Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Mauk, M. R., & Girotti, A. W. (1973) *Biochemistry 12*, 3187–3193.
Mims, M., Olson, J. S., Russu, I. M., Miura, S., Cedel, T. E., & Ho, C. (1983) *J. Biol. Chem. 258*, 6125–6134.
Noble, R. W., Rossi, G. L., & Berni, R. (1972) *J. Mol. Biol. 70*, 689–696.
Otting, G., Widmer, H. Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.
Phillips, S. E. V. (1980) *J. Mol. Biol. 142*, 531–554.
Ramaprasad, S., Johnson, R. D., & La Mar, G. N. (1984) *J. Am. Chem. Soc. 106*, 5330–5335.
Rance, M., & Wright, P. E. (1986) *J. Magn. Reson. 66*, 372–378.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Ringe, D., Petsko, G. A., Kerr, D. E., & Ortiz de Montellano, P. R. (1984) *Biochemistry 23*, 2–4.
Sebban, P., Copper, M., Alpert, B., Lindqvist, L., & Jameson, D. M. (1980) *Photochem. Photobiol. 32*, 727–731.
Simonneaux, G., Bondon, A., Sodano, P., & Sinbandhit, S. (1989) *Biochim. Biophys. Acta 999*, 42–45.
Springer, B. A., Egeberg, K. D., Sligar, S. G., Rohlfs, R. J., Mathews, A. J., & Olson, J. S. (1989) *J. Biol. Chem. 264*, 3057–3060.
Takano, T. (1977a) *J. Mol. Biol. 110*, 537–568.
Takano, T. (1977b) *J. Mol. Biol. 110*, 569–584.
Teale, F. W. J. (1955) *Biochim. Biophys. Acta 35*, 289.
Varadarajan, R., Zewert, T., Gray, H. B., & Boxer, S. G. (1989a) *Science 243*, 69–72.
Varadarajan, R., Lambright, D. G., & Boxer, S. G. (1989b) *Biochemistry 28*, 3771–3781.

Characterization of Hydrophobic Cores in Apomyoglobin: A Proton NMR Spectroscopy Study†

Melanie J. Cocco and Juliette T. J. Lecomte*

Department of Chemistry, 152 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802

Received July 23, 1990; Revised Manuscript Received August 30, 1990

ABSTRACT: A proton nuclear magnetic resonance spectroscopic study of horse apomyoglobin was undertaken in order to define the regions of myoglobin that are and that are not structurally affected by the binding of the prosthetic group. It was found that, in spite of the poor spectral resolution, a number of spin systems could be identified by using standard correlated methods. Four clusters consisting mostly of hydrophobic residues were detected by nuclear Overhauser spectroscopy, two of which involved the tryptophan side chains. Extensive similarities to nuclear Overhauser spectroscopy data collected on the carbonmonoxy form of holomyoglobin suggested tentative assignments for several residues. It appeared that distinct cores of side chains on the distal side of the binding pocket and between the A, B, G, and H helices maintain the same packing as they do in holomyoglobin and apomyoglobin reconstituted with protoporphyrin IX.

Myoglobin is a monomeric heme protein of 17 kDa that in its native state is ca. 70% helical (Takano, 1977a,b; Evans & Brayer, 1988). The regular secondary structure is organized into eight helices, denominated A through H, with the heme contained in a cavity defined mostly by the E and F helices. Apomyoglobin (apoMb),[1] the proteinic component of myoglobin, is known to possess fewer residues in helical conformation than the holoprotein (Harrison & Blout, 1965; Kawamura-Konishi et al., 1988). Various physical techniques have contributed to a fragmentary description of the apoprotein structure. Fluorescence depolarization and lifetime measurements of sperm whale apoMb (Anderson et al., 1970) have shown that the environment of the two tryptophans is practically unchanged on going from the holo- to the apoprotein. These fluorescence results along with folding studies of human apoMb mutated at position 110 (Hughson & Baldwin, 1989) indicate that the A, B, G, and H helices are at least partially formed and dock properly through hydrophobic interactions. The presence of an extensive hydrophobic core is consistent with differential scanning calorimetry experiments reporting the cooperative thermal unfolding of apoMb (Griko et al., 1988).

In the preceding paper (Lecomte & Cocco, 1990), we presented an NMR study of apomyoglobin combined with protoporphyrin IX and confirmed the observation of Breslow et al. (1967) that the hydrophobic and ionic interactions involving the prosthetic group are sufficient to restore the native fold. In order to describe the conformational change triggered by heme (or protoporphyrin IX) insertion within the protein matrix, we initiated the characterization of the apoprotein by the same methods. We show here that limited but valuable information can be obtained on apoMb by standard two-di-

† Acknowledgement is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, for support of this work.

* To whom correspondence should be addressed.

[1] Abbreviations: apoMb, apomyoglobin; CD, circular dichroism; des-Fe Mb, des-iron myoglobin; COSY, correlated spectroscopy; 2D, two-dimensional; DQF-COSY, double-quantum-filtered COSY; holoMb, holomyoglobin; Mb, myoglobin; MbCO, carbonmonoxymyoglobin; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional proton nuclear Overhauser spectroscopy; TOCSY, total correlation spectroscopy; 2Q, two-quantum.

0006-2960/90/0429-11067$02.50/0

mensional procedures and describe regions of the protein that do not require prosthetic group interactions to organize into a structure resembling that of the holoprotein.

Materials and Methods

Apomyoglobin. Apomyoglobin was prepared from horse skeletal myoglobin (Sigma Type I) by using the methyl ethyl ketone method of Teale (1955) as described in the previous paper (Lecomte & Cocco, 1990).

NMR Samples. NMR samples were 2 mM in protein and kept at 4 °C when not in use. The pH of all samples was adjusted to 5.7 (except where noted) with a Beckman Φ71 pH meter equipped with an Ingold combination microelectrode. For ^2H_2O samples, exchange was accomplished by dissolving the sample in 1–2 mL of ^2H_2O, lyophilizing after a few hours, and repeating the process two to three times. The pH of ^2H_2O samples was not corrected for isotope effect and is reported as pH*. All NMR experiments were run at 298 K.

NMR Measurements. Proton NMR spectra were acquired at 500 MHz on a Bruker AM-500 spectrometer as reported for des-Fe Mb except that the sweep width could be reduced to 6024 Hz for DQF-COSY (Rance et al., 1983) and 2Q (Braunschweiler et al., 1983; Rance & Wright, 1986) experiments. All two-dimensional spectra were recorded in the phase-sensitive mode (Drobny et al., 1979; Marion & Wüthrich, 1983). Acquisition parameters were as follows: for DQF-COSY spectra, 2048 complex points, 64 transients per each of 512 t_1 increments, and ω_1 sweep width of 6024 Hz; and for 2Q spectra, 2048 complex points, 96 transients per each of 512 t_1 increments, ω_1 sweep width of 11 904 Hz, and τ_m = 22, 40, and 80 ms. The ω_2 sweep width was 12 500 Hz for Hahn-echo NOESY (Kumar et al., 1980; Bodenhausen et al., 1984; Davis, 1989) and TOCSY (Braunschweiler & Ernst, 1983) spectra. Other parameters were 4096 complex points, 96 transients per each of 512 t_1 increments, and ω_1 sweep width of 6024 Hz. Mixing times were between 110 and 150 ms (NOESY) and 75 ms with ca. 6-kHz MLEV-17 locking (TOCSY; Bax & Davis, 1985). Data were processed on a micro Vax II computer with FTNMR (Hare Research, Inc., Woodinville, WA) as described previously (Lecomte & Cocco, 1990). Chemical shifts are referenced to the water line at 4.76 ppm.

Results

Assignment Strategy. Since removal of the heme from horse Mb results in partial loss of structure (Crumpton & Polson, 1965; Kawamura-Konishi et al., 1988), it is not surprising that the one-dimensional ^1H NMR spectrum of horse apoMb contains mostly broad and overlapping lines. One-dimensional spectra of sperm whale apoMb (Griko et al., 1988) share those characteristics. However, in both apoproteins, the observed chemical shift dispersion is indicative of incomplete unfolding (Bundi & Wüthrich, 1979) and suggests the presence of some organized structure amenable to 2D NMR study. The limited resolution of amide and C^α proton resonances prevents the systematic application of homonuclear sequential backbone procedure. Thus, attempts were not made to define the secondary structure. Instead, selected side chains that are likely to form hydrophobic cores and be involved in tertiary interactions were targeted.

In the aromatic region of the spectrum, combining the results of DQF-COSY, TOCSY, and 2Q experiments in ^2H_2O and 90% ^1H_2O/10% ^2H_2O allowed for the identification of most spin systems at pH 5.7. For example, horse Mb contains two tryptophans (Dayhoff, 1976). All resonances of one of the tryptophan rings were located, and only one signal is missing for the other. Of the two tyrosines, one is resolved; the other appears to resonate in a crowded region of the spectrum. All seven phenylalanines and nine of the 11 histidines were found. The aliphatic region is more complex because of heavy spectral overlap. Nevertheless, some of the resonances belonging to the two methionines and several valines, leucines, and isoleucines could be recognized.

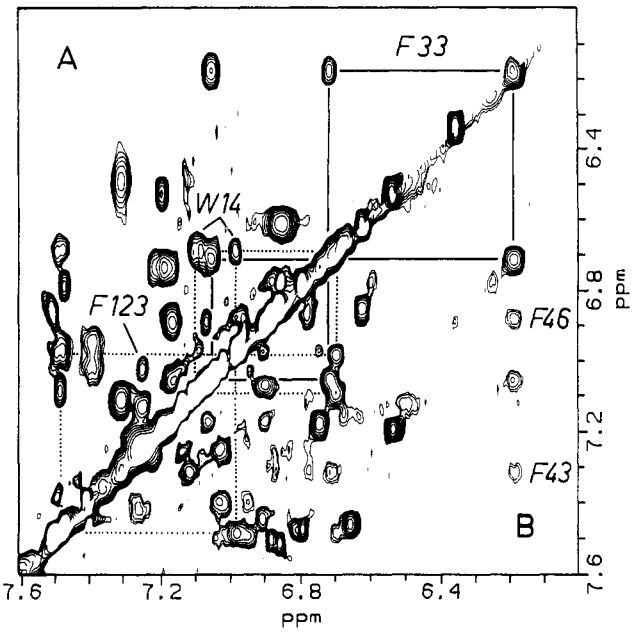

FIGURE 1: Combined TOCSY–NOESY spectra of a sample of horse apoMb in 90% ^1H_2O/10% ^2H_2O at pH 5.7 and 298 K. (A) TOCSY spectrum (τ_m = 75 ms). Two spin systems are outlined, Phe 33 (—) and Trp 14 (···). A well-resolved signal of Phe 123 is also marked. (B) NOESY spectrum (τ_m = 110 ms). Note the NOEs of Phe 33 to two other phenylalanine spin systems.

The number of identified side chains and the documented presence of native hydrophobic cores around the tryptophan residues justifies an attempt to proceed with further spectral analysis based on NOE data. As will become obvious, in several instances a comparison to holoMb and des-Fe Mb spectra yields a consistent interpretation of the apoprotein spectrum.

Trp 14, Val 17, His 24, Leu 115, Phe 123, and Met 131. Figure 1 presents TOCSY (part A) and NOESY (part B) spectra recorded on horse apoMb in 90% ^1H_2O/10% ^2H_2O at pH 5.7. A resolved tryptophan pattern is identified at 7.02, 7.13, 6.69, and 6.94 ppm. The phase-sensitive NOESY cross peaks included in Figure 2 (marked b) show that this side chain is near a group yielding a sharp signal at 2.16 ppm. No scalar connectivity can be found to this group, which is assigned as a methionine $C^\epsilon H_3$. The chemical shifts and NOE patterns for this Trp–Met pair are reminiscent of Figure 5 in the previous paper (Lecomte & Cocco, 1990) and analogous data on the holoprotein, where they arise from Trp 14 and Met 131. These represent the first secure assignments in the apoprotein. Chemical shifts and important NOEs are listed in Table I. The surroundings of Trp 14 in the apoprotein can be probed by analyzing other NOEs to which it gives rise in the aliphatic region upfield from the methionine methyl. Shown in Figure 3 is a relevant section of the phase-sensitive NOESY spectrum. Trp 14 is in tight dipolar contact with a residue resonating at −0.38 and 0.10 ppm. DQF-COSY and TOCSY connectivities demonstrate the side chain to be a valine. This valine is near a histidine imidazole ring with $C^\delta H$ and $C^\epsilon H$ resonances at 6.36 ppm and 8.00 ppm. By analogy to both holo and des-Fe Mb [Figure 8, previous paper (Lecomte & Cocco,

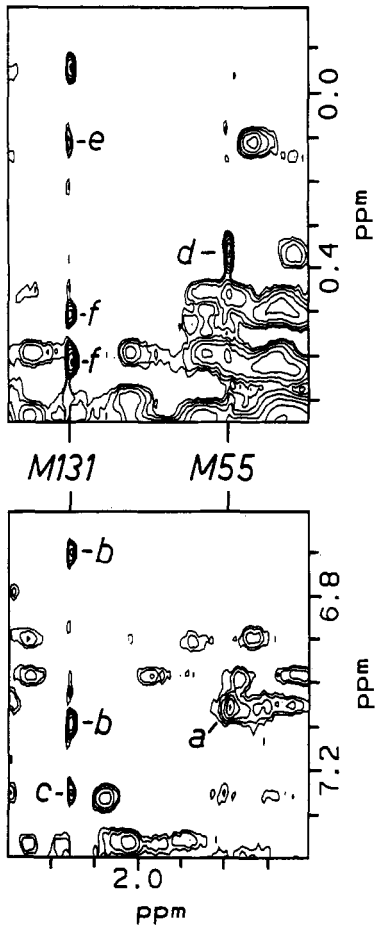

FIGURE 2: Region of the same phase-sensitive NOESY spectrum as shown in Figure 1 illustrating the connectivities involving the two Met C^eH_3 groups in horse apoMb. Cross peaks are labeled as follows: a, Met 55 to Phe 33; b, Met 131 to Trp 14; c, Met 131 to Phe 123; d, Met 55 to Leu 40; e, Met 131 to Val 17; f, Met 131 to Leu 115. This figure is to be compared with Figure 5 of the previous paper (Lecomte & Cocco, 1990).

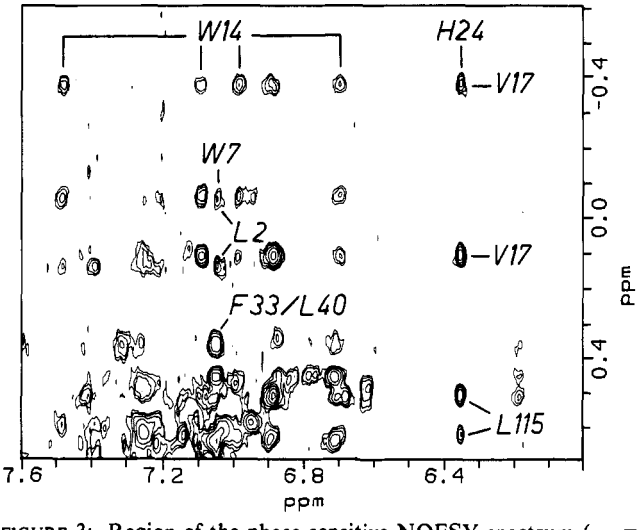

FIGURE 3: Region of the phase-sensitive NOESY spectrum (τ_m = 110 ms) of horse apoMb, recorded in ^2H_2O at pH* 5.7 and 298 K. Some of the connectivities involving Trp 7, Trp 14, His 24, and Phe 33 are marked. This figure is to be compared with Figure 8 of the previous paper (Lecomte & Cocco, 1990).

1990)], these are assigned as Val 17 and His 24. Figure 4 is to be compared with Figure 7 of the des-Fe Mb study (Lecomte & Cocco, 1990), which illustrates histidine–histidine dipolar contacts. His 24 is near another histidine side chain at 6.90 and 8.52 ppm (cross peak a). Assignment to His 119

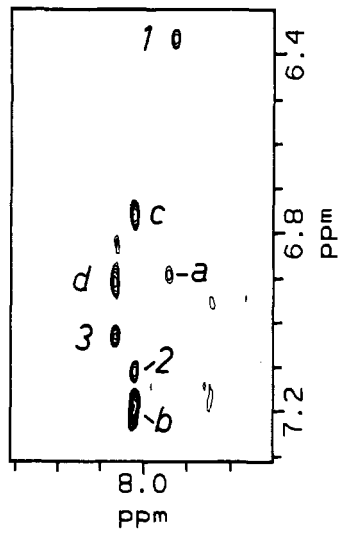

FIGURE 4: Region of a phase-sensitive NOESY spectrum showing the connectivities involving His residues 24, 36, 48, and 119 in horse apoMb recorded in ^2H_2O and at pH* 6.4, τ_m = 110 ms. The cross peaks are numbered as in Figure 7 of the previous paper (Lecomte & Cocco, 1990): 1, His 24 $C^\delta H/C^\epsilon H$; 2, His 36 $C^\delta H/C^\epsilon H$; 3, His 48 $C^\delta H/C^\epsilon H$; a, His 24 $C^\epsilon H$ to His 119 $C^\delta H$; b, His 36 $C^\epsilon H$ to Phe 106 $C^\epsilon H$, $C^\zeta H$; c, His 36 $C^\epsilon H$ to Phe 106 $C^\delta H$; d, His 48 $C^\epsilon H$ to Phe 46 $C^\zeta H$.

follows in keeping with a well-formed core in that region of the protein. Titration data indicate that between pH 5.6 and 8.4 the chemical shift of His 24 $C^\delta H$ remains unchanged.

The direct environment of Trp 14 contains an additional aliphatic residue, whose two methyl groups yield NOEs to Val 17, His 24 (Figure 3), His 119, and Met 131 (Figure 2, cross peak f). Although the side chain could not be traced completely, assignment is suggested to Leu 115 by comparison of chemical shift and NOE pattern with those of des-Fe Mb and MbCO. Figure 2 also illustrates that Met 131 lies near an aromatic ring (cross peak c); analysis of the corresponding scalar connectivities in Figure 1A and 2Q spectra point to a phenylalanine spin system at 7.14, 7.25, and 7.03 ppm. This residue matches well the description of Phe 123. NOEs are observed between Phe 123 and Leu 115, which completes the description of the core.

Leu 2 and Trp 7. The second tryptophan, Trp 7, also gives rise to a number of resonances close to their positions in the holoprotein spectrum. Trp 7 has NOEs to the methyl groups of an upfield-shifted side chain (Figure 3), assigned as Leu 2.

Phe 33, Leu 40, Phe 43, Phe 46, Leu 49, and Met 55. The second methionine (Met 55) C^eH_3 is recognizable in Figure 2 by its sharp NOEs along δ_2 = 1.79 ppm. Near this methyl group is found an aromatic ring with protons resonating at 6.72 ppm (cross peak a). The TOCSY spectrum of Figure 1A demonstrates the signal to arise from a phenylalanine ring. Similar NOEs are observed in the holo and des-Fe proteins between Phe 33 and Met 55 [Figure 5, previous paper (Lecomte & Cocco, 1990)]. Thus, the assignments are carried over to the apoprotein, with Met 55 resonating within 0.08 ppm of its holoprotein position and Phe 33 with shifted resonances indicative of external ring current. Phe 33 indeed has NOEs to two other phenylalanine spin systems, assumed to be Phe 43 and Phe 46 (Figure 1B). These two phenylalanines have their own set of diagnostic NOE connectivities: the C^δ protons of one of them present a strong NOE to the $C^\epsilon H$ of a histidine residue resonating at 8.18 ppm (Figure 4). This pair is consequently identified as Phe 46–His 48, leaving the assignment of Phe 43 to the ring resonances at 7.30, 7.10, and

Table I: Proton Resonance Assignments and NOEs of Horse Apomyoglobin[a]

residue	proton[b]	chemical shift	NOEs[c]	residue	proton[b]	chemical shift	NOEs[c]
Leu 2	$C^{\beta}H$	1.12		Phe 43	$C^{\epsilon}H$	7.10	F33$C^{\zeta}H$
	$C^{\gamma}H$	1.17			$C^{\zeta}H$	6.49	
	$C^{\delta 1}H_3$	0.13	W7$C^{\eta}H$	Phe 46	$C^{\alpha}H$	5.01	F46$C^{\delta}H$, H48$C^{\epsilon}H$, L49$C^{\gamma}H$, L49$C^{\delta 1}H_3$
	$C^{\delta 2}H_3$	−0.05	W7$N^{\epsilon}H$, W7$C^{\zeta 2}H$, W7$C^{\eta}H$				
Trp 7	$C^{\delta}H$	7.05			$C^{\delta}H$	6.87	F33$C^{\epsilon}H$, F33$C^{\zeta}H$, L40$C^{\delta 1}H_3$, L40$C^{\delta 2}H_3$, F43$C^{\delta}H$, F46$C^{\alpha}H$, H48$C^{\epsilon}H$
	$N^{\epsilon}H$	10.04	L2$C^{\delta 2}H_3$				
	$C^{\zeta 2}H$	7.40	L2$C^{\delta 2}H_3$				
	$C^{\eta}H$	7.01	L2$C^{\delta 1}H_3$, L2$C^{\delta 2}H_3$		$C^{\epsilon}H$	6.77	F43$C^{\delta}H$
	$C^{\zeta 3}H$	6.95			$C^{\zeta}H$	6.81	
Trp 14	$C^{\eta}H$	6.99		His 48	$C^{\delta}H$	7.13	
	$N^{\epsilon}H$	10.28			$C^{\epsilon}H$	8.18	F46$C^{\alpha}H$, F46$C^{\delta}H$, L49$C^{\gamma}H$, L49$C^{\delta 1}H_3$, L49$C^{\delta 2}H_3$
	$C^{\zeta 2}H$	7.58	V17$C^{\gamma 1}H_3$				
	$C^{\eta}H$	6.98	V17$C^{\gamma 1}H_3$	Leu 49	$C^{\gamma}H$	1.14	F46$C^{\alpha}H$, H48$C^{\epsilon}H$
	$C^{\zeta 3}H$	6.69	V17$C^{\gamma 1}H_3$, V17$C^{\gamma 2}H_3$, L115$C^{\delta 2}H_3$, M131$C^{\epsilon}H_3$		$C^{\delta 1}H_3$	0.98	F46$C^{\alpha}H$, H48$C^{\epsilon}H$
	$C^{\epsilon 2}H$	7.09	V17$C^{\gamma 1}H_3$, V17$C^{\gamma 2}H_3$, M131$C^{\epsilon}H_3$		$C^{\delta 2}H_3$	0.70	H48$C^{\epsilon}H$
Val 17	$C^{\alpha}H$	1.73		Met 55	$C^{\epsilon}H_3$	1.79	F33$C^{\epsilon}H$, L40$C^{\delta 1}H_3$, L40$C^{\delta 2}H_3$
	$C^{\beta}H$	0.93		Tyr 103	$C^{\delta}/C^{\epsilon}H$	6.98	F106$C^{\delta}H$, F106$C^{\epsilon,\zeta}H$
	$C^{\gamma 1}H_3$	−0.38	W14$C^{\zeta 2}H$, W14$C^{\eta}H$, W14$C^{\zeta 3}H$, W14$C^{\epsilon 3}H$, H24$C^{\delta}H$, H24$C^{\epsilon}H$, L115$C^{\delta 1}H_3$, L115$C^{\delta 2}H_3$		$C^{\epsilon}/C^{\delta}H$	6.90	F106$C^{\delta}H$, F106$C^{\epsilon,\zeta}H$
				Phe 106	$C^{\delta}H$	6.75	H36$C^{\delta}H$, H36$C^{\epsilon}H$, Y103$C^{\delta}H$, Y103$C^{\epsilon}H$
	$C^{\gamma 2}H_3$	0.10	W14$C^{\epsilon 3}H$, W14$C^{\zeta 3}H$, H24$C^{\delta}H$, H24$C^{\epsilon}H$, L115$C^{\delta 1}H_3$, L115$C^{\delta 2}H_3$, H119$C^{\delta}H$		$C^{\epsilon,\zeta}H$	7.18	H36$C^{\delta}H$, H36$C^{\epsilon}H$, Y103$C^{\delta}H$, Y103$C^{\epsilon}H$
His 24	$C^{\delta}H$	6.36	V17$C^{\gamma 1}H_3$, V17$C^{\gamma 2}H_3$, L115$C^{\delta 1}H_3$, L115$C^{\delta 2}H_3$, H119$C^{\delta}H$	Leu 115	$C^{\gamma}H$	1.69	
					$C^{\delta 1}H_3$	0.49	V17$C^{\gamma 1}H_3$, V17$C^{\gamma 2}H_3$, H24$C^{\delta}H$, H119$C^{\delta}H$, F123$C^{\epsilon}H$, M131$C^{\epsilon}H_3$
	$C^{\epsilon}H$	8.00	V17$C^{\gamma 1}H_3$, V17$C^{\gamma 2}H_3$, H119$C^{\delta}H$, H119$C^{\epsilon}H$				
Phe 33	$C^{\delta}H$	7.05	L40$C^{\delta 1}H_3$, L40$C^{\delta 2}H_3$		$C^{\delta 2}H_3$	0.61	W14$C^{\zeta 3}H$, V17$C^{\gamma 1}H_3$, V17$C^{\gamma 2}H_3$, H24$C^{\delta}H$, H119$C^{\delta}H$, F123$C^{\epsilon}H$, M131$C^{\epsilon}H_3$
	$C^{\epsilon}H$	6.72	L40$C^{\delta 1}H_3$, L40$C^{\delta 2}H_3$, F43$C^{\delta}H$, F46$C^{\delta}H$, M55$C^{\epsilon}H_3$				
	$C^{\zeta}H$	6.19	F43$C^{\delta}H$, F43$C^{\epsilon}H$, F46$C^{\delta}H$				
His 36	$C^{\delta}H$	7.11	F106$C^{\delta}H$, H48$C^{\epsilon}H$,	His 119	$C^{\delta}H$	6.90	V17$C^{\gamma 2}H_3$, H24$C^{\delta}H$, H24$C^{\epsilon}H$, L115$C^{\delta 1}H_3$, L115$C^{\delta 2}H_3$
	$C^{\epsilon}H$	8.01	F106$C^{\delta}H$, F106$C^{\epsilon,\zeta}H$				
Leu 40	$C^{\beta}H$	1.63			$C^{\epsilon}H$	8.52	H24$C^{\epsilon}H$
	$C^{\beta'}H$	1.38		Phe 123	$C^{\delta}H$	7.14	M131$C^{\epsilon}H_3$
	$C^{\gamma}H$	1.24			$C^{\epsilon}H$	7.25	L115$C^{\delta 1}H_3$, L115$C^{\delta 2}H_3$, M131$C^{\epsilon}H_3$
	$C^{\delta 1}H_3$	0.37	F33$C^{\delta}H$, F33$C^{\epsilon}H$, F43$C^{\delta}H$, F46$C^{\delta}H$, M55$C^{\epsilon}H_3$		$C^{\zeta}H$	7.03	
	$C^{\delta 2}H_3$	0.35	F33$C^{\delta}H$, F33$C^{\epsilon}H$, F43$C^{\delta}H$, F46$C^{\delta}H$, M55$C^{\epsilon}H_3$	Met 131	$C^{\epsilon}H_3$	2.16	W14$C^{\epsilon 3}H$, W14$C^{\zeta 3}H$, L115$C^{\delta 1}H_3$, L115$C^{\delta 2}H_3$, F123$C^{\delta}H$, F123$C^{\epsilon}H$
Phe 43	$C^{\delta}H$	7.30	F33$C^{\epsilon}H$, F33$C^{\zeta}H$, L40$C^{\delta 1}H_3$, L40$C^{\delta 2}H_3$, F46$C^{\delta}H$, F46$C^{\epsilon}H$				

[a] Chemical shifts in parts per million with respect to the water line at 4.76 ppm, determined at pH* 5.7, 298 K. [b] The denomination 1 or 2 for methyl groups of valines and leucines is chosen according to holoprotein solid-state structure predictions (Hanson & Schoenborn, 1981). [c] Only the NOEs essential to assignment are listed. Determined at pH* 5.7, 298 K, with mixing time of 110 ms.

6.59 ppm. NOEs between Phe 43 and Phe 46 are also observed, which confirms that the three residues form a tight cluster. Strong NOEs from Met 55 and Phe 33 and weak NOEs from Phe 43 and Phe 46 to a leucine side chain suggest that Leu 40 is involved in this hydrophobic core as it is in the holoprotein. The Phe 46 C^{δ} protons have a strong NOE to a C^{α} proton at 5.01 ppm, which is in close contact with His 48 $C^{\epsilon}H$. In MbCO and des-Fe Mb, it is assigned as Phe 46 $C^{\alpha}H$ in agreement with the solid-state structure geometry. To reinforce this view, His 48 $C^{\epsilon}H$ and Phe 46 $C^{\alpha}H$ both are in dipolar contact with an isopropyl-containing side chain tentatively assigned as Leu 49.

His 36, Tyr 103, and Phe 106. Inspection of Figure 4 and Figure 1A provides another set of histidine–phenylalanine NOEs. In this instance, the phenylalanine is adjacent to an aromatic residue whose DQF-COSY and 2Q connectivities are unequivocally those of a tyrosine. The NOE network is also seen in MbCO and des-Fe Mb [Dalvit & Wright, 1987; Figure 7, previous paper (Lecomte & Cocco, 1990)]; it enables us to assign the signals of His 36, Phe 106, and Tyr 103.

Discussion

In spite of the partial unfolding caused by prosthetic group removal, several side-chain signals are reasonably well resolved and can be identified in the 2D spectra of apoMb. Because of the size of the apoprotein and the relatively low content of secondary structure, our first approach concentrates on those residues readily recognized and on their environment. Although signal identification is unambiguous, the assignments reached here are based on a pronounced correspondence between the NOE patterns observed in the holoprotein and the apoprotein and are consequently tentative. It is noted, however, that the interpretation of the results resting on a structure similar to that of the holoprotein is entirely self-consistent.

The two tryptophans constitute reliable starting points for spectral analysis. They belong to the A helix and are normally in contact with the G, B, and H helices (Takano, 1977a). Many of the stabilizing interactions among the residues of the A–B–G–H interface are detected through NOE spectroscopy. Figure 5 summarizes the connectivities. Trp 14 (A12), Val 17 (A15), Leu 115 (G16), Phe 123 (GH5), and Met 131 (H8) form a tightly packed hydrophobic cluster within apoMb. The chemical modification of Trp 7 and Trp 14 to the more polar oxindolylalanine destabilizes significantly the holoprotein (Radding, 1987) and reaffirms the active role played by these side chains. It appears that in the apoprotein, too, they contribute to the fastening of structurally and thermodynamically important hydrophobic cores.

Recent studies on human apomyoglobin and mutants (Hughson & Baldwin, 1989) have suggested that the stability of "native" apoMb and its acid unfolding intermediate is sensitive to the integrity of the G–B helix pairing site. Mutation of Cys 110 (G11) in the G–B interface has similar consequences for the folding of both the holo- and the apoprotein. Although residue 110 (an alanine in horse Mb) was not assigned in the spectrum of the apoprotein, the environment

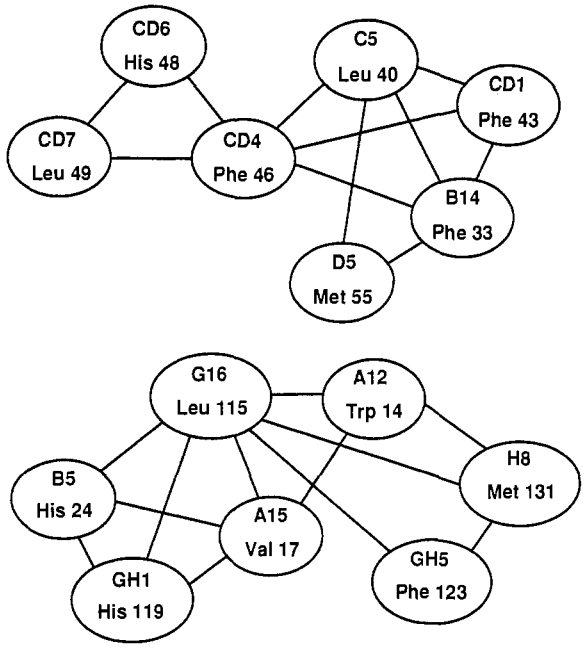

FIGURE 5: Schematic diagram of the observed NOE connectivities in two extensive hydrophobic cores of horse apomyoglobin. More detailed information is listed in Table I.

of other side chains was characterized in the G segment. Leu 115 (G16), a residue of the Trp 14 cluster, is correctly docked against His 24 (B5), while NOEs between Phe 106 (G7) and His 36 (C1) are consistent with the proper orientation of the extremities of the B and G helices. Thus, the spatial relationship of residues in the G–B interface is maintained farther from Ala 110.

Of structural significance are the spectral properties of His 24. His 24 displays a number of NOEs to assigned side chains, in particular to His 119 (GH1). The ring protons of His 24 resonate at chemical shifts close to their values in MbCO and des-Fe Mb. As in the latter two forms of the protein, the $C^{\delta}H$ chemical shift is independent of pH, indicating like states of protonation. In the holoprotein, this is interpreted as a manifestation of a hydrogen bond between His 24 and His 119. The 24–119 hydrogen bond, which has been proposed to stabilize the GH corner (Takano, 1977a), is therefore present in the apoprotein as well. In the H helix, only the terminal methyl group of Met 131 (H8) could readily be found in the spectrum. In connection with potential native structure in this segment, it is interesting to note that the peptide fragment comprising residues 132 (H9) to 153 (C-terminus) has been shown by NMR spectroscopy to have helix-forming propensity (Waltho et al., 1990).

As described by Kawamura-Konishi et al. (1988), the first step in the formation of the holoprotein from the apoprotein and hemin dicyanide is the clamping of the prosthetic group without coordination to the proximal histidine. Des-Fe Mb studies (Breslow et al., 1967; Lecomte & Cocco, 1990) also indicate that the hydrophobic contacts established by the porphyrin ring are crucial and sufficient to complete the folding. The advantage of NMR spectroscopy over the other techniques so far applied to apoMb lies in the ability to describe in detail the environment of a variety of hydrophobic residues, especially phenylalanines. In the NOESY spectra, a set of three interacting phenylalanine rings is recognized, which can be assigned as Phe 33 (B14), Phe 43 (CD1), and Phe 46 (CD4) by analogy to the holoprotein. Nearby residues Leu 40 (C5) and Met 55 (D5) are packed along in the same region (Figure 5). This hydrophobic cluster extends over part of the CD corner and, with Phe 43, defines a wide distal edge for the heme cavity. Phe 43 is a rigorously conserved residue (Bashford et al., 1987) in contact with the heme group in the holoprotein. The preorganization of this region of apoMb into a holoprotein structure may aid in the process of prosthetic group recognition and binding, not only through the formation of some hydrophobic surface but also through the proper positioning of residue CD3, which is thought to interact with a propionate side chain in the first encounter of apoMb with hemin (La Mar et al., 1989). Inspection of the holoMb structure shows that Tyr 103 (G4), Phe 106 (G7), and His 36 (C1), which pack together in the apoprotein, constitute an extension of this heme cavity edge. With Tyr 103 normally in contact with the 4-vinyl side chain, it seems again that the apoprotein offers important native features to the incoming prosthetic group.

CONCLUSIONS

Four clusters of residues that do not require the prosthetic group (hemin or protoporphyrin IX) to fold into a stable conformation have been identified within apoMb. They are likely to provide some stability and an organized structure capable of recognizing the heme group, which introduces the additional hydrophobic interactions necessary to complete the folding. Two of the detected cores contain the tryptophan residues; their presence is consistent with data obtained through optical methods and in folding experiments. Comparison to the holoprotein spectra suggests strongly that these hydrophobic clusters form elements of structure similar to those observed in the holoprotein, and they are therefore thought to report on the A, B, G, and H segments of myoglobin. The other two cores share spectral characteristics with regions of the holoprotein nearer to the heme binding site. They appear to form the border of the cavity, by the CD corner and the G edge. Additional work is in progress to confirm the spectral assignments independently of the holoprotein data and to describe the position of residues of the E and F segments that normally constitute the heme pocket.

Registry No. Trp, 73-22-3; His, 71-00-1.

REFERENCES

Anderson, S. R., Brunori, M., & Weber, G. (1970) *Biochemistry 9*, 4723–4729.

Bashford, D., Chothia, C., & Lesk, A. M. (1987) *J. Mol. Biol. 196*, 199–216.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983) *Mol. Phys. 48*, 535–560.

Breslow, E., Koehler, R., & Girotti, A. W. (1967) *J. Biol. Chem. 242*, 4149–4156.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285–297.

Crumpton, M. J., & Polson, A. (1965) *J. Mol. Biol. 11*, 722–729.

Dalvit, C., & Wright, P. E. (1987) *J. Mol. Biol. 194*, 313–327.

Davis, D. G. (1989) *J. Magn. Reson. 81*, 603–607.

Dayhoff, M. O. (1976) *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 2, National Biomedical Research Foundation, Washington, DC.

Drobny, G., Pines, A., Sinton, S., Weitekamp, D. P., & Wemmer, D. (1979) *Symp. Faraday Soc. 13*, 49–55.

Evans, S. V., & Brayer, G. D. (1988) *J. Mol. Biol. 263*, 4263–4268.

Griko, Y. V., Privalov, P. L., Venyaminov, S. Y., & Kutyshenko, V. P. (1988) *J. Mol. Biol. 202*, 127–138.
Hanson, J. C., & Schoenborn, B. P. (1981) *J. Mol. Biol. 153*, 117–146.
Harrison, S. C., & Blout, E. R. (1965) *J. Biol. Chem. 240*, 299–303.
Hughson, F. M., & Baldwin, R. L. (1989) *Biochemistry 28*, 4415–4422.
Kawamura-Konishi, Y., Kihara, H., & Suzuki, H. (1988) *Eur. J. Biochem. 170*, 589–595.
Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.
La Mar, G. N., Pande, U., Hauksson, J. B., Pandey, R. K., & Smith, K. M. (1989) *J. Am. Chem. Soc. 111*, 485–491.
Lecomte, J. T. J., & Cocco, M. J. (1990) *Biochemistry* (preceding paper in this issue).
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Privalov, P. L., Griko, Y. V., Venyaminov, S. Y., & Kutyshenko, V. P. (1986) *J. Mol. Biol. 190*, 487–498.
Radding, J. A. (1987) *Biochemistry 26*, 3530–3536.
Rance, M., & Wright, P. E. (1986) *J. Magn. Reson. 66*, 372–378.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Würthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Takano, T. (1977a) *J. Mol. Biol. 110*, 537–568.
Takano, T. (1977b) *J. Mol. Biol. 110*, 569–584.
Teale, F. W. J. (1955) *Biochim. Biophys. Acta 35*, 289.
Waltho, J. P., Feher, V. A., & Wright, P. E. (1990) in *Current Research in Protein Chemistry* (Villafranca, J. J., Ed.) Section III, pp 283–293, Academic Press, New York.

Refolding and Aggregation of Bovine Carbonic Anhydrase B: Quasi-Elastic Light Scattering Analysis†

Jeffrey L. Cleland and Daniel I. C. Wang*
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
Received April 16, 1990; Revised Manuscript Received August 28, 1990

ABSTRACT: Bovine carbonic anhydrase B (CAB) is chosen as the model protein to study the phenomenon of protein aggregation, which often occurs during the refolding process. Refolding of CAB from 5 M GuHCl has been observed by quasi-elastic light scattering (QLS), which confirms the formation of a molten globular protein structure as reported previously [Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J., & Ptitsyn, O. B. (1987) *FEBS Lett. 224*, 9–13]. QLS analysis reveals the formation of multimeric species prior to precipitation. Activity and cross-linking studies have confirmed the presence of inactive multimeric protein species. The dimer formation has been determined to be the initiating step in the aggregation of CAB during refolding. Activity studies have indicated that the first intermediate observed in the refolding pathway of CAB aggregates to form the inactive dimer. The rate of formation of the dimer has a stoichiometric dependence on the final protein concentration. The dimer formation rate is a function of the final guanidine hydrochloride (GuHCl) concentration to the inverse 6.7 power, which correlates well with the binding of GuHCl to the native protein in 0.60–0.80 M GuHCl. These rate dependencies require the refolding of CAB to be performed at high GuHCl concentrations (1 M GuHCl) and low protein concentrations (less than 1 mg/mL) to avoid the formation of aggregates. Alternatively, refolding can be performed by allowing the first intermediate to form the second intermediate prior to further dilution or dialysis. The aggregation of a hydrophobic first intermediate species is likely to be common to the refolding of other molten globular proteins.

Purification of recombinant proteins expressed in *Escherichia coli* often requires the solubilization and renaturation of proteins that are expressed as insoluble inclusion bodies (Marston, 1986). During the refolding of proteins from denatured states, aggregation of partially refolded protein occurs resulting in decreased recovery of the native protein. Early studies of protein refolding by Anfinsen and others have indicated that many proteins aggregate during refolding from unfolded states (Anfinsen & Haber, 1961; Epstein & Goldberger, 1963). The aggregation of protein during refolding is strongly dependent on the final protein and denaturant concentration (Zettlmeissl et al., 1979). Therefore, proteins are usually refolded at very low protein concentrations (μg/mL) or high denaturant concentrations (1 M guanidine hydrochloride or 4 M urea) to avoid aggregation.

To investigate the phenomenon of protein aggregation during refolding, the study of the model protein system, bovine carbonic anhydrase B (CAB),[1] is undertaken since CAB has a well-characterized refolding pathway (Stein & Henkens, 1978; Doligkh et al., 1984; Semisotnov et al., 1987). During refolding of CAB from its denatured state in 5 M GuHCl, the protein will rapidly form a compact molten globule structure with exposed hydrophobic clusters. These clusters will then

† This work was supported by National Science Foundation Grant CDR-88-03014.
* To whom correspondence should be addressed.

[1] Abbreviations: CAB, bovine carbonic anhydrase B; GuHCl, guanidine hydrochloride; QLS, quasi-elastic light scattering; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; pNPA, *p*-nitrophenol acetate; pNP, *p*-nitrophenol; DMS, dimethyl suberimidate dihydrochloride; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; CONTIN, constrained regularization method; SDP, size distribution processor.

0006-2960/90/0429-11072$02.50/0