805.

Griko, Y.V., Privalov, P.L., Venyaminov, S.Y., and Kutyshenko, V.P. 1988. Thermodynamic study of the apomyoglobin structure. *J. Mol. Biol.* **202:** 127–138.

Hargrove, M.S., Barrick, D., and Olson, J.S. 1996a. The association rate constant for heme binding to globin is independent of protein structure. *Biochemistry* **35:** 11293–11299.

Hargrove, M.S., Wilkinson, A.J., and Olson, J.S. 1996b. Structural factors governing hemin dissociation from metmyoglobin. *Biochemistry* **35:** 11300–11309.

Hirst, J.D. and Brooks, C.L. 1994. Helicity, circular dichroism and molecular dynamics of proteins. *J. Mol. Biol.* **243:** 173–178.

———. 1995. Molecular dynamics simulations of isolated helices of myoglobin. *Biochemistry* **34:** 7614–7621.

Hodsdon, M.E. and Cistola, D.P. 1997. Discrete backbone disorder in the nuclear magnetic resonance structure of apo intestinal fatty acid-binding protein: Implications for the mechanism of ligand entry. *Biochemistry* **36:** 1450–1460.

Hubbard, S.J. 1998. The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta* **1382:** 191–206.

Hubbard, S.J., Eisenmenger, F., and Thornton, J.M. 1994. Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci.* **3:** 757–768.

Hughson, F.M. and Baldwin, R.L. 1989. Use of site directed mutagenesis to destabilize native apomyoglobin relative to folding intermediates. *Biochemistry* **28:** 4415–4422.

Hughson, F.M., Wright, P.E., and Baldwin, R.L. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science* **249:** 1544–1548.

Iakoucheva, L.M., Kimzey, A.L., Masslon, C.D., Bruce, J.E., Garner, E.C., Brown, C.J., Dunker, A.K., Smith, R.D., and Ackerman, E.J. 2001. Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Sci.* **10:** 560–571.

Jenkins, A.E., Hockenberry, J.A., Nguyen T, and Bernlohr, D.A. 2002. Testing of the portal hypothesis: Analysis of a V32G, F57G and K58G mutant of the fatty acid binding protein of the murine adipocyte. *Biochemistry* **41:** 2022–2027.

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

Johnson, W.C. 1990. Protein secondary structure and circular dichroism: A practical guide. *Proteins* **7:** 205–214.

Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F., and 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.

Kelly, S.M. and Price, N.C. 2000. The application of circular dichroism to studies of protein folding and unfolding. *Biochim. Biophys. Acta* **1338:** 161–185.

Kendrew, J.C., Dickerson, R.E., Dintzis, H.M., Parrish, R.G., Wycoff, H., and Phillips, D.C. 1958. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **185:** 422–427.

Kitahara, R., Yamada, H., Akasaka, K., and Wright, P.E. 2002. High pressure NMR reveals that apomyoglobin is an equilibrium mixture from the native to the unfolded. *J. Mol. Biol.* **320:** 311–319.

Kriwacki, R.W. and Siuzdak, G. 1998. Combined use of proteases and mass spectrometry in structural biology. *J. Biomol. Techniques* **9:** 5–15.

Kuriyan, J., Wilz, S., Karplus, M., and Petsko, G.A. 1986. X-ray structure and refinement of carbon-monoxy (FeII)-myoglobin at 1.5 Å resolution. *J. Mol. Biol.* **192:** 133–154.

Lecomte, J.T.J., Kao, Y.-H., and Cocco, M.J. 1996. The native state of apomyoglobin described by proton NMR spectroscopy: The A-B-G-H interface of wild-type sperm whale apomyoglobin. *Proteins* **25:** 267–285.

Lecomte, J.T.J., Sukits, S.F., Bhattacharya, S., and Falzone, C.J. 1999. Conformational properties of native sperm whale apomyoglobin in solution. *Protein Sci.* **8:** 1484–1491.

Lin, L., Pinker, R.J., and Kallenbach, N.R. 1993. α-Helix stability and the native state of myoglobin. *Biochemistry* **32:** 12638–12643.

Lin, L., Pinker, R.J., Forde, K., Rose, G.D., and Kallenbach, N.R. 1994a. Molten globular characteristics of the native state of apomyoglobin. *Nat. Struct. Biol.* **1:** 447–451.

Lin, L., Pinker, R.J., Phillips, G.N., and Kallenbach, N.R. 1994b. Stabilization of myoglobin by multiple alanine substitutions in helical positions. *Protein Sci.* **3:** 1430–1435.

Liong, E.C., Dou, Y., Scott, E.E., Olson, J.S., and Phillips, G.N. 2001. Waterproofing the heme pocket: Role of proximal amino acid side chains in preventing hemin loss from myoglobin. *J. Biol. Chem.* **276:** 9093–9100.

Low, S.N., Kay, M.S., and Baldwin, R.L. 1995. Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc. Natl. Acad. Sci.* **92:** 5446–5450.

Luo, J. and Baldwin, R.L. 2001. How Ala → Gly mutations in different helices affect the stability of the apomyoglobin molten globule. *Biochemistry* **40:** 5283–5289.

Lyn, P.C., Liff, M.I., Marky, L.A., and Kallenbach, N.R. 1990. Side chain contributions to the stability of α-helical structure in peptides. *Science* **250:** 669–673.

Merutka, G., Lipton, W., Shalongo, W., Park, S., and Stellwagen, E. 1990. Effect of central-residue replacements on the helical stability of a monomeric peptide. *Biochemistry* **29:** 7511–7515.

Mihalyi, E. 1978. *Application of proteolytic enzymes to protein structure studies*. CRC Press, Boca Raton, FL.

Muñoz, V. and Serrano, L. 1997. Development of the multiple sequence approximation within the AGADIR model of α-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* **41:** 495–509.

Nishimura, C., Dyson, H.J., and Wright, P.E. 2002. The apomyoglobin folding pathway revisited: Structural heterogeneity in the kinetic burst phase intermediate. *J. Mol. Biol.* **322:** 483–489.

Novotny, J. and Bruccoleri, R.E. 1987. Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. *FEBS Lett.* **211:** 185–189.

O'Neil, K.T. and De Grado, W.F. 1990. A thermodynamic scale for the helix forming tendencies of the commonly occurring amino acids. *Science* **250:** 646–650.

Onufriev, A., Case, D.A., and Bashford, D. 2003. Structural details, pathways and energetics of unfolding apomyoglobin. *J. Mol. Biol.* **325:** 555–567.

Pace, C.N. and Scholz, J.M. 1998. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **75:** 422–427.

Padmanabhan, S., Marqusee, S., Ridgeway, T., Lane, T.M., and Baldwin, R.L. 1990. Relative helix forming tendencies of nonpolar amino acids. *Nature* **344:** 268–270.

Pinker, R.J., Lin, L., Rose, G.D., and Kallenbach, N.R. 1993. Effects of alanine substitutions in α-helices of sperm whale myoglobin on protein stability. *Protein Sci.* **2:** 1099–1106.

Price, N.C. and Johnson, C.M. 1990. Proteinases as probes of conformation of soluble proteins. In *Proteolytic enzymes: A practical approach* (eds. R.J. Beynon and J.S. Bond), pp 163–180. IRL Press, Oxford, UK.

Ptitsyn, O.B. 1995. Molten globule and protein folding. *Adv. Protein Chem.* **47:** 83–229.

Radivojac, P., Obradovic, Z., Smith, D.K., Zhu, G., Vucetic, S., Brown, C.J., Lawson, J.D., and Dunker, A.K. 2004. Protein flexibility and intrinsic disorder. *Protein Sci.* **13:** 71–80.

Rechsteiner, M. 1990. PEST sequences are signals for rapid intracellular proteolysis. *Semin. Cell. Biol.* **1:** 433–440.

Rechsteiner, M. and Rogers, S.W. 1996. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21:** 267–271.

Reymond, M.T., Merutka, G., Dyson, H.J., and Wright, P.E. 1997. Folding propensities of peptide fragments of myoglobin. *Protein Sci.* **6:** 706–716.

Richardson, J. and Richardson, D.C. 1988. Amino acids preferences for specific locations at the end of helices. *Science* **240:** 1646–1652.

Rogers, G., Wells, R., and Rechsteiner, M. 1986. Amino acid sequences common to rapidly degraded proteins: The PEST hypothesis. *Science* **234:** 364–368.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. 2001. Sequence complexity of disordered proteins. *Proteins* **42:** 38–48.

Sacchettini, J.C., Gordon, J.I., and Banaszak, L.J. 1989. Crystal structure of rat intestinal fatty-acid-binding protein: Refinement and analysis of the *Escherichia coli*-derived protein with bound palmitate. *J. Mol. Biol.* **208:** 327–339.

Schägger, H. and von Jagow, G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166:** 368–379.

Scott, E.E., Paster, E.V., and Olson, J.S. 2000. The stabilities of mammalian apomyoglobins vary over a 600-fold range and can be enhanced by comparative mutagenesis. *J. Biol. Chem.* **275:** 27129–27136.

Shiro, Y., Iizuka, T., Marubayashi, K., Ogura, T., Kitagawa, T., Balasubramanian, S., and Boxer, S.G. 1994. Spectroscopic study of Ser92 mutants of human myoglobin: Hydrogen bonding effect of Ser92 to proximal His93 on structure and property of myoglobin. *Biochemistry* **33:** 14986–14992.

Smerdon, S.J., Krzywda, S., and Wilkinson, A.J. 1993. Serine 92 (F7) contributes to the control of heme reactivity and stability in myoglobin. *Biochemistry* **32:** 5132–5138.

Springer, B.A. and Sligar, S.G. 1987. High level expression of sperm whale myoglobin in *E. coli*. *Proc. Natl. Acad. Sci.* **84:** 8961–8965.

Stella, L., Di Iorio, E.E., Nicotra, M., and Ricci, G. 1999. Molecular dynamics simulations of human glutathione transferase P1-1: Conformational fluctuations of the apo structure. *Proteins* **37:** 10–19.

Takano, T. 1977a. Structure of myoglobin refined at 2.0 Å resolution. I. Crystallographic refinement of sperm whale met-myoglobin. *J. Mol. Biol.* **110:** 537–568.

———. 1977b. Structure of myoglobin refined at 2.0 Å resolution. II. Structure of deoxymyoglobin from sperm whale. *J. Mol. Biol.* **110:** 569–584.

Tanaka, N., Ikeda, C., Kanaori, K., Hiraga, K., Konno, T., and Kunugi, S. 2000. Pressure effect on the conformational fluctuation of apomyoglobin in the native state. *Biochemistry* **39:** 12063–12068.

Tcherkasskaya, O. and Uversky, V.N. 2001. Denatured collapsed states in protein folding: Example of apomyoglobin. *Proteins* **44:** 244–254.

Teague, S.J. 2003. Implications of protein flexibility for drug discovery. *Nat. Drug Disc.* **2:** 527–541.

Teale, F.W.J. 1959. Cleavage of heme-protein link by acid methylethylketone. *Biochim. Biophys. Acta* **35:** 543.

Tirado-Rives, J. and Jorgensen, W.L. 1993. Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry* **32:** 4175–4184.

Tompa, P. 2000. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27:** 527–533.

Uversky, V.N. 2002. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* **11:** 739–756.

Uversky, V.N., Gillespie, J.R., and Fink, A.L. 2000. Why are "natively unfolded" proteins unstructured under physiological conditions? *Proteins* **41:** 415–427.

Venyaminov, S.Y. and Yang, J.T. 1996. Determination of protein secondary structure. In *Circular dichroism and conformational analysis of biomolecules* (ed. G.D. Fasman), pp. 69–107, Plenum Press, New York.

Woody, R.W. 1995. Circular dichroism. *Methods Enzymol.* **246:** 34–71.

Wright, P.E. and Dyson, H.J. 1999. Intrinsically unstructured proteins: Reassessing the protein structure-function paradigm. *J. Mol. Biol.* **203:** 321–331.

Yao, J., Chung, J., Eliezer, D., Wright, P.E., and Dyson, H.J. 2001. NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry* **40:** 3561–3571.