PROTEINS: Structure, Function, and Genetics 25:267–285 (1996)

RESEARCH ARTICLES

The Native State of Apomyoglobin Described by Proton NMR Spectroscopy: The A-B-G-H Interface of Wild-Type Sperm Whale Apomyoglobin

Juliette T.J. Lecomte, Yung-Hsiang Kao,* and Melanie J. Cocco†
Department of Chemistry and the Center for Biomolecular Structure and Function, Pennsylvania State University, University Park, Pennsylvania 16802

ABSTRACT **Proton nuclear magnetic resonance spectroscopy was applied to sperm whale apomyoglobin to describe the conformation adopted by the protein under native conditions. The study focused on the A-B-G-H interface, a region known to form a compact subdomain in the apoprotein (Hughson and Baldwin, Biochemistry 28:4415–4422, 1989). Two histidine residues located in this subdomain, His24 and His119, interact and are thought to play a role in the acid denaturation process (Barrick et al., J. Mol. Biol. 237:588–601, 1994). A stable double mutant at these positions (His24Val/His119Phe sperm whale apomyoglobin) was compared with wild-type apomyoglobin. The amino acid replacements result in chemical shift perturbations near the mutations, in particular in the AB interhelical region, and in a deceleration of backbone amide hydrogen exchange in the B helix from position 27 to position 33. The double mutant data were used to expand and confirm the wild-type spectral analysis. Signals from the D helix were identified that demonstrate the formation of holoprotein-like structure. The assigned wild-type nuclear Overhauser effects, although in small number, were sufficient to construct a model of the compact subdomain of the apoprotein. This was achieved by using the structure of the holoprotein and restraining it with the geometrical information on the apoprotein in a simulated annealing procedure. The experimental restraints define a low-resolution model of the A-B-G-H interface in apomyoglobin.**

Key words: myoglobin, histidine, hydrophobic core

INTRODUCTION

Myoglobin, the *b* hemoprotein responsible for oxygen storage in red muscles, is a small, stable, and readily purified protein. It functions as a monomer and contains a single heme group. Interest in apoMb (the protein without the heme group) has grown with the investigation of its thermodynamic properties. "Native" apoMb undergoes a two-state cooperative heat denaturation. The changes in enthalpy and heat capacity accompanying this denaturation are typical of a one-domain water-soluble *globular* protein with specific hydrophobic interactions.[1,2] ApoMb is also one of the few proteins for which a nearly complete, cooperative, and reversible cold denaturation can be detected.[1] The maximal stability of apoMb (pH 5.0 and 304 K) is 11 kJ/mol, lower than that of holoMb by ~20 kJ/mol.[1]

Early circular dichroism studies suggested that partial unfolding accompanies heme removal.[3,4] Although the helical content of the apoprotein remains only approximately determined,[5] a value of about 55% helicity has been generally accepted for native apoMb, and compared with a value of about 80% in the holoprotein. The partial unfolding is reversible: free heme is rapidly captured by the native apopro-

Abbreviations: ANS, 8-anilino-1-naphthalenesulfonic acid; apoMb, apomyoglobin; CD, circular dichroism; DIPSI, decoupling in the presence of scalar interactions; DG, distance-geometry; DQF-COSY, two-dimensional double-quantum-filtered correlated spectroscopy; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulfonate; EPR, electron paramagnetic resonance; holoMb, holomyoglobin; HyTEMPO, 4-hydroxy-2,2,6,6,-tetramethyl piperidinyl-1-oxy; Mb, myoglobin; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser spectroscopy; rms, root mean square; SA, simulated annealing; TOCSY, total correlation spectroscopy; TPPI, time proportional phase incrementation.

Received August 4, 1995; revision accepted December 21, 1995.

Address reprint requests to Juliette T.J. Lecomte, Department of Chemistry, 152 Davey Laboratory, Pennsylvania State University, University Park, PA 16802.
*Current address: Department of Analytical Chemistry, Genentech, Inc., South San Francisco, CA 94080.
†Current address: Department of Biochemistry, University of Connecticut Health Center, Farmington, CT 06032.

tein,[6-8] and all structure is recovered upon heme binding.[3] These properties make apoMb a protein with a characterizable thermodynamic state, neither fully folded nor fully unfolded, yet existing under native conditions. Recent reports on apoMb have included predictions of pK_a values,[9] calculated trajectories upon heme removal,[10,11] molecular dynamic simulations of isolated helices,[12,13] and stability observations.[14,15] The lack of an appropriate structural basis (i.e., the structure of the apoprotein itself as opposed to that of the related but not precisely pertinent holoprotein) for the interpretation of these studies and an interest in marginally stable native states prompted us to extend our current description of apoMb in solution.

The proton NMR spectrum of native apoMb is that of a species with detectable tertiary interactions, possibly confined in a few regions of the structure. The most readily identified of these interactions are holoprotein-like and involve several aliphatic residues.[16] Previous NMR experiments also revealed secondary structure in the A, B, C, E, G, and H regions of native apoMb[17] (Fig. 1A illustrates the structure of holoMb and the helix nomenclature). Interhelix contacts (A-H, B-G) support the formation of a compact subdomain in the A-B-G-H interface. Pulse-labeling NMR experiments,[5,18] fluorescence of the two tryptophans in the A helix,[19] and thermodynamic consequences of mutations[20] all concur independently with this view.

When poor resolution and conformational exchange hinder the analysis of proton data, mutant proteins offer one recourse. The selection of a useful Mb mutant was guided by the acid response of apoMb.[5,20-24] At low pH, it forms a stable species that can be further stabilized by salt.[25,26] This "I" state has the physical characteristics of a molten globule,[21,27] whereas the native state of apoMb does not.[2] Baldwin and coworkers[24] performed thermodynamic studies of the I state and the native state of apoMb mutants in an attempt to identify the residues responsible for the distinctive acid response. Among the 12 histidines that can be mutated in systematic explorations of pH effects, two participate in the A-B-G-H interface: His24 (B5) and His119 (GH1). Their imidazole rings are buried and within hydrogen-bonding distance of each other in the holoprotein[28-30] (Fig. 1) and native apoMb.[30] The single mutations His24Val and His119Phe are moderately destabilizing[24]; on the other hand, the double mutation His24Val–His119Phe stabilizes native and intermediate states.[24] It also decreases the apparent number of protons linked to the N-to-I transition by one unit.[24] Structural information on native His24Val–His119Phe apoMb is desirable to complement these thermodynamic data. The stable double mutant was therefore chosen for NMR investigations of the A-B-G-H interface in apomyoglobin.

In an earlier study of the native apoprotein state,[17] we proposed as a testable hypothesis that unfolding caused by heme removal might occur in the D helix, the beginning of the E helix, the F helix, and the carboxy-end of the H helix. In this paper, we expand our proton spectral assignments to include signals from the D helix and refine the conformational description of the A-B-G-H compact subdomain with the help of His24Val–His119Phe apoMb data. We construct a model of the readily detectable elements of tertiary and secondary structure to summarize our knowledge of the stable features of apomyoglobin.

MATERIALS AND METHODS

Protein Samples

Wild-type sperm whale myoglobin was purchased from Sigma and used without further purification. Mutants of sperm whale myoglobin were prepared from myoglobin genes provided by Drs. Baldwin, Barrick, and Hughson (Stanford University). The growth, isolation, and purification procedures were per the standard protocols detailed by these authors.[24] All apoMb samples were prepared with the cold butanone method as described previously[16]; final protein concentrations were 1 to 2 mM for two-dimensional NMR experiments. HyTEMPO was purchased from Aldrich. A 1 M HyTEMPO solution was prepared and an appropriate microliter volume was added to the 0.5 mL protein solution.

NMR Experiments

NMR spectra were acquired at a proton frequency of 500 MHz on a Bruker AM-500 spectrometer at 304 or 298 K. Two-dimensional data were collected in the quadrature mode and with TPPI[31,32] by using parameters already established.[16,17,30] The experiments include DQF-COSY,[33] Hahn-echo NOESY,[34-36] and TOCSY.[37,38] A DIPSI-2 pulse train[39] (with a 90° pulse of ~25 μs) was used for spin-locking in the latter; the train was repeated to achieve mixing times ranging between 35 and 70 ms. Mixing times in NOESY experiments were 100 or 110 ms. In addition to these experiments, NOESY and clean TOCSY[40] data were collected in 90% H_2O/10% ^2H_2O with pulsed field gradient on a Bruker AMX2-500 spectrometer. For these experiments, a self-shielded z-gradient probe and WATERGATE sequence[41] were used to suppress the water signal. Acquisition parameters for the clean TOCSY data collection were as for the TOCSY data, with a 90° locking pulse of 23 μs and a delay for relaxation compensation corresponding to a 143.89° locking pulse.[40]

A total of 96 transients was collected for each of the 480 or 512 t_1 values. Both TOCSY and NOESY AM-500 experiments were sine-modulated in the t_1 domain,[42] and the t_1 delay was compensated for pulse length.[43] Spectral widths were as follows:

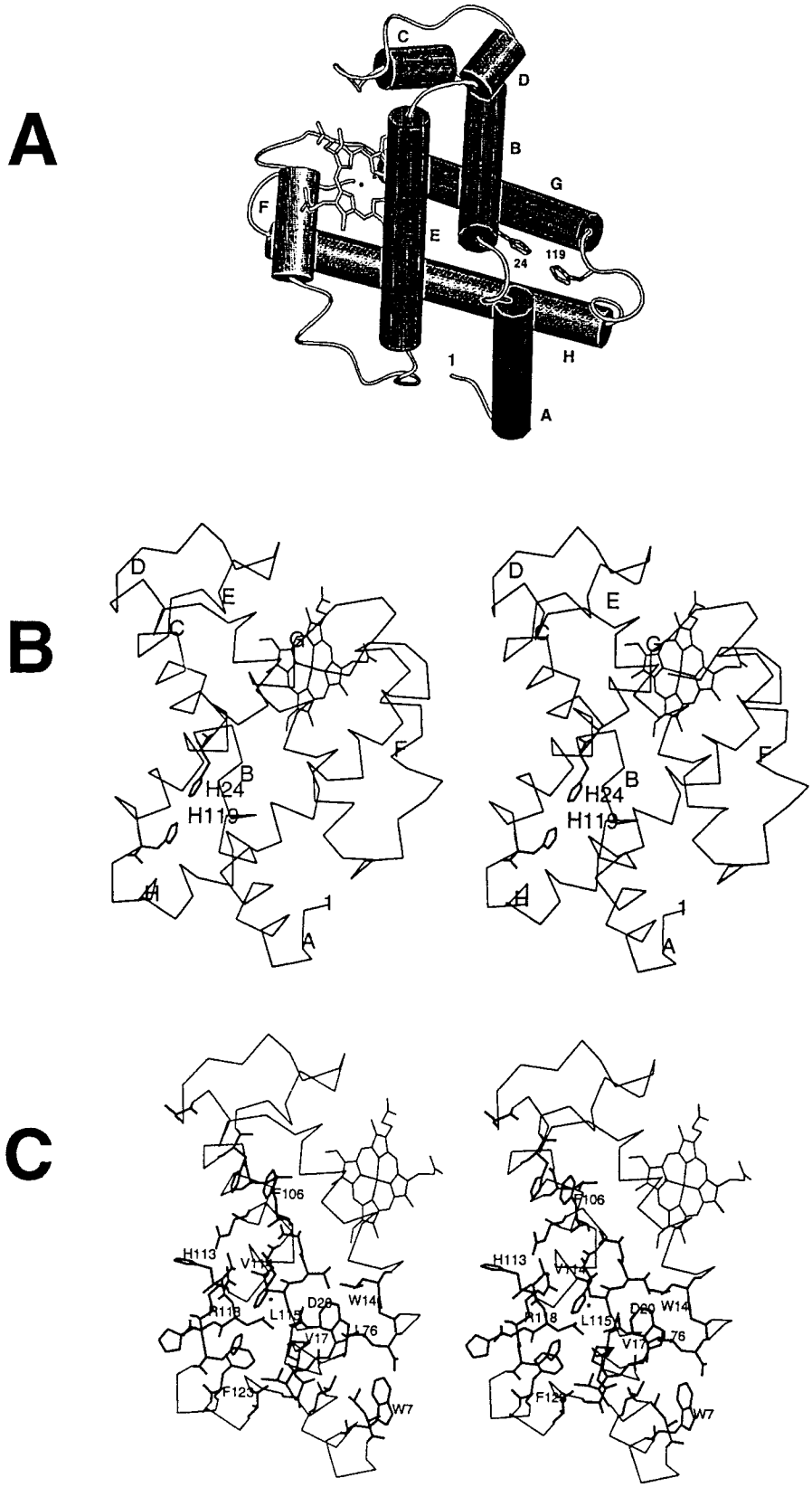


Fig. 1. **A:** Schematic diagram of sperm whale metaquomyoglobin.[44] The N-terminus of the protein is marked with "1." The eight helices are labeled with the letter code and are defined as follows: A1–A16: Ser3–Glu18; B1–B16: Asp20–Ser35; C1–C7: His36–Lys42; CD1–CD8: Phe43–Leu49; D1–D7: Thr51–Ala57; E1–E20: Ser58–Lys77; EF1–EF8: Lys78–Glu85; F1–F9: Leu86–Ala94; FG1–FG5: Thr95–Ile99; G1–G19: Pro100–Arg118; GH1–GH5: His119–Phe123; H1–H26: Gly124–Leu149. The side chains of His24 and His119 are included. The orientation of the imidazole rings is consistent with the formation of a hydrogen bond.[28] The double mutant studied here has a valine at position 24 (B5) and a phenylalanine at position 119 (GH1), as found in the β chain of human hemoglobin. **B:** Stereo view of the Cα trace with the 24 and 119 side chains. **C:** Portion of the structure in the same orientation as in B, with some of the side chains discussed in the text. The asterisk indicates His24.

DQF-COSY, 6,250 Hz in both t_2 (2,048 complex points) and t_1 (470–512 increments); NOESY and TOCSY, 12,500 Hz in t_2 (4,096 complex points) and 6,250 Hz in t_1 (512 increments). Only the central 6,250 Hz of the transformed F2 dimension were kept in the final matrix. All two-dimensional proton data were zero-filled in the indirect dimension to give a data matrix of 2,048 × 2,048 real points. Data were processed using the program FELIX from Biosym Technologies running on an Indigo XZ4000 Silicon Graphics workstation. A squared sine-bell window with a $\pi/4$ shift was applied in the F1 and F2 dimensions to most data sets. NOESY data for the purpose of volume integration were subjected to a $\pi/3$ shift in the F1 dimension. Chemical shifts are referenced through DSS to the water line at 4.76 ppm (298 K) or 4.695 ppm (304 K). pH* indicates electrode reading without correction for isotope effects.

Ring Current Shift Calculations

Ring current shift calculations were performed with the refined neutron diffraction structure of sperm whale MbCO (PDB file 2mb5)[28] and the X-ray structure of metaquomyoglobin (PDB file 4mbn).[44,45] The ring current shifts induced by aromatic side chains were evaluated as reported elsewhere[46,47] with a FORTRAN program employing the ring current intensities listed by Ösapay and Case.[48] The Haigh-Mallion 8-loop method[47] was applied in the calculations of the heme ring current shift.

Structural Calculations

The NOESY data in ^2H_2O and H_2O provided a total of 457 interproton restraints to calculate a three-dimensional model of the A-B-G-H interface. Of these, 276 were interresidue. A total of 101 restraints were long range ($|i - j| > 5$, where i and j represent amino acid numbers), 71 were short range ($1 < |i - j| \leq 5$), and 104 were sequential ($|i - j| = 1$). These constraints apply to 87 residues out of a total of 153. Large regions entirely devoid of constraints stretch from position E20 (Lys77) to G3 (Lys102), and past H14 (Leu137). Approximate interproton distances were calibrated with Trp Cζ2H-CηH and Tyr CδH-CεH cross peak volumes. The observed NOEs were assigned to three categories corresponding to strong, medium, or weak intensities. The respective distance ranges for these three categories were 1.8 to 2.7 Å, 1.8 to 3.3 Å, and 1.8 to 5.0 Å. Methyl groups were treated as X-PLOR pseudo-atoms[49] with $<r^{-6}>$ averaging; 0.5 Å was added to the upper bound[50] and the allowable lower limit was unchanged. Likewise, no distinction was made between Cδ1H and Cδ2H (Cε1H and Cε2H) of Tyr or Phe rings, and $<r^{-6}>$ averaging was applied. Stereospecific assignments were made arbitrarily and then adjusted on the basis of resulting violations. At this low level of determination, no $^3J_{HN\text{-}C\alpha H}$ or H-bond constraints were considered.

All three-dimensional structures were calculated by using the program X-PLOR 3.1.[49] The calculations were performed with two protocols: one based on the hybrid DG-SA method,[51] and the other based on restrained (ab initio) simulated annealing starting from non-random template coordinates.[52] We used for templates an all-helical structure and the X-ray coordinates of metaquoMb. The DG-SA method led to structures with several violations, even after several rounds of simulated annealing refinement. The long unconstrained segment extending from residue 77 to residue 102 formed a loop stemming out of a distorted A-B-G-H nucleus (see Results). This DG protocol did not yield an acceptable set of structures with the limited number of available constraints, and the ab initio SA method was preferred. The results obtained from the all-helical conformation, when heating and cooling durations are multiplied by 4 in the first SA round,[49,52] are comparable to those from the X-ray structure. Unconstrained helices do unfold and the topology of the E-A-B-G-H segment is achieved. This demonstrates that there is no bias from starting with the X-ray coordinates and reducing the calculation time. We present the results obtained from the X-ray structure template.

All structures generated from the holoprotein coordinates with the ab initio SA protocol were subjected to an SA refinement (2,000 slow cooling steps (Δt = 0.003 ps, ΔT = 50 K) from the initial annealing temperature of 1,000 K to the final temperature of 100 K). This was followed by 300 steps of restrained minimization with effective van der Waals radii set to 0.75 times the standard values used in the CHARMm empirical energy function.[53] The majority of the resulting structures showed no NOE violations greater than 0.2 Å. Out of the 100 structures, twenty with lowest energy and fewest violations were accepted to prepare a low-resolution model of the apoprotein. Further refinement did not improve the quality of the structure, as judged by the relative total energy values.

RESULTS

His24Val–His119Phe Apomyoglobin

Mutant proteins, if stable and well behaved from the NMR standpoint, are generally helpful for confirming some assignments in the wild-type spectrum. We have inspected data from several histidine mutants of sperm whale myoglobin and among these, the His24Val–His119Phe double mutant yields spectra that are slightly but significantly better resolved than those of the wild-type protein. This resolution characteristic, the improved stability,[24] and the acid denaturation behavior[24] make the double mutant an interesting protein to complement our wild-type proton data analysis.

HyTEMPO accessibility

To probe the structure around the mutated sites, we have used HyTEMPO. This spin label broadens the lines of the residues it contacts on the accessible surface of the protein and therefore allows for the discrimination of comparatively sharper signals from residues buried in hydrophobic cores.[54] NMR spectra of the wild-type apoprotein show selective and pronounced broadening of resonances from the distal side of the heme binding site at low radical/protein concentration ratios.[17] In parallel, the EPR signals of free nitroxide decrease in amplitude as the protein is added to a radical solution. The broad asymmetrical signals of "bound" HyTEMPO could not be detected with the latter technique. It is assumed that the same interaction between HyTEMPO and distal residues takes place with the double mutant, and no effort was made to evaluate the residence time of the label on the protein.

Among the 11 aromatic residues of the wild-type apoprotein, only 4 are detected in the presence of a 10:1 molar ratio of radical to protein. Phe123 (GH5) is the most buried, and there is partial broadening of Trp7 (A5), Trp14 (A12), and Phe106 (G7).[17] Figure 2 illustrates the effect of HyTEMPO on the DQF-COSY spectrum of the double mutant. The pattern of protection is qualitatively the same as for the wild-type spectra, except for one additional protected phenylalanine spin system. Likewise, the aliphatic region with and without HyTEMPO (not shown) compares well with the wild-type region,[17] indicating at least partial protection of Leu2 (NA2), Val10 (A8), Val13 (A11), Val17 (A15), Val21 (B2), Ala22 (B3), Leu76 (E19), Leu115 (G16), Ala127 (H4), and Met131 (H8). Some of these residues are shown in Figure 1C. In the double mutant, an additional valine spin system is protected. NOESY data collected in the presence of the radical were particularly useful to reduce spectral overlap. In these relaxation-edited data, it is apparent that the additional protected phenylalanine and valine side chains are close to one another, and therefore the assignment to Val24 (B5) and Phe119 (GH1). The interpretation is supported by other NOE connectivities detected in the absence of HyTEMPO (see below). The strongest Val24–Phe119 dipolar contacts are between one methyl group of the former and the ε and ζ protons of the latter. The HyTEMPO experiments verify the integrity of hydrophobic cores in the double mutant and demonstrate that the side chains replacing the imidazole groups are buried near each other. They also confirm that the double mutant apoprotein resembles the wild-type apoprotein sufficiently for comparative spectral analysis.

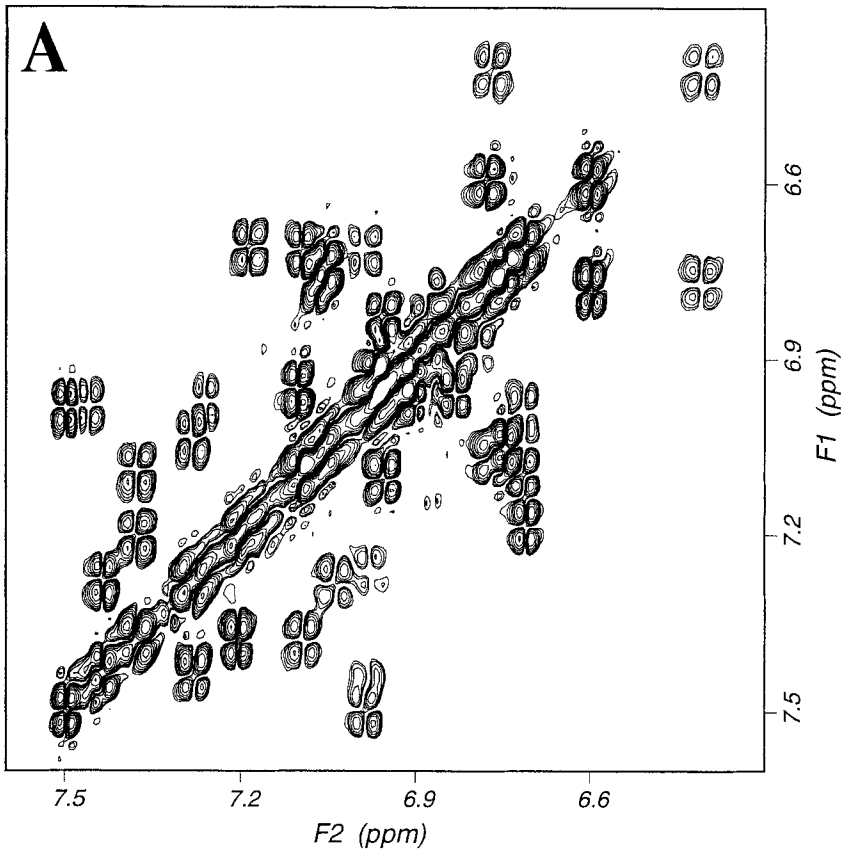
Chemical shift analysis

Spectra collected in the absence of the nitroxide radical can be analyzed like the wild-type spectra, and many assignments are readily transferred by comparison. Figure 3 presents TOCSY and NOESY data collected on the double mutant apoprotein in 90% H_2O/10% ^2H_2O. The same region is shown in Figure 4 for the wild-type apoprotein. Selected chemical shifts for the double mutant are listed in Table I and the deviations from wild-type values are illustrated in Figure 5, as a plot of chemical shift difference versus sequence number. Distinction is made between backbone amide protons and other protons.

On the GH side of the mutations, the available chemical shift perturbations are small. Phe119 interacts with core residues other than 24, namely Val17 (A15) and Leu115 (G16). The latter is moderately affected by the mutations at the β position. Several NOEs between the ring of Phe119 and residues 17, 24, and 115 are of medium intensity, stronger than expected if Phe119 is made to occupy as closely as possible the position of His119 in the holoprotein. In both wild-type and mutant proteins the α proton of 119 is shifted downfield near 5.3 ppm (Figs. 3A, 4A), and the $^3J_{HN\text{-}C\alpha H}$ value is larger than 8 Hz. Thus, overall the backbone conformation and environment of residue 119 are maintained upon mutation, and the NOE data hint at local repacking.

In the A helix, chemical shift deviations are noted for residues normally in contact with residue 24, in particular for Val17 (A15), a residue packing near Trp 14 (A12)[16] (Fig. 5). The ring current generated by aromatic residues at position 17 of the wild-type apoprotein is large; therefore even a minor rearrangement caused by mutation at 24 could lead to shifts. The introduction of Phe119 might also contribute some ring current at that location, if Phe119 relocates closer. In the B helix, Asp20 (B1) and Asp27 (B8) are affected. This can also be rationalized with contact disruptions not necessarily caused by extreme structural reorganization. The wild-type solid state structure[44] shows that the backbone carbonyl of Asp20 is within hydrogen bonding distance of the Nδ1 atom of His24, and that both Asp20 and Asp27 are within close range of Arg118 (G19). The side chain of Arg118 stretches on the surface of the holoprotein (Fig. 1C), and its conserved location in solution is manifested in NOEs from its δ protons to the ring of His24 and detectable guanidinium group signals (NεH, 6.5 ppm; NηH, 6.2 ppm). In the wild-type apoprotein, the 24–118 NOEs and guanidinium group signals are also observed. The mutations at 24 and 119 and a consequent displacement of Arg118 could account for the effects on Asp20 and Asp27. Direct perturbation due to the ring current generated by Phe119, although small if the structure at 119 is maintained, could increase upon repacking of the interface.

Several amide hydrogens are affected by the mutation: the stretch 17 to 27 experiences large variations upfield and downfield with a 1.5 ppm differ-


A
7.5
7.2
6.9
6.6
F2 (ppm)
6.6
6.9
7.2
7.5
F1 (ppm)

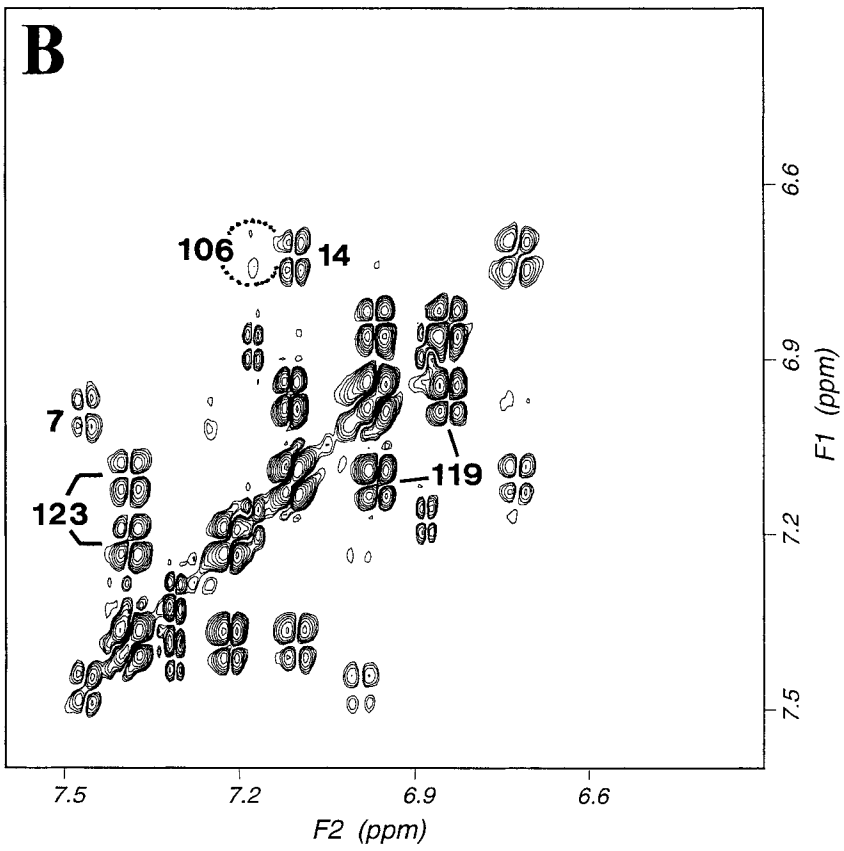

B
106
14
7
123
119
7.5
7.2
6.9
6.6
F2 (ppm)
6.6
6.9
7.2
7.5
F1 (ppm)

Fig. 2.

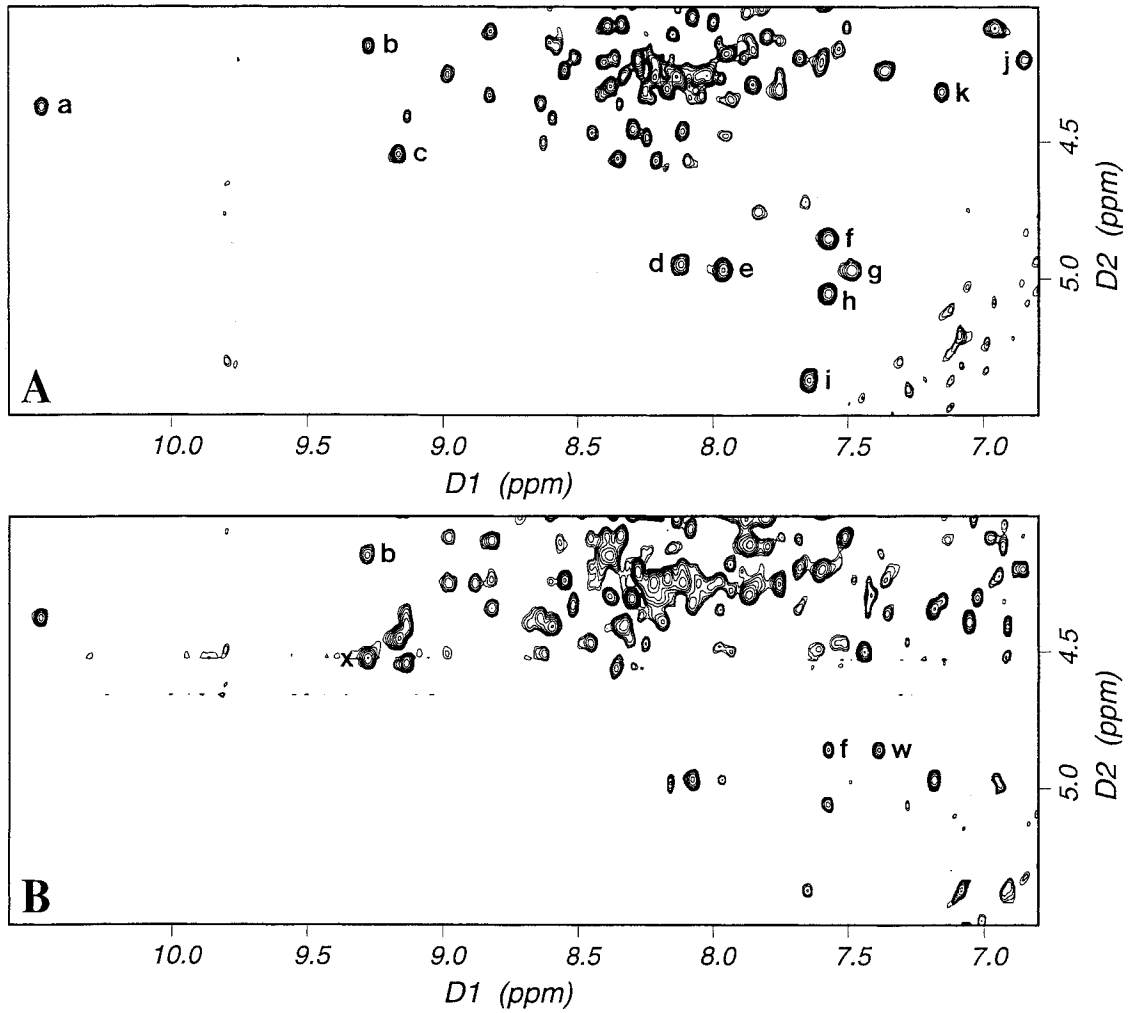


Fig. 3. Section of TOCSY (**A**) and NOESY (**B**) spectra of His24Val–His119Phe sperm whale apoMb in 90% H_2O/10% ^2H_2O at pH 5.7 and 304 K. The TOCSY data were recorded with 53 ms mixing time. The intraresidue NH-CαH cross peaks are: a, Glu38; b, Glu59; c, Ser3; d, His36; e, Phe123; f, Asp20; g, Phe46; h, Asp122; i, Phe119; j, Ala19; k, Ala57. The NOESY data were recorded with 100 ms mixing time. Interresidue connectivities are marked with letter: w, Val21 NH to Asp20 CαH, indicative of the AB kink; x, Glu59 NH to Ser 58 CαH, indicative of the D-E kink.

ence for Val21 (B2) (Figs. 3B, 4B). In the holoprotein solid state structure,[28] the NH of 21 does not participate in hydrogen bonding, either with a backbone carbonyl, or with a capping side chain. It is within 2.4 Å of the α proton of Asp20. A strong NOE is observed between these two protons in the wild-type holoprotein (not shown), wild-type apoprotein (Fig. 4B, marked w), and double mutant apoprotein (Fig. 3b, marked w). In all the proteins studied, the α protons of Asp20 and Val21 are found at chemical shifts consistent with extended strand and helix, respectively,[55] and indicate the normal formation of the AB hinge. The amide-to-amide NOEs characteristic of the A and B helices are maintained in the mutant, except perhaps at position 16, where shift caused by Phe119 and spectral overlap with aromatic signals prevent unambiguous detection. One interpretation for the shift at position 21 is that the introduction at position 24 of a residue with a relatively low helical propensity and no ability to form a side-chain H-bond causes a reorientation of the preceding Asp20 side chain.

Assigned NOEs and chemical shifts in other regions of the double mutant apoprotein are similar to those in the wild-type form. The general appearance

Fig. 2. The aromatic region of DQF-COSY spectra of sperm whale His24Val–His119Phe apoMb in ^2H_2O at pH* 5.7 and 304 K. **A:** Reference spectrum, collected in the absence of HyTEMPO. The aromatic groups consist of 3 Tyr, 2 Trp, and 7 Phe. **B:** Spectrum of the same sample to which HyTEMPO was added to a concentration of 15 mM (molar ratio of about 10). At this concentration of radical only signals from Trp7, Trp14, Phe106, and Phe123 remain detectable in the wild-type protein.[17] In the double mutant protein, the same residues are protected. The signal from Phe106 (circled) is weak but detectable. An additional phenylalanine spin system is seen in its entirety. Analysis of the NOE connectivities confirms that this residue is the mutated Phe119. The aliphatic region of the spectrum (not shown) allows for the assignment of Val24, which is also protected from the broadening.

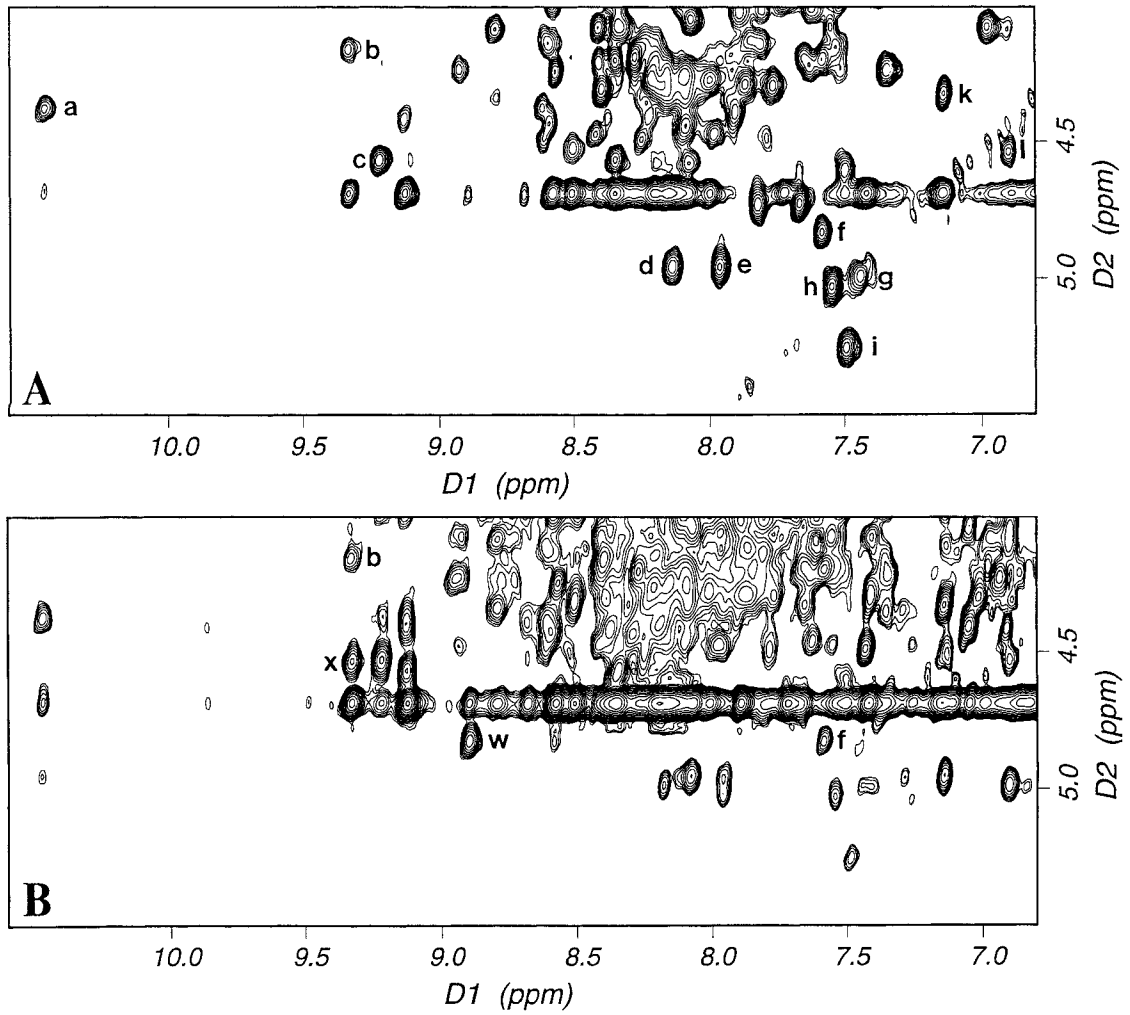


Fig. 4. Section of TOCSY (4A) and NOESY (4B) spectra of wild-type sperm whale apoMb in 90% H_2O/10% ^2H_2O at pH 5.7 and 304 K. Clean TOCSY data were recorded with 43 ms mixing time; the NOESY data, with 110 ms mixing time. Both wild-type experiments used pulsed-field gradients (z-axis) for water suppression; the data were processed with a softer sine-bell window than in Figure 3. Labeling in A is as in Figure 3A with, in addition: l, Ser58; j is outside the boundaries of the plotted region. Labeling in B is as in Figure 3B.

of the spectra suggests that local and probably minor rearrangements occur in the A-B-G-H interface.

Backbone amide hydrogen exchange

The double mutant shows altered backbone hydrogen exchange properties compared with the wild-type apoprotein. Figure 6 displays the amide region of a NOESY spectrum collected on the double mutant dissolved in ^2H_2O at pH* 5.7 and 304 K. The data set required approximately a day of acquisition time. A few NH-NH connectivities tracing a turn of a helix are clearly detectable and are marked in the figure. Data sets collected in H_2O display unique amide cross peaks arising from Val24 and Phe119, and the former allows for the assignment of the labile protons whose exchange is relatively slow; they are located in the B helix, at positions 27 (B8) to 33 (B14). Other slowly exchangeable protons are assigned to Trp14 (A12), Val17 (A15), Gly73 (E16), Val114 (G15), and Ala134 (H11). Wild-type data collected in the same fashion display the same amides, although their intensity is much lower in the B helix as the exchange occurs more rapidly.

The incompletely exchanged spectrum in Figure 6 simplifies an otherwise crowded region, and scalar connectivities can be verified for the B helix. This aids in the interpretation of the wild-type data. For example, the persistence of Arg31 NH in the ^2H_2O spectra leads to the characterization of this residue. From the amide NH, a series of NOEs are observed that match NOEs observed from His113 δ2H. The same NOEs are observed in holomyoglobin. The assignment of the 113 δ2H NOEs to Arg31 in the double mutant and in the wild-type apoprotein is supported by data on mutants at position 113 (His113Glu and His113Gln, not shown): in these

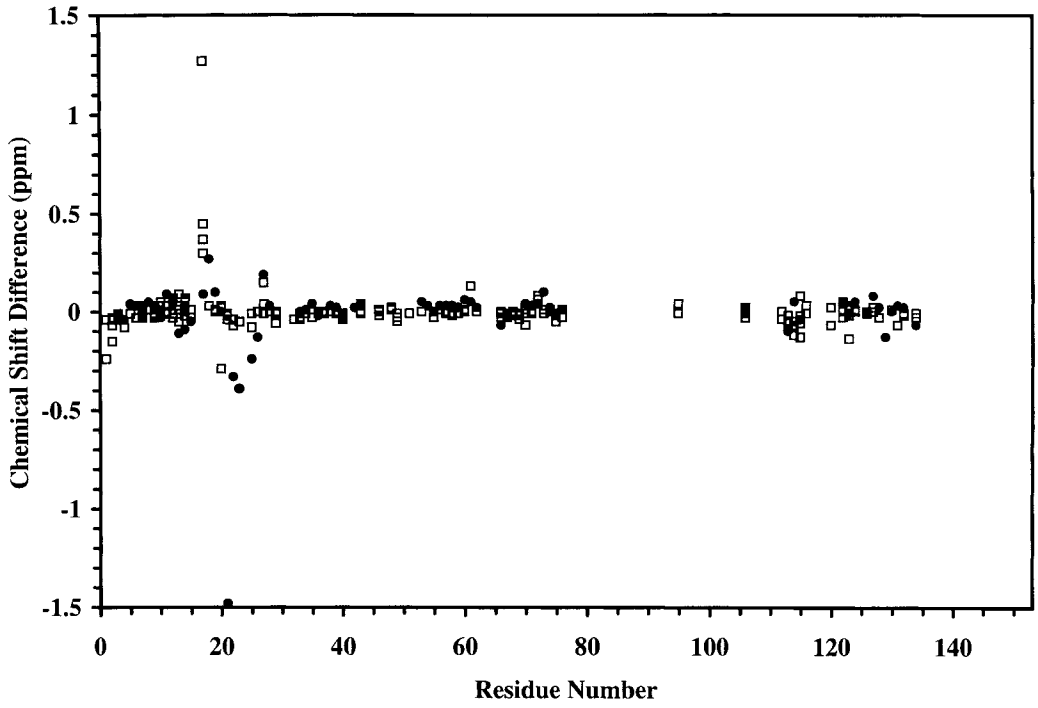


Fig. 5. Chemical shift difference plot for apoMb in 90% H_2O/10% ^2H_2O at pH 5.7 and 304 K. The chemical shifts of the wild-type apoprotein were subtracted from those of the His24Val/His119Phe mutant where available, and plotted versus residue number. Filled circles represent backbone amide data; open squares, other signals. The chemical shift difference is small overall, except around the mutated B5 (Val24) residue. The deviation observed for the first three residues arises from the terminal Met, which is not cleaved in the mutant.

mutants, the δ-δ′ cross peak of Arg31 is shifted to a different position. The proximity of Arg31 (B12) and His113 (G14) is expected on the basis of the X-ray structure of the wild-type holoprotein and is assumed to play a role in the depressed pK_a of His113 in both holo[30,56] and apoprotein.[30]

Wild-Type Apomyoglobin

D helix assignment

For wild-type apoMb, data collected with pulsed-field-gradient water suppression (Fig. 4) helped to describe a hitherto unexplored region of the structure. A shifted amide proton at 9.38 ppm, which exchanges efficiently with the solvent, has two strong NOEs to α-protons: its own, at 4.15 ppm (Fig. 4B, cross peak b), and that of another residue, at 4.53 ppm (Fig. 4B, cross peak x). The latter is *J*-coupled to an amide proton at 6.90 ppm. From 6.90 ppm, an NOE to the NH of an Ala is observed (7.15 ppm), and NH-NH connectivities follow, to 8.49 ppm, 8.14 ppm, 7.73 ppm, and 8.36 ppm, which arises from another Ala spin system. The sequence A-X-X-X-A-X-X is uniquely matched to A53-E54-M55-K56-A57-S58-E59. Although not all correlation signals could be unambiguously attributed past the β protons, the assignments are consistent with this analysis. Beyond Glu59, an NH-NH NOE is observed to Asp60. This new stretch corresponds to residues in the D and E helices (D3 to E3) and precedes the previously reported stretch in the E helix (residues E10 to E19). The NH-NH connectivity is not detected between Ser58 and Glu59, where the chain takes a bend between the D and E helices in the holoprotein X-ray structure.[44]

Chemical shift analysis

The chemical shift of the proton resonances arising from residues in structured regions provides a criterion to judge the similarity of wild-type holo and apo structures.[17] Since the D and E helices are located close to the heme group, the contribution from the heme ring current must be subtracted from the holoprotein shift value for the comparison to be valid. This correction is necessary when the shifts in the diamagnetic carbonmonoxy protein are compared. When comparing with paramagnetic metaquoMb, both the heme ring current shift and the dipolar shift must be accounted for.[57] The ring current generated by the heme group was calculated with the X-ray coordinates of the holoproteins.[57] In the D3-E3 segment of the protein the heme ring current effects are negative and small (<0.15 ppm), so that the deviation from statistical coil in the holoprotein is largely due to the polypeptide fold. After correction for heme effects, the correspondence of chemical shift values between holoprotein, either in

TABLE I. Selected Chemical Shifts in Wild-Type and His24Val–His119Phe Sperm Whale apoMb*

Residue			NH	αH	βH	Others
Leu2	NA2	dm[a]		4.45	0.60	γH 1.10; δH_3 0.21,−0.31
Ser3	A1	dm	9.19	4.53		
Glu4	A2	dm	9.17	4.41	2.08, 2.17	
Gly5	A3	dm	8.71	3.83		
Glu6	A4	dm	7.81	3.83		
Trp7	A5	dm	8.65	4.50	3.39, 3.13	δ1H 6.91; NεH 9.90; ζ2H 7.24; η2H 6.96; ζ3H 6.96; ε3H 7.44
Gln8	A6	dm	8.36	4.08		
Leu9	A7	dm	7.61	4.18		γH 1.92; δH_3 0.72, 0.84
Val10	A8	dm	7.93	3.76	2.46	γH_3 0.90, 1.01
Leu11	A9	dm	9.02	4.25		γH 2.01; δH_3 0.79, 0.87
His12	A10	dm	8.48	4.46	3.47	δ2H 7.29; ε1H 8.44
Val13	A11	dm	7.53	3.70	2.41	γH_3 0.96, 1.13
		wt	7.64	3.69	2.38	γH_3 1.03, 1.04[†]
Trp14	A12	dm	8.86	3.91	3.16, 3.41	δ1H 6.97; NεH 10.19; ζ2H 7.48; η2H 6.92; ζ3H 6.69; ε3H 7.07
Ala15	A13	dm	7.50	4.08	1.38	
Val17	A15	dm	6.94	3.00	1.23	γH_3−0.07, 0.47
Glu18	A16	dm	7.69	3.45		
Ala19	AB1	dm	6.86	4.19	1.39	
Asp20	B1	dm	7.58	4.85	2.60, 2.47	
		wt	7.58	4.83	2.89, 2.44	
Val21	B2	dm	7.41	3.47	2.03	γH_3 0.94, 0.97
		wt	8.89	3.48	2.05	γH_3 0.98, 0.98
Ala22	B3	dm	8.25	3.98	1.25	
Gly23	B4	dm	7.73	3.92		
Val24	B5	dm	8.16	3.34	1.82	γH_3 0.76, 0.97
Gly25	B6	dm	8.60	3.39, 3.57		
Gln26	B7	dm	7.93	3.75		
Asp27	B8	dm	7.78	4.24	2.51, 2.82	
		wt	7.59	4.20	2.67, 2.52	
Ile28	B9	dm	8.43			
Leu29	B10	dm	7.42			γH 1.32; δH_3 0.38, 0.78
Ile30	B11	dm	8.28			
Arg31	B12	dm	7.76	3.97		γH 1.50, 1.62; δH 2.99, 2.68
		wt	7.58	3.98		γH 1.52, 1.66; δH 3.01, 2.65
Leu32	B13	dm	8.16	4.01		
Phe33	B14	dm	8.21	4.40		δH 7.05; εH 6.76; ζH 6.40
Lys34	B15	dm	8.42	4.07		
Ser35	B16	dm	8.11	4.02	3.63, 3.34	
His36	C1	dm	8.13	4.97	2.75, 2.91	δ2H 7.04; ε1H 8.06
Glu38	C3	dm	10.50	4.38		
Thr39	C4	dm	8.37	4.20	4.22	$\gamma 2H_3$ 1.34
Leu40	C5	dm	7.34	3.82	1.60, 1.34	γH 1.25; δH_3 0.28, 0.38
Phe43	CD1	dm	7.86		2.95, 3.02	δH 7.41; εH 7.26; ζH 7.01
Phe46	CD4	dm	7.44	5.00		δH 6.93; εH 6.85; ζH 6.85
Ala53	D3	dm	8.41	3.98	1.36	
		wt	8.36	3.98	1.36	
Glu54	D4	dm	7.76			
		wt	7.73	3.97		
Met55	D5	dm	8.14	3.95		εH_3 1.83
		wt	8.14	3.95		εH_3 1.86
Lys56	D6	dm	8.52	3.77		
		wt	8.49	3.76		
Ala57	D7	dm	7.17	4.31	1.43	
		wt	7.14	4.32	1.43	
Ser58	E1	dm	6.93	4.52	4.11	
		wt	6.90	4.54	4.12	
Glu59	E2	dm	9.32	4.14		
		wt	9.38	4.15	2.11, 2.38	
Asp60	E3	dm	8.41	4.57	2.72	
		wt	8.35	4.56	2.72	

(continued)

TABLE I. Selected Chemical Shifts in Wild-Type and His24Val–His119Phe Sperm Whale apoMb* (Continued)

Residue			NH	αH	βH	Others
Val66	E9	dm	7.95	3.52	2.29	γH_3 0.90, 1.07
Thr67	E10	dm	8.36	3.89	4.39	$\gamma 2H_3$ 1.25
Thr70	E13	dm	8.83	4.08	4.27	γH_3 1.32
Ala71	E14	dm	7.65	4.19	1.55	
Leu72	E15	dm	7.90			γH 1.05; δH_3 0.42, 0.61
Gly73	E16	dm	8.84	3.89, 2.59		
Ala74	E17	dm	7.54	3.98	1.43	
Leu76	E19	dm	7.97			γH 1.41; δH_3 −0.17, 0.61
His113	G14	dm		4.23	3.15, 3.23	δ2H 6.80; ε1H 7.88
Val114	G15	dm	8.55	3.59	1.98	γH_3 0.93, 1.03
Leu115	G16	dm	8.53	3.71	0.22, 1.15	γH 1.72; δH_3 0.52, 0.61
Arg118	G19	wt	8.22			δH 2.29, 2.73; εH 6.72; ηH 6.40
Phe119	GH1	dm	7.66	5.38	2.88, 3.40	δH 7.07; εH 6.93; ζH 6.81
Asp122	GH4	dm	7.59	5.08	2.35, 2.81	
		wt	7.55	5.03	2.38, 2.81	
Phe123	GH5	dm	7.95	4.98	2.65, 3.28	δH 7.19; εH 7.37; ζH 7.07
Asp126	H3	dm	8.57	4.23	2.50, 2.61	
Ala127	H4	dm	8.32	3.90	1.50	
Gln128	H5	dm	8.44	3.35		
Met131	H8	dm	8.65	4.31		εH_3 2.15
Ala134	H11	dm	8.22	3.74	1.31	

*Chemical shifts are listed at 304 K and pH 5.7 with respect to the water line at 4.70 ppm. The abbreviations "wt" and "dm" refer to wild-type and His24Val–His119Phe apoproteins, respectively. Other wild-type assignments can be found in Refs. 16, 17, and 30.
†Mislisted in ref. 17.

the carbonmonoxy form or the metaquo form, and apoprotein is remarkably good (sum of squared residuals <0.1 ppm^2 with the apoprotein, compared with >4 ppm^2 with the statistical coil). The detection of organized structure in the D helix and the beginning of the E helix through NOEs and chemical shifts is noteworthy because this section of the myoglobin backbone contains no slowly exchanging labile amide hydrogens that can serve as probes of the secondary structure.[5,18]

NOE analysis

The NMR data on wild-type apoMb restrict the number of possible conformations in solution. The structural implications can be visualized with a three-dimensional model accommodating all the assigned NOEs. To construct such a model, the methods of NMR structural determination were applied. The volumes of assigned cross peaks were measured and scaled to the volume of resolved Trp and Tyr ring cross peaks. The contacts were then organized in three distance bins with allowance for equivalent protons from freely rotating groups. The interresidue distances were compiled in a graphical form to prepare the contact map in Figure 7A, which illustrates that at this stage of analysis the geometrical constraints are unevenly distributed in the apoprotein structure. For example, the A helix displays many contacts within itself and with residues from the E helix, and the H helix. The B helix is docked onto the G helix. On the other hand, the F helix and the carboxy end of H helix remain without confirmed interresidue connectivities. For comparison with the NMR constraints, Figure 7B shows a contact map of the holoprotein. This map was obtained by using all non-hydrogen atoms in the X-ray structure and a maximum distance of 5 Å. The symmetry of the plot illustrates that the identified interfaces of the apoprotein resemble those in the holoprotein. The interactions in the A-B-G-H interface are in relatively large number, and building a model of this region was attempted.

Structural calculations were performed over the whole sequence (153 residues) with the available apoprotein constraints. Because of the small number of restraints (an average of only 3 per residue over the whole structure, and 5 over the 87 residues used in Table I), the hybrid DG-SA calculations resulted in an unrealistic model of apoMb. The best structures exhibited a high number of large constraint violations (even though low-energy, small-violation structures are obtainable with the same constraint list) and did not account for the experimental data on secondary structure content,[17] histidine pK_a s,[30] protection from HyTEMPO,[17] and radius of gyration.[58] In spite of these limitations, the A-B-G-H interface did emerge from the DG-SA structures, and the A helix was recognizable. For all its shortcomings, this model, which will not be discussed further, announces the trend to expect when the calculations can be improved with additional constraints and dynamics information.

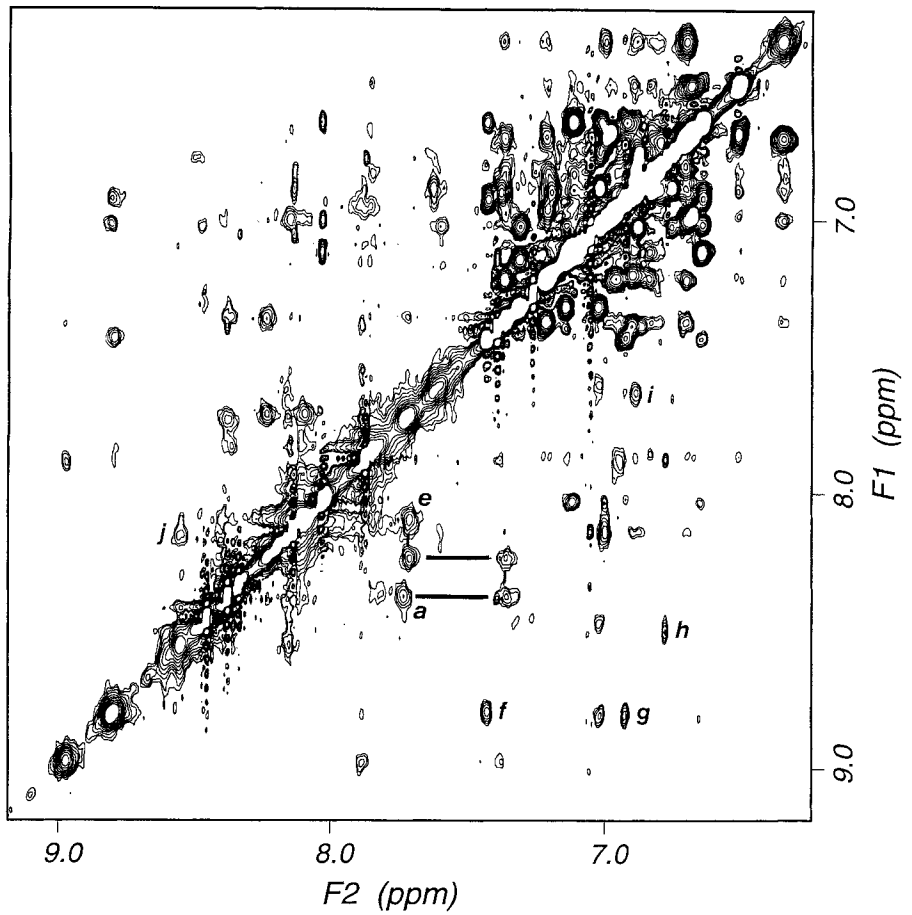


Fig. 6. NOESY spectrum of sperm whale His24Val–His119Phe apoMb in ^2H_2O at pH* 5.7 and 304 K. The mixing time was 100 ms. This spectrum was collected immediately after dissolution in deuterated water and a few amide protons remain detectable throughout the experiment. Long-lasting connectivities are detected in the B helix from cross peak a (Asp27NH to Ile28NH) to cross peak e (Arg31NH to Leu32NH). Other slowly exchanging amides, also observed in the wild-type apoprotein, are labeled (F1 dimension to F2 dimension): f, Gly73 (E16) NH to Trp14 (A12) Cζ2H; g, Trp14 NH to Trp14 Cδ1H; h, Val114 (G15) to His113 (G14) Cδ2H; i, Glu18 (A16) NH to Val17 (A15) NH; j, Ala134 (H11) NH to Lys133 (H10) NH.

When starting with the X-ray structure (or an all-α conformation) and the identical set of constraints, the ab initio SA protocol produces structures that have low energies and a small number of violations. Unfolding is observed where the chain is unconstrained, as expected when efficient conformational sampling occurs. At this stage, the unfolded regions carry little information; they reflect either the lack of knowledge about the structure (incomplete assignments, insufficient resolution) or lack of ordered, stable structure. It is likely that tighter convergence will be reached in some of these regions as the spectral resolution is improved with the use of isotopic labels, and more constraints can be added to the list.

Of greater interest are the segments that do converge to a folded structure. These are presented in Figures 8 and 9. In Figure 8, the 20 structures of lowest energy among a set of 100 structures are included, superimposed by their α-carbons with a least squares procedure onto the α-carbons of the A-B-G-H interface in the holoprotein (arbitrarily defined as residues 9 to 27 and 113 to 122). The rms deviation of the α-carbons averaged over the 20 structures is 2.0 Å compared with the holoprotein structure. The experimental data are consistent with a compact subdomain[21] at the A-B-G-H interface, and the calculations produce a well-formed interface. His24 and His119 are the sole side chains included in Figure 8. They assume a similar orientation in all structures. The A helix is folded to the amino end, although with a higher rms deviation than at the carboxy end. The chemical shift values at the N-terminus correlate well with the holoprotein values, and a similar extent of folding is therefore expected. The kink between the A and the B helix at residue 19 (AB1) is observed and the contacts between the GH turn and the A-B segments organize the chain in the proper orientation. The B helix is folded at its N-terminus, up to residue 30. Past this residue, the

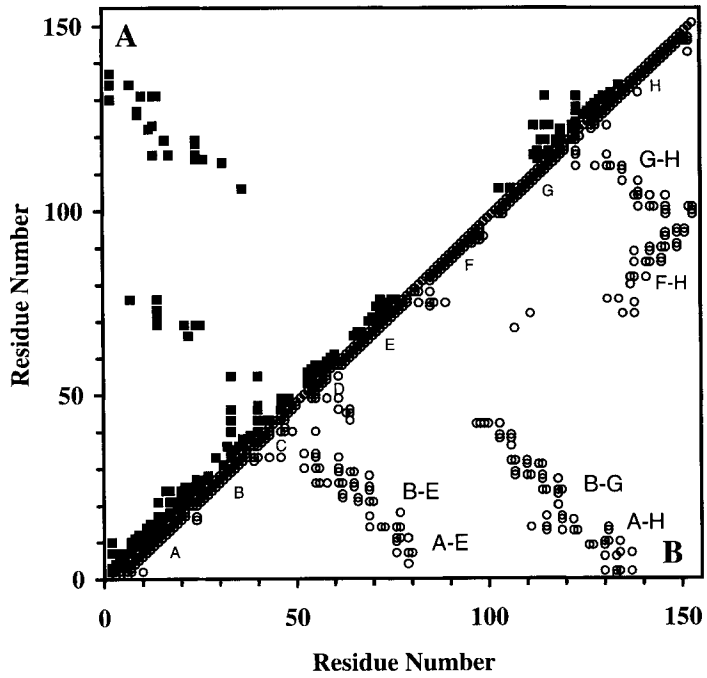


Fig. 7. Dual constraint and contact plot for apoMb and holomyoglobin. The interresidue proton-proton dipolar contacts observed in native apoMb are represented above the diagonal (**A**, filled squares), whereas a contact map calculated from the X-ray structure of metaquoMb[44,45] is offered for comparison below the diagonal (**B**, open circles). Holoprotein helix nomenclature is listed in Table II and Figure 1. In the apoprotein, many long-range contacts are observed that mirror those in the holoprotein.

constraints are few and fraying is seen. Although the manifold solution to the constraint problem might be artificial, the entrance into the B helix is not. Several side chains give rise to strong dipolar contacts that imply order and limited fluctuations. Likewise, the helical turns immediately preceding and following the GH corner are detected and structural ambiguity is observed past them.

The additional element of structure included in Figure 8B is a section of the E helix. Even though the α-carbons of this section were left out of the least squares alignment, the positioning with respect to the A-B-G-H subdomain is correct. This is due to A-E and B-E contacts, in particular those involving Trp14 and Ala22. The compact subdomain, from this structural point of view, could be extended to include this ten-residue E segment.

Figure 9 shows three small segments, each superimposed independently on the holoprotein structure. In Figure 9A, the 20 structures were aligned with the α-carbon of residues 9 to 18. These last two turns of the A-helix are well defined, with short-range interactions, and the rms deviation for 9 to 18 alone is only 0.7 Å. Figure 9B displays the D helix, with alignment over the α-carbons of 53 to 60 (rms deviation of 2.0 Å), and Figure 9C displays the end of the E helix with alignment over the α-carbons of 67 to 77 (rms deviation of 2.3 Å). In contrast, the F helix, beginning of the G-helix, and end of H-helix, for which no constraints are available, give large average rms deviations and large excursions from the average.

DISCUSSION

The Val B5—Phe GH1 pair is naturally found at the A-B-G-H interface of human hemoglobin β-chains, where the shortest carbon interatomic distance between the two residues is ~6 Å.[59] In sperm whale apoMb when the hydrogen-bonded histidines at B5 and GH1 are replaced by the hemoglobin pair, there are two types of effects apparent in the proton NMR spectra: 1) structural distortions; and 2) deceleration of hydrogen exchange. The extent of structural distortions appears minimal. Hydrophobic cores are conserved; side chains are oriented as in the wild-type protein; and elements of secondary structure are formed with similar boundaries. NOE and chemical shift parameters are consistent with a shorter B5-GH1 distance than in hemoglobin. Since the Val-Phe pair occupies a smaller volume than the His-His pair, crevice formation might occur upon mutation, as in bovine pancreatic trypsin inhibitor[60] and in T4 lysozyme.[61-63] The double replacement His24Val–His119Phe leads to a small stabilization of the native state ($\Delta\Delta G^{\circ}_{N-U} \cong 4$ kJ mol^{-1} at pH > 7 and 0°C).[24] Cavity formation is likely to be destabilizing, but the formation of a "hydrophobic bridge" between Val24 and Phe119 is expected to have the opposite effect at pH values below 7.[24,64] These two possibilities need to be explored at higher

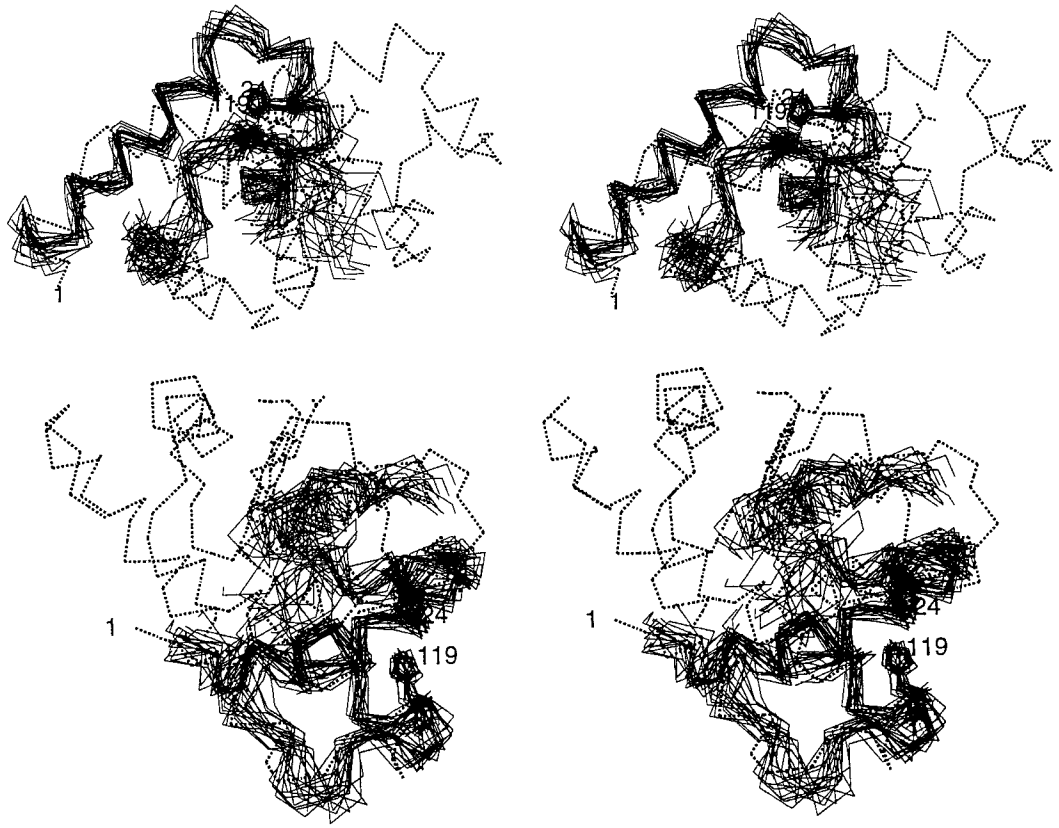


Fig. 8. **A,B:** Modeling of the A-B-G-H interface of wild-type apoMb. Only Cα traces and His24 and His119 are shown. A family of 20 structures (thin continuous lines) is superimposed over the holoprotein structure (thick dashed lines, with heme group included). The superimposition was achieved using the Cα of residues 9 to 18, 20 to 27, and 112 to 122. The apoMb Cα traces are shown from residue 3 to 32 and 113 to 130. In addition to these sections, a piece of the E helix (67 to 76) is included in B.

resolution than offered here, with consideration of intervening solvent molecules.

The second effect of the double mutation in apoMb, retardation of hydrogen exchange exhibited several residues away from the mutations, suggests that the dynamics of the polypeptide chain are affected by repacking around the mutated sites. Sharpening of the proton lines and hydrogen exchange deceleration are also observed when the wild-type apoprotein binds ANS.[17] It is possible that the mutations, like ANS binding, cause a shift of the equilibrium population toward a more restricted set of native-like conformations.

The implication of the double mutant results is that the hydrogen bond between B5 and GH1 in myoglobin is not essential for docking of the GH turn onto the A and B helices. The structural similarity in spite of the mutations allowed for an improved analysis of A-B-G-H signatures in the wild-type apoprotein. The parallel study yielded enough dipolar contacts to define the three-dimensional structure of the interface. The calculated model satisfies all the available geometry requirements. It is also consistent with the experimental consensus that either the structure is very similar to that found in the holoprotein, or more or less unfolded, and that there is no evidence for packing resulting from an unrelated topology. At this stage of assignment, there remain non-statistical coil resonances of unknown origin. These, and dynamic characteristics, will eventually be incorporated for a complete structural determination.

Can properties of apoMb be explained with the model? It is premature to put this incomplete representation of the protein to specific uses but some features are worth mentioning. For example, the model supports the interpretation of thermodynamic data[1] that there is a well-defined hydrophobic core sequestering side chains from the solvent. The NMR data also require that Phe123 be buried in the A-H interface. Its replacement is known to be deleterious to the A-B-G-H compact subdomain,[21] indicating that there is little tolerance at this position. However, the connection between structure and thermodynamic properties cannot be formulated if the side chain is not significantly buried. As reported in the context of holomyoglobin,[65] surface replacements do not follow simple rules. In the case of His113, interactions are detected with Arg31, an exposed side chain from the B helix, yet various residues can be

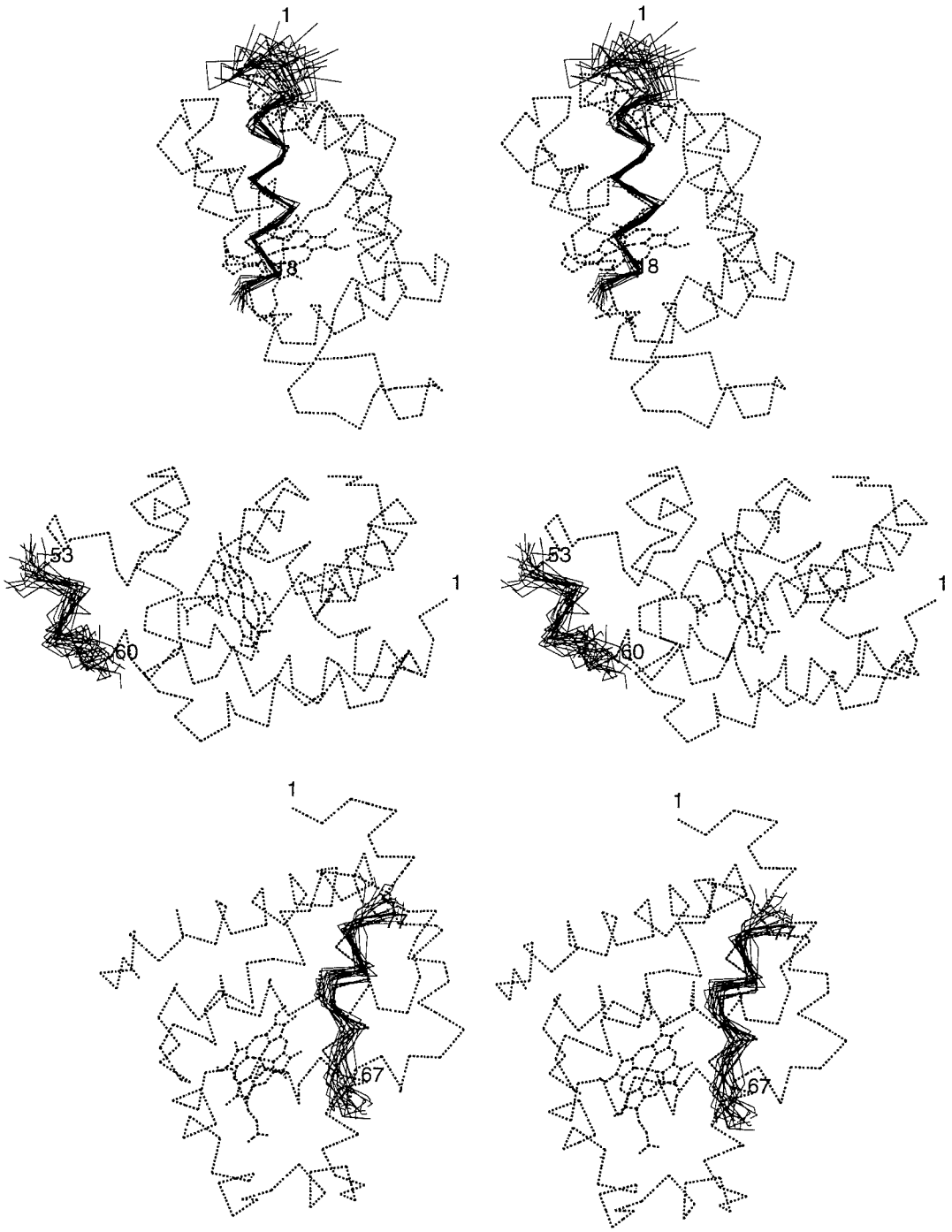


Fig. 9. Three segments of structure in wild-type apoMb. Only Cα traces are shown. In each case, a family of 20 structures (thin continuous lines) is superimposed over the holoprotein structure (thick dashed lines, with heme group included). The superimposition was achieved using the Cα of residues 9 to 18 (part of A helix, shown in **A**, 3–18); 53 to 60 (mostly D helix, **B**); and 67 to 77 (part of the E helix, **C**). Three different orientations are used.

accommodated at position 113 without severe consequences.[24]

In addition to a stability role, the A-B-G-H interface elements might participate in heme recognition. The time course of myoglobin folding in the presence of hemindicyanide has been followed by CD, absorption, and fluorescence rapid mixing experiments.[66] When myoglobin is refolded from the urea-unfolded mixture of apoprotein and heme at pH 9.1, four distinct events appear to take place se-

quentially, the fastest of which can be interpreted as the formation of an intermediate containing less than 40% α-helix. It can be speculated that this intermediate corresponds to the species observed in the refolding experiments without hemindicyanide[18] and therefore that it is folded in parts of the A, G, H, and B sections of the protein. This intermediate is capable of capturing the heme group, although where the hemindicyanide binds is not known. It is interesting that a truncated form of apoMb spanning residues 32 (B13) to 139 (H16) (therefore lacking the A helix, the first residues of the B helix, and the last residues of the H helix) adopts in solution a structure whose ellipticity at 222 nm indicates 39% helix.[67] This "minimyoglobin" also binds heme. Thus, the A helix and the structure it induces elsewhere are not necessary in the first stage of recognition. The G and leftover H segments, which make contact with the heme group in the holoprotein, and the structure that they organize would assume a dominant role in this species.

The kinetics of folding of apoMb have been studied by pulse labeling methods in the absence of heme.[18] According to these stopped-flow experiments, the folding of apoMb from the urea-unfolded state is a sequential process. Amide hydrogen protection, indicative of secondary structure, is achieved rapidly (within a few milliseconds) for sections of the A, B, G, and H helices. The C-terminal folding of the A and B helices of apoMb is completed within 1 second, whereas protection is observed for C, CD, and E amides at later times, within a few seconds.[18] Little is known about the folding of the D and F helices because these segments of structure contain no slowly exchangeable amide hydrogens and therefore are not observed in the pulse-labeling experiments.[5,18] The equilibrium NMR data demonstrate that the D helix is also structured in the native apoprotein, where interactions are detected with the CD corner, the C helix (slow folding), and a turn of B helix (fast folding).[18]

The equilibrium NMR results, when combined with the published pulse-labeling data, limit which parts of the protein undergo complete secondary structure unfolding. A list of folded regions is provided in Table II. If the residues adopting helical conformation are added up, a lower bound limit is calculated around 60% at pH 5.7, higher than the 55% helicity derived from CD data. The validity of the CD interpretation in the case of apoMb has been questioned,[5] and a recent theoretical study carried out to evaluate the effect of fluctuations and breaks in secondary structure proposed that these effects could result in a reduction of the ellipticity at 222 nm.[68]

Although it is not possible to define precisely the unfolded regions, experimental data suggest their location. Enzymatic digestion profiles show that the C-terminus of the H helix is more susceptible to proteolysis in the apoprotein than in the holoprotein.[15] Chemical shift estimates based on the holoprotein values[57,69] for the carboxy end of the H helix predict shifted resonances where none are observed. This negative evidence, when balanced against the excellent correlation obtained in the structured regions described above and combined with the lack of exchange protection past Ala143 (H20),[5,18] points to altered conformational distribution at the carboxy end of helix H. Likewise, the EF turn and F helix give calculated resonance positions within spectral windows devoid of signals, and mutagenesis data indicate dissimilarity with the holoprotein.[24] Since the end of the H helix is in contact with the F helix in the holoprotein structure (Fig. 7B), the concomitant disruption of these two regions is plausible.

TABLE II. Secondary Structure in Wild-Type apoMb*

Location	Method
A3-A16	Pulse labeling†
A2-A5, A8-A16	Direct observation‡
B1-B7, B10	Direct observation
B6-B15	Pulse labeling
C1-C5	Pulse labeling
C3-C5	Direct observation
D3-D7	Direct observation
E1-E3, E10, E12-E17, E19	Direct observation
E3-E20	Pulse labeling
G1-G16	Pulse labeling
G15-G16	Direct observation
H4-H8, H11, H15	Direct observation
H6-H20	Pulse labeling

*Elements of secondary structure in the holoprotein are as follows: A1–A16: Ser3–Glu18; B1–B16: Asp20–Ser35; C1–C7: His36–Lys42; CD1–CD8: Phe43–Leu49; D1–D7: Thr51–Ala57; E1–E20: Ser58–Lys77; EF1–EF8: Lys78–Glu85; F1–F9: Leu86–Ala94; FG1–FG5: Thr95–Ile99; G1–G19: Pro100–Arg118; GH1–GH5: His119–Phe123; H1–H26: Gly124–Leu149.

†As reported in Ref. 18.

‡This work and as reported in Refs. 16, 17, and 30.

A prediction of the apoMb structure was recently published by Srinavasan and Rose.[70] These authors applied a hierarchic condensation algorithm to determine the structure encoded by the amino acid sequence. The secondary structure proposed by the algorithm resembles that of the holoprotein in all regions, including the F and H helices. Other standard algorithms also arrive at similar predictions. This indicates that the amino acid sequence has the specificity necessary to form the helices of the hologlobin fold, independently of heme influences. The heme would provide stability to the encoded structure without dictating it.

Although the three-dimensional structure of many *b* hemoproteins has been determined by X-ray diffraction methods, in only one instance, that of apocytochrome b_{562}, has the structural role of the

prosthetic group been characterized.[71] Cytochrome b_{562} is a 4-α helical bundle with the heme pinched in at one end of the structure. Removal of the heme leads to partial distortion of four of the four helices. In the case of holocytochrome b_5, which contains a 5-stranded β-sheet and 6 α-helices,[72] the heme group is enclosed in a small irregular 4-α bundle. NMR data collected on apocytochrome b_5 indicate that the β-sheet is only 4-stranded in the absence of the heme.[73] The fifth, short strand has a low intrinsic propensity to form, a property not noticeable once the heme induces additional folding in the polypeptide chain. Holomyoglobin has yet a different fold, all helical, and its apoprotein is also endowed with a large extent of holoprotein structure. In all three apoproteins, sufficient tertiary interactions are established for the thermal unfolding to be cooperative, apparently two-state, and to involve excess enthalpy.[1,74,75]

The characterization of non-native species is often problematic. The disruption of native intramolecular interactions is accompanied by an enhanced tendency to establish intermolecular interactions and by an increase in conformational sampling.[76] Non-native forms are seen to exhibit various degrees of secondary structure, tertiary structure, and internal motions, in large or small, contiguous or discontinuous portions of the polypeptide. Their thermodynamic properties also vary. The structural view that emerges from the NMR analysis of the N state of apoMb is that it, like that of apocytochrome b_5, is a partially folded state closely related to the N state of the holoprotein and is best depicted as a heterogeneous structure. It contains well-folded regions remarkably similar to the corresponding regions of the native holoprotein and amenable to the traditional methods of structural determination in solution. Juxtaposed to these, other, less defined regions carrying the specificity for the ultimate fold, depend on the heme to fix their encoded conformation. The high affinity for the heme group and the rapid kinetics of heme insertion make apoMb a potential template for the design of heme-binding proteins. Efforts along those lines have focused on heme ligation[77] and could gain by incorporating the hybrid structural features of apoMb.

ACKNOWLEDGMENTS

Acknowledgment is made to The Petroleum Research Fund, administered by the American Chemical Society, for the partial support of this research. Partial support was also provided by the National Institutes of Health, grant DK43101 and its continuation GM54217. The authors thank Drs. Baldwin, Hughson, and Barrick for the generous gift of the myoglobin mutant genes, Dr. Barrick for advice on the purification methods, Drs. Barrick and Falzone for useful discussions, and Dr. Matthews for critical comments and advice.

REFERENCES

1. Griko, Y.V., Privalov, P.L., Venyaminov, S.Y., Kutyshenko, V.P. Thermodynamic study of the apomyoglobin structure. J. Mol. Biol. 202:127–138, 1988.
2. Griko, Y.V., Privalov, P.L. Thermodynamic puzzle of apomyoglobin unfolding. J. Mol. Biol. 235:1318–1325, 1994.
3. Breslow, E., Koehler, R. Combination of protoporphyrin IX with sperm whale myoglobin. J. Biol. Chem. 240:PC2266–PC2268, 1965.
4. Harrison, S.C., Blout, E.R. Reversible conformational changes of myoglobin and apomyoglobin. J. Biol. Chem. 240:299–303, 1965.
5. Hughson, F.M., Wright, P.E., Baldwin, R.L. Structural characterization of a partly folded apomyoglobin intermediate. Science 249:1544–1548, 1990.
6. Antonini, A., Brunori, M.: "Hemoglobin and Myoglobin in Their Reactions With Ligands." Amsterdam: North-Holland, 1971.
7. Adams, P.A. The kinetics and mechanism of the recombination reaction between apomyoglobin and haemin. Biochem. J. 159:371–376, 1976.
8. Adams, P.A. The kinetics of the recombination reaction between apomyoglobin and alkaline haematin. Biochem. J. 163:153–158, 1977.
9. Yang, A.-S., Honig, B. Structural origins of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin. J. Mol. Biol. 237:602–614, 1994.
10. Tirado-Rives, J., Jorgensen, W.L. Molecular dynamics simulations of the unfolding of apomyoglobin in water. Biochemistry 32:4175–4184, 1993.
11. Brooks, C.L. Characterization of "native" apomyoglobin by molecular dynamics simulation. J. Mol. Biol. 227:375–380, 1992.
12. Soman, K.V., Karimi, A., Case, D.A. Unfolding of an α-helix in water. Biopolymers 31:1351–1361, 1991.
13. Hirst, J.D., Brooks, C.L., III Molecular dynamics simulations of isolated helices of myoglobin. Biochemistry 34: 7614–1621, 1995.
14. Hargrove, M.S., Krzywda, S., Wilkinson, A.J., Dou, Y., Ikeda-Saito, M., Olson, J.S. Stability of myoglobin: A model for the folding of heme proteins. Biochemistry 33: 11767–11775, 1994.
15. Zhong, M., Lin, L., Kallenbach, N.R. A method for probing the topography and interactions of proteins: Footprinting of myoglobin. Proc. Natl. Acad. Sci. USA 92:2111–2115, 1995.
16. Cocco, M.J., Lecomte, J.T.J. Characterization of hydrophobic cores in apomyoglobin: A proton NMR spectroscopy study. Biochemistry 29:11067–11072, 1990.
17. Cocco, M.J., Lecomte, J.T.J. The native state of apomyoglobin described by proton NMR spectroscopy: Interaction with the radical HyTEMPO and the fluorescent dye ANS. Protein Sci. 3:267–281, 1994.
18. Jennings, P.A., Wright, P.E. A molten globule intermediate formed early on the kinetic folding pathway of apomyoglobin. Science 262:892–896, 1993.
19. Kirby, E.P., Steiner, R.F. The tryptophan environment in apomyoglobin. J. Biol. Chem. 245:6300–6306, 1970.
20. Hughson, F.M., Barrick, D., Baldwin, R.L. Probing the stability of a partly folded apomyoglobin intermediate by site-directed mutagenesis. Biochemistry 30:4113–4118, 1991.
21. Hughson, F.M., Baldwin, R.L. Use of site-directed mutagenesis to destabilize native apomyoglobin relative to folding intermediates. Biochemistry 28:4415–4422, 1989.
22. Bismuto, E., Colonna, G., Irace, G. Unfolding pathway of myoglobin. Evidence for a multistate process. Biochemistry 22:4166–4174, 1983.
23. Irace, G., Bismuto, E., Savy, F., Colonna, G. Unfolding pathway of myoglobin: Molecular properties of intermediate forms. Arch. Biochem. Biophys. 244:459–469, 1986.
24. Barrick, D., Hughson, F.M., Baldwin, R.L. Molecular mechanisms of acid denaturation. The role of histidine residues in the partial unfolding of apomyoglobin. J. Mol. Biol. 237:588–601, 1994.
25. Bismuto, E., Sirangelo, I., Irace, G. Salt-induced refolding of myoglobin at acidic pH: Molecular properties of a partially folded intermediate. Arch. Biochem. Biophys. 298: 624–629, 1992.
26. Goto, Y., Fink, A.L. Phase diagram for acidic conforma-

tional states of apomyoglobin. J. Mol. Biol. 214:803–805, 1990.
27. Barrick, D., Baldwin, R.L. The molten globule intermediate of apomyoglobin and the process of protein folding. Protein Sci. 2:869–876, 1993.
28. Cheng, X., Schoenborn, B.P. Neutron diffraction study of carbonmonoxymyoglobin. J. Mol. Biol. 220:381–399, 1991.
29. Dalvit, C., Wright, P.E. Assignment of resonances in the ^{1}H nuclear magnetic resonance spectrum of the carbon monoxide complex of sperm whale myoglobin by phase-sensitive two-dimensional techniques. J. Mol. Biol. 194: 313–327, 1987.
30. Cocco, M.J., Kao, Y.-H., Phillips, A.T., Lecomte, J.T.J. Structural comparison of apomyoglobin and metaquomyoglobin: pH titration of histidines by NMR spectroscopy. Biochemistry 31:6481–6491, 1992.
31. Drobny, G., Pines, A., Sinton, S., Weitekamp, D.P., Wemmer, D. Fourier transform multiple quantum nuclear magnetic resonance. Symp. Faraday Soc. 13:49–55, 1979.
32. Marion, D., Wüthrich, K. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. Biochem. Biophys. Res. Commun. 113:967–974, 1983.
33. Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., Wüthrich, K. Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. Biochem. Biophys. Res. Commun. 117:479–485, 1983.
34. Kumar, A., Ernst, R.R., Wüthrich, K. A 2D NOE experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem. Biophys. Res. Commun. 95:1–6, 1980.
35. Bodenhausen, G., Kogler, H., Ernst, R.R. Selection of coherence-transfer pathways in NMR pulse experiments. J. Magn. Reson. 58:370–388, 1984.
36. Davis, D.G. Elimination of baseline distortions and minimization of artifacts from phased 2D NMR spectra. J. Magn. Reson. 81:603–607, 1989.
37. Braunschweiler, L., Ernst, R.R. Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. J. Magn. Reson. 53:521–528, 1983.
38. Rance, M. Improved techniques for homonuclear rotating frame and isotropic mixing experiments. J. Magn. Reson. 74:557–564, 1987.
39. Shaka, A.J., Lee, C., Pines, A. Iterative schemes for bilinear operators; application to spin decoupling. J. Magn. Reson. 77:274–293, 1988.
40. Cavanagh, J., Rance, M. Suppression of cross-relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence. J. Magn. Reson. 96:670–678, 1992.
41. Piotto, M., Saudek, V., Sklenar, V. Gradient-taylored excitation for single-quantum NMR spectroscopy of aqueous solutions. J. Biomol. NMR 2:661–665, 1992.
42. Otting, G., Widmer, H., Wagner, G., Wüthrich, K. Origin of t_1 and t_2 ridges in 2D NMR spectra and procedures for suppression. J. Magn. Reson. 66:187–193, 1986.
43. Bax, A., Ikura, M., Kay, L.E., Zhu, G. Removal of F1 baseline distortion and optimization of folding in multidimensional NMR spectra. J. Magn. Reson. 91:174–178, 1991.
44. Takano, T. Structure of myoglobin refined at 2.0 Å resolution. I. Crystallographic refinement of sperm whale metmyoglobin. J. Mol. Biol. 110:537–568, 1977.
45. Takano, T. Refinement of myoglobin and cytochrome *c*. In: "Methods and Applications in Crystallographic Computing." Hall, S.R., Ashida, T. (eds.). Oxford: Oxford University Press, 1984:262–272.
46. Cross, K.J., Wright, P.E. Calibration of ring-current models for the heme ring. J. Magn. Reson. 64:220–231, 1985.
47. Haigh, C.W., Mallion, R.B. Ring current theories in nuclear magnetic resonance. Prog. NMR Spec. 13:303–344, 1980.
48. Ösapay, K., Case, D.A. A new analysis of proton chemical shifts in proteins. J. Am. Chem. Soc. 113:9436–9444, 1991.
49. Brünger, A.T. "X-PLOR. Version 3.1. A System for X-Ray Crystallography and NMR." New Haven: Yale University Press, 1992.
50. Wagner, G., Braun, W., Havel, T.S., Schaumann, T., Go, N., Wüthrich, K. Protein structure in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. J. Mol. Biol. 196:611–639, 1987.
51. Nilges, M., Clore, G.M., Gronenborn, A.M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. FEBS Lett. 229:317–324, 1988.
52. Nilges, M., Kuszewski, J., Brünger, A.T. Sampling properties of simulated annealing and distance geometry. In: "Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy." Hoch J.C., Poulsen, F.M., Redfield, C. (eds). New York: Plenum Press, pp. 451–455, 1991.
53. Brooks B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy minimization and dynamics calculations. J. Comput. Chem. 4:187–217, 1983.
54. Petros, A.M., Mueller, L., Kopple, K.D. NMR identification of protein surfaces using paramagnetic probes. Biochemistry 29:10041–10048, 1990.
55. Wishart, D.S., Sykes, B.D., Richards, F.M. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J. Mol. Biol. 222:311–333, 1991.
56. Gurd, F.R.N., Friend, S.H., Rothgeb, T.M., Gurd, R.S., Scouloudi, H. Electrostatic stabilization in sperm whale and harbor seal myoglobins. Biophys. J. 10:65–75, 1980.
57. Kao, Y.-H., Lecomte, J.T.J. Determination of the zero-field splitting constant for proton NMR chemical shift analysis in metaquomyoglobin. The dipolar shift as a structural probe. J. Am. Chem. Soc. 115:9754–9762, 1993.
58. Nishii, I., Kataoka, M., Tokunaga, F., Goto, Y. Cold denaturation of the molten globule states of apomyoglobin and a profile for protein folding. Biochemistry 33:4903–4909, 1994.
59. Fermi, G., Perutz, M.F., Shaanan, B., Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74 Ångstroms resolution. J. Mol. Biol. 175:159–174, 1984.
60. Kim, K.-S., Tao, F., Fuchs, J., Danishefsky, A.T., Housset, D., Wlodawer, A., Woodward, C. Crevice-forming mutants of bovine pancreatic trypsin inhibitor: Stability changes and new hydrophobic surfaces. Protein Sci. 2:588–596, 1993.
61. Eriksson, A.E., Baase, W.A., Zhang, X.J., Heinz, D.W., Blaber, M., Baldwin, E.P., Matthews, B.W. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255:178–183, 1992.
62. Baase, A.W., Eriksson, A.E., Zhang, X.J., Heinz, D.W., Sauer, U., Blaber, M., Baldwin, E.P., Wozniak, J.A., Matthews, B.W. Dissection of protein structure and folding by site-directed mutagenesis. Faraday Discuss. 93:173–181, 1992.
63. Eriksson, A.E., Baase, W.A., Matthews, B.W. Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme and different structural and dynamic consequences. J. Mol. Biol. 229:747–769, 1993.
64. Hendsch, Z.S., Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci. 3:211–226, 1994.
65. Pinker, R.J., Lin, L., Rose, G.D., Kallenbach, N.R. Effects of alanine substitutions in α-helices of sperm whale myoglobin on protein stability. Protein Sci. 2:1099–1105, 1993.
66. Chiba, K., Ikai, A., Kawamura-Konishi, Y., Kihara, H. Kinetic study on myoglobin refolding monitored by five optical probe stopped-flow methods. Proteins 19:110–119, 1994.
67. De Sanctis, G., Ascoli, F., Brunori, M. Folding of apominimyoglobin. Proc. Natl. Acad. Sci. USA 91:11507–11511, 1994.
68. Hirst, J.D., Brooks, C.L., III. Helicity, circular dichroism and molecular dynamics of proteins. J. Mol. Biol. 243:173–178, 1994.
69. Theriault, Y., Pochapsky, T.C., Dalvit, C., Chiu, M.L., Sligar, S.G., Wright, P.E. ^{1}H and ^{15}N resonance assignments and secondary structure of the carbon monoxide complex of sperm whale myoglobin. J. Biomol. NMR 4:491–504, 1994.
70. Srinivasan, R., Rose, G.D. LINUS: A hierarchic procedure to predict the fold of a protein. Proteins 22:81–99, 1995.

71. Feng, Y., Sligar, S.G., Wand, A.J. Solution structure of apocytochrome b_{562}. Nature Struct. Biol. 1:30–35, 1994.
72. Mathews, F.S., Czerwinski, E.W., Argos, P.: The X-ray crystallographic structure of calf liver cytochrome b_5, 107–147. "The Porphyrins." New York: Academic Press, 1979.
73. Moore, C.D., Lecomte, J.T.J. Characterization of an independent structural unit in apocytochrome b_5. Biochemistry 32:199–207, 1993.
74. Pfeil, W. Thermodynamics of apocytochrome b_5 unfolding. Protein Sci. 2:1497–1501, 1993.
75. Feng, Y., Sligar, S.G. Effect of heme binding on the structure and stability of *Escherichia coli* apocytochrome b_{562}. Biochemistry 30:10150–10155, 1991.
76. Baldwin, R.L. Molten globules: Specific or non-specific intermediates. Chemtracts: Biochem. Mol. Biol. 2:379–389, 1991.
77. Robertson, D.E., Farid, R.S., Moser, C.C., Urbauer, J.L., Mulholland, S.E., Pidikiti, R., Lear, J.D., Wand, A.J., DeGrado, W.F., Dutton, P.L. Design and synthesis of multihaem proteins. Nature 368:425–432, 1994.