Biochemistry

 Volume 35, Number 21 May 28, 1996

Accelerated Publications

Design Challenges for Hemoproteins: The Solution Structure of Apocytochrome b_5[†,‡]

Christopher J. Falzone,* Michael R. Mayer, Eileen L. Whiteman, Cathy D. Moore,[§] and Juliette T. J. Lecomte*

Department of Chemistry and the Center for Biomolecular Structure and Function, The Pennsylvania State University, University Park, Pennsylvania 16802

Received February 29, 1996; Revised Manuscript Received April 15, 1996[⊗]

ABSTRACT: In order to characterize the structural and dynamic factors that determine the assembly in *b* hemoproteins, the solution structure of the 98-residue protein apocytochrome b_5 was determined by NMR methods. Over 800 experimental restraints derived from a series of two- and three-dimensional experiments were used. Holocytochrome b_5, the protein with iron protoporphyrin-IX liganded to His-39 and His-63, contains in sequence the following elements of secondary structure: $\beta1-\alpha1-\beta4-\beta3-\alpha2-\alpha3-\beta5-\alpha4-\alpha5-\beta2-\alpha6$ [Mathews, F. S., Czerwinski, E. W., & Argos, P. (1979) *The Porphyrins*, Vol. 7, pp 107−147, Academic Press, New York]. The folded holoprotein possesses two hydrophobic cores: an extensive, functional core around the heme (core 1), and a smaller, structural core remote from the heme (core 2). The apoprotein was found to contain a stable four-stranded β-sheet encompassing $\beta1$, $\beta2$, $\beta3$, and $\beta4$ and three α-helices, corresponding to α1, α2, and α6. Two short α-helices (α3 and α5) appear to form partially, and α4 is not detected. These three helices and $\beta5$ border the heme binding pocket and are disordered in the apoprotein NMR structure. According to backbone $^1H-^{15}N$ NOE results, the most flexible region of the apoprotein, except for the termini, extends from Ala-50 (in $\beta5$) to Glu-69 (in α5). The polypeptide segment bearing His-63 (located immediately prior to α5) exhibits faster internal motions than that bearing His-39 (at the C-terminal end of α2). The latter imidazole samples a restricted region of space, whereas the former can adopt many orientations with respect to the stable core. It was concluded that heme removal affects the structure and dynamics of most of core 1 whereas it leaves core 2 largely intact. The results provide guidelines for the rational design of *b* hemoproteins: a modular structure including a packed, stable core and a partially folded binding site is anticipated to present strong kinetic and thermodynamic advantages compared to approaches relying on the complete formation of secondary structure prior to heme binding.

The *de novo* design of a protein endowed with a specified function has become a realistic goal of biomolecular engineering. Of particular interest is the construction of a polypeptide scaffold to support a chemical moiety capable of interactions different from those offered by amino acid side chains alone. Hemoproteins illustrate this chemical versatility as they use the same heme group to transport electrons or bind small ligands. There are two main structural challenges in the design of hemoproteins: those associated with the production of a polypeptide matrix that will fold spontaneously into the desired shape, and those

† This work was supported by the National Institutes of Health Grant DK-43101 and its continuation GM-54217.

‡ Coordinates have been submitted to the Brookhaven Protein Data Bank and are listed under file names 1IET (minimized average) and 1IEU (ten models).

* To whom correspondence should be addressed at the Department of Chemistry, 152 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802.

§ Current address: Department of Biochemistry, Brandeis University, Waltham, MA 02254.

⊗ Abstract published in *Advance ACS Abstracts*, May 15, 1996.

S0006-2960(96)00501-6 CCC: $12.00

associated with the unique binding of a prosthetic group. Current minimalist efforts are inspired by simple motifs, such as the pair of parallel helices of cytochrome bc_1 (Robertson et al., 1994), and emphasize the engineering of stable folded structure with judiciously located histidyl ligands. In general, these designs are likely to be complicated by a critical dependence on the internal hydrophobic residues, which must simultaneously satisfy the requirements of secondary structure, tertiary structure, and binding.

Rigid shape complementation does not ensure that the desired product will be the sole thermodynamic or kinetic outcome: developing a hydrophobic anchor independently of binding has an entropic benefit but can lead to indiscriminate non-reversible association through apolar groups. Building from malleable elements is an alternative where the interacting surface is induced only upon binding. An example is found in the dimerization-linked folding of coiled coil peptides (Zitzewitz et al., 1995). To emulate the exquisite control exercised by natural systems through flexible pieces requires a deep understanding of structure and folding determinants, and such information would be useful in the design of hemoproteins.

A solution to this design problem involves the decoupling, at least to an extent, of folding and heme-binding considerations and calls for modules dedicated to each of these tasks. Modular construction is a convenient avenue for optimization of solubility and stability and provides a readily and independently tunable template for binding alternative prosthetic groups. Natural archetypes for modular design are *b* hemoproteins of complex architectures, such as myoglobin and cytochrome b_5 (cyt b_5).[1] The necessary clues for sound design cannot be obtained from inspection of the finished holoprotein products but are displayed in the apoproteins. These are neither fully folded nor fully unfolded precursors existing under native conditions, and unlike their holoprotein counterparts, they contain both dynamic and structural information. Thus, apo *b* hemoproteins offer unique systems to explore how evolution has devised a prosthetic group-binding protein, for they have inscribed in their fold the key properties for correct assembly. It is the description of one such apoprotein, from cytochrome b_5, that is undertaken here.

The water-soluble fragment of cyt b_5 contains a single heme group and functions as an electron transport protein. The bovine variant has a known three-dimensional structure in the crystalline state (Mathews et al., 1979). This small protein contains a five-stranded β-sheet of $(+3x,-1,+2,-3x)$ topology and six α-helices. The β-sheet and part of $\alpha1$ make a β-barrel whose cylindrical center is filled with hydrophobic side chains. These, from $\alpha1$, $\beta2$, $\beta3$, and $\beta4$, pack with a residue from $\alpha6$ to constitute crystallographic "core 2." The heme prosthetic group (6% of the total mass) is intimately associated with the polypeptide matrix on one side of this core. Two imidazole rings, His-39 and His-63, coordinate to the iron through their N^{ϵ} atom. The outer face of $\beta3$ and $\beta4$, along with the central segment $\alpha2-\alpha3-\beta5-\alpha4-\alpha5$, contribute 14 side chains in contact with the heme group. These residues are in majority apolar and form the larger crystallographic "core 1." Whereas hydrophobic core 1 has a clear functional role, the highly conserved hydrophobic core 2 appears to have only a structural role. Recombinant rat liver cyt b_5 (the protein used here) has been studied by NMR spectroscopy (Guiles et al., 1993) and displays NOEs defining a secondary and tertiary structure in solution closely similar to that of the beef holoprotein in the crystal.

In the absence of the heme group, the cytochrome fold is less stable (Pfeil, 1993) and less structured (Huntley & Strittmatter, 1972). This destabilized species can be investigated in solution by NMR techniques at room temperature. In previous ^{1}H NMR studies, rat liver apocyt b_5 was shown to possess a stable core corresponding to core 2 (Moore & Lecomte, 1990; Moore et al., 1991). Stopped-flow kinetic experiments monitoring the fluorescence of one of its residues, Trp-22, demonstrated that this core is capable of rapid, unaided folding from denaturing conditions in the absence of the heme group. Further characterization using 2D heteronuclear NMR experiments failed to reveal typical signatures of regular secondary structure in a stretch of residues located in the heme-binding region (Moore & Lecomte, 1993). The undetected elements were assumed missing, and apocyt b_5 was depicted as a two-module fold, core 2 acting as an independent structural unit, loosely organizing the binding site.

Herein, NMR spectroscopy is applied to determine the solution structure of apocyt b_5. Preliminary backbone dynamics data are used to complement the description of the heme-binding site. The new data confirm the validity of a modular approach to hemoprotein mimicry and identify specific features to be incorporated into future designs.

MATERIALS AND METHODS

Protein Preparation and Purification. The original pUC-19 construct for the rat cyt b_5 gene (Beck von Bodman et al., 1986) was provided by Dr. S. G. Sligar. This gene is flanked by restriction sites for *Pst* I and *Eco*RI. By applying PCR methods, these sites were changed to *Nco*I and *Bam*HI sites, for insertion into a pET-3d plasmid. The product of the ligation was transformed into the *Escherichia coli* cell strain DH5α and colonies were screened for the cytochrome gene on ampicillin plates. The plasmid, pET-3d(*cytb*$_5$), was found and isolated, and its sequence was verified. It was then transformed into BL21(DE3) *E. coli* cell line; pET-3d(*cytb*$_5$) utilizes the T7 promoter (Studier et al., 1990) and produces large amounts of protein in this cell line.

Cells harboring pET-3d(*cytb*$_5$) were induced with IPTG after approximately 3 h of growth in M9 medium (Maniatis et al., 1982); 8 h later the faintly pink cells were harvested and lysed with lysozyme. With this T7 expression system, the majority of the protein is found in the apoprotein form, but purification was carried out on the holoprotein to improve

[1] Abbreviations: 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional; ANS, 8-anilino-1-naphthalenesulfonic acid; CD, circular dichroism; cyt b_5, cytochrome b_5; DG, distance geometry; DIPSI, decoupling in the presence of scalar interactions; 2QF-COSY, two-dimensional double-quantum-filtered correlated spectroscopy; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulfonate; GARP, globally optimized alternating phase rectangular pulse; HyTEMPO, 4-hydroxy-2,2,6,6,-tetramethylpiperidinyl-1-oxy; IPTG, isopropyl β-thiogalactoside; MD, molecular dynamics; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser spectroscopy; PCR, polymerase chain reaction; pH*, pH meter reading, uncorrected for isotope effects; PP-IX, protoporphyrin IX; rms, root mean square; rmsd, root mean square deviation; SA, simulated annealing; TOCSY, total correlation spectroscopy; TPPI, time proportional phase incrementation.

Table 1: Structural Statistics and Atomic RMS Deviations

	Structural Statistics	
	{SA}[a]	<SA>$_r$[b]
rms deviation from experimental constraints		
distance constraints (Å) (767)[c]	0.0184 ± 0.0018	0.0148
dihedral constraints (deg) (50)[d]	0.15 ± 0.08	0.036
rms deviation from ideal geometry		
bonds (Å)	0.0030 ± 0.0002	0.0028
angles (deg)	0.376 ± 0.018	0.35
impropers (deg)	0.302 ± 0.019	0.285
number of violations[e]		
distance constraints	76 ± 5	71
dihedral constraints	5 ± 2	2
	RMS Deviations from the Average Structure (Å)[f]	
	all residues[g]	limited set[h]
backbone (N-C$^\alpha$-C)	2.19 ± 0.37	0.50 ± 0.11
side chains (heavy atom)	2.60 ± 0.26	0.92 ± 0.11

[a] {SA} is the family of final 15 simulated annealing structures; mean values and standard deviations are shown for these 15 structures. [b] <SA>$_r$ is the structure obtained by applying 5000 steps of restrained minimization to the mathematical average structure. The mathematical average structure was obtained by first aligning the 15 structures with a least-squares fit over the regular secondary structure [(N-C$^\alpha$-C) of residues 5−13, 21−25, 33−37, 73−76 and 82−84] and then individually averaging the atomic coordinates. [c] A total of 767 distance restraints were used in the structure calculation: 192 were long-range ($|i - j| > 5$); 76 were short-range ($1 < |i - j| \leq 5$); 222 were sequential ($|i - j| = 1$); 255 were intraresidue; and 11 × 2 were H-bond constraints. [d] A total of 50 dihedral constraints were used in the structure calculations: 33 were backbone ϕ angles; 17 were side chain χ_1 angles. The values were $\phi = -50 \pm 40°$ or $-150 \pm 50°$, and $\chi_1 = 60$, -60, or 180°, all ±30°. [e] *All* distance violations are less than 0.31 Å, and *all* dihedral violations are less than 1.6°. [f] The rms deviations of the atomic positions were calculated for each structure in {SA} against the mathematical average structure. The average values and standard deviations for the 15 final structures are reported. [g] The rmsd for residues 3−88. [h] The rmsd for residues 5−13, 21−25, 33−37, 73−76, and 82−84. The side chain rmsd's are calculated using nonsurface residues from these regions and include 4, 6−9, 22−25, 32, 33, 35, 36, 73−76, 80, 81, and 84.

stability and ease of detection. To convert the apoprotein into the holoprotein, a saturated solution of heme in 1 M NaOH was titrated to excess into the supernatant. This solution was purified according to the usual procedure (Beck von Bodman et al., 1986). The yield of pure protein is ca. 100 mg per liter in M9 medium. The same procedure was applied for ^{15}N labeling, yielding about 90 mg of pure protein per liter with [^{15}N]ammonium chloride (Isotec, Inc.) as the sole nitrogen source. Apocyt b_5 was prepared by the method of Teale (1959; Moore & Lecomte, 1990). Apoprotein concentrations were estimated by using an extinction coefficient of 10.6 mM^{-1} cm^{-1} at 280 nm.

ANS Binding. The binding of ANS to apocyt b_5 was monitored by fluorescence spectroscopy (Spex Fluorolog 1681). Excitation was performed at 370 nm, and emission was recorded from 400 to 600 nm. Samples were 3 μM in apocyt b_5 at room temperature, pH 7.5 (10 mM phosphate buffer), and contained amounts of ANS ranging from 0 to 150 μM. To verify the integrity of the apoprotein, Trp fluorescence spectra were recorded (excitation at 295 nm and emission between 305 and 460 nm) as well as CD spectra. Horse apomyoglobin [prepared as in Cocco and Lecomte (1990)], to which ANS is known to bind (Stryer, 1965), was used as a control protein.

NMR Spectroscopy. ^{1}H NMR spectra were acquired as previously described on a Bruker AM-500 spectrometer at 298 K (Moore & Lecomte, 1993). Additional data were collected on a Bruker AMX2-500 spectrometer in 95% H_2O/5% ^2H_2O with a self-shielded *z*-gradient probe and modified WATERGATE sequence (Piotto et al., 1992; Sklenár et al., 1993) to suppress the water signal. Three-dimensional ^{1}H−^{15}N NOESY-HSQC and TOCSY-HSQC data were collected on both Bruker AM and AMX2-500 spectrometers at 298 K. AM-500 experiments are described in Falzone et al. (1994b). The mixing times were 100 ms (NOESY) and 40 ms (TOCSY with relaxation-compensated DIPSI mixing scheme; Cavanagh & Rance, 1992). The number of complex data points collected in each dimension was 1024 (F3), 32 (F2), and 128 (F1), with respective spectral widths of 7042, 1500, and 6900 Hz. Quadrature was achieved in the indirect ^{1}H dimension by using TPPI (Drobny et al., 1979) and in the ^{15}N dimension by using the TPPI-states method (Marion et al., 1989). The data were zero-filled once in each of the indirect dimensions and a squared sine-bell window with shift of 45−60° multiplied the data in each dimension.

Two- and three-dimensional NOESY gave 745 interproton restraints (Table 1). Approximate interproton distances were calibrated with the single Trp C$^{\zeta 3}$H-C$^\eta$H cross-peak volume. The NOEs were qualified as strong, medium, or weak, with respective distance ranges of 1.8−2.7 Å, 1.8−3.3 Å, and 1.8−5.5 Å. Methyl groups were treated as X-PLOR pseudoatoms (Brünger, 1992) with $<r^{-6}>$ averaging; 0.5 Å was added to the upper bound, and the allowable lower limit was unchanged. No distinction was made between C$^{\delta 1}$H and C$^{\delta 2}$H (C$^{\epsilon 1}$H and C$^{\epsilon 2}$H) of Tyr or Phe rings, and $<r^{-6}>$ averaging was applied to these as well.

Protected backbone NH's were identified with a crude exchange experiment. A total of 12 mg of lyophilized ^{15}N-labeled apoprotein was dissolved in 450 μL of ^2H_2O at pH* 6.7, and after a rapid adjustment of spinning shims, an initial ^{1}H 1D spectrum was collected. A ^{1}H−^{15}N HSQC spectrum was then acquired in 5.1 min (4 transients, 64 t_1 increments). Collection of the first data set was completed 5.8 min after dissolution. Subsequent HSQC spectra were collected and after 11 min, no intensity was detectable. These data yielded 11 H-bond restraints.

J-modulated ^{1}H−^{15}N HSQC data with ^{15}N decoupling (Neri et al., 1990; Billeter et al., 1992) were obtained for 12 *J* evolution times τ between 20 and 100 ms in order to determine ϕ backbone angles. Cross-peak intensities were measured as a function of τ and fit to an exponentially dampened oscillation. A total of 33 ϕ angle restraints were

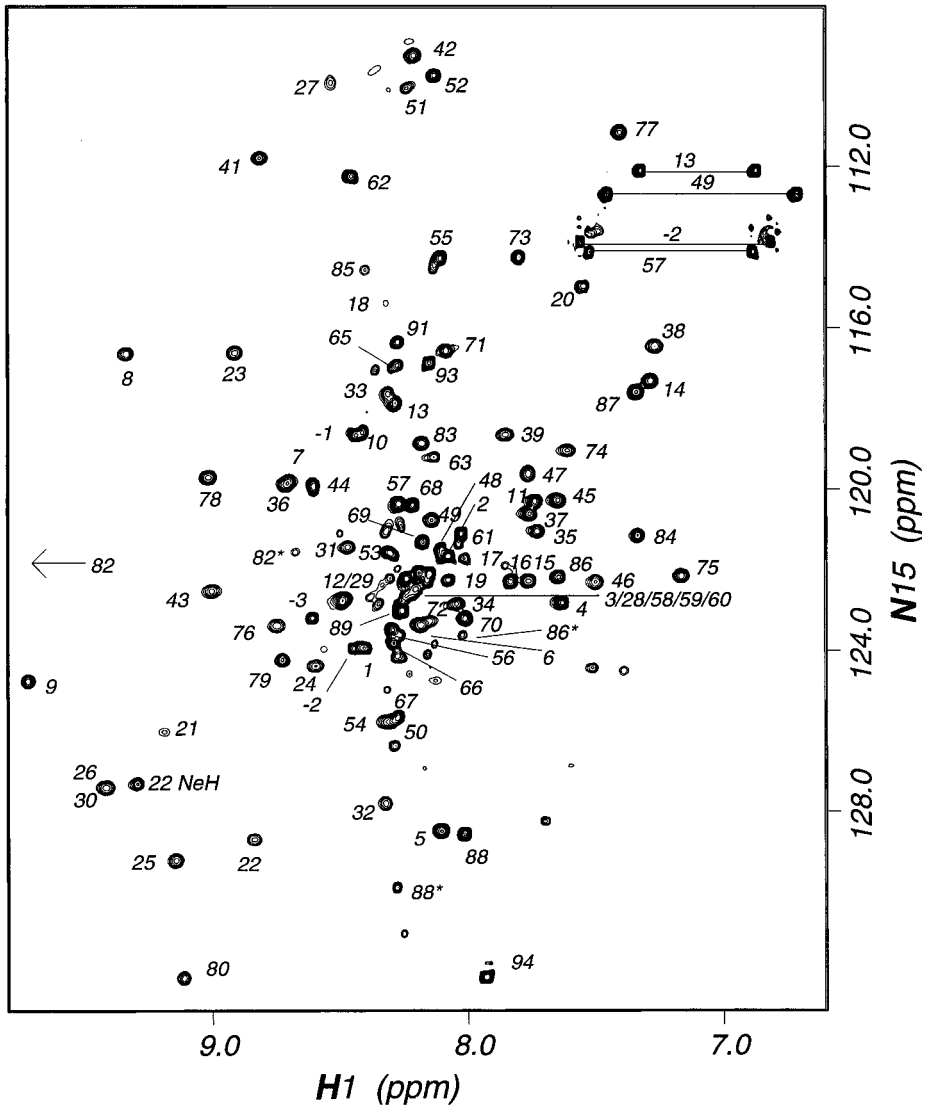

FIGURE 1: Assigned $^1H-^{15}N$ HSQC spectrum of apocyt b_5 (pH 6.2, 298 K). The $N^{\epsilon}H$ of Trp-22 is marked W22; horizontal lines connect the two cross-peaks of side chain NH_2's. Assigned minor form signals are marked with an asterisk.

extracted from the $^3J_{NH-C\alpha H}$ outside the 6−8 Hz window. In addition, 17 side chain χ_1 dihedral angles were obtained with 2D H_NNH_{AB} COSY data, 3D ^{15}N NOESY-HSQC data, 3D ^{15}N TOCSY-HSQC data, and 2QF-COSY data as described in Falzone et al. (1994a) and references therein.

To survey the dynamics of the backbone, $^1H-^{15}N$ NOE data were obtained by acquiring two $^1H-^{15}N$ data sets, one without pre-irradiation of the proton resonances and one with pre-irradiation (Kay et al., 1989). For the irradiated spectrum, a 1-s relaxation period was followed by a 3-s GARP-1 pulse train (5.6 kHz centered on H_2O). Cross-peak intensities were measured in both spectra, and their ratios yielded the NOE factor.

Structure Calculations. Residues −4 to −1 were omitted from the cytochrome sequence as no medium- or long-range NOEs were detected from these amino acids. Structures were calculated with X-PLOR 3.1 (Brünger, 1992). A total of 767 distance restraints (including H-bonds), 33 ϕ, and 17 χ_1 angle restraints were applied. Two starting points were used: a straight chain (protocol detailed in Falzone et al., 1994a) and the X-ray structure (PDB file 3b5c; Mathews et al., 1979), with sequence adjusted for the C- and N-terminal extensions and the nine amino acid differences in the common segment. Either way yielded similar results, although the latter required minimal computational effort. The extended chain calculations will not be discussed further.

Simulated annealing was performed to produce 30 structures by heating the X-ray model to 1000 K and cooling to 100 K in 2000 steps in the presence of all experimental restraints with square well potentials, van der Waals terms, and the covalent geometry terms. Refinement of these 30 structures by using 5000 steps of restrained minimization (with final effective van der Waals radii set to 0.80 times the standard CHARMm energy function; Brooks et al., 1983) yielded the final 15 structures shown in Figure 3. These structures have distance violations smaller than 0.31 Å and dihedral violations smaller than 1.6°. In addition, X-PLOR 3.2 was used to assess the restrained MD trajectory of the flexible loop identified with heteronuclear NOE data (residues 42−68). An $<r^{-3}>$ time-average refinement protocol was applied (Gros et al., 1990; Bonvin et al., 1994) to generate a representative family of conformers.

RESULTS

Solution Structure of Apocytochrome b_5. Figure 1 presents a $^1H-^{15}N$ HSQC spectrum of apocyt b_5 at 298 K and pH 6.2. Sets of 3D TOCSY and NOESY data were collected

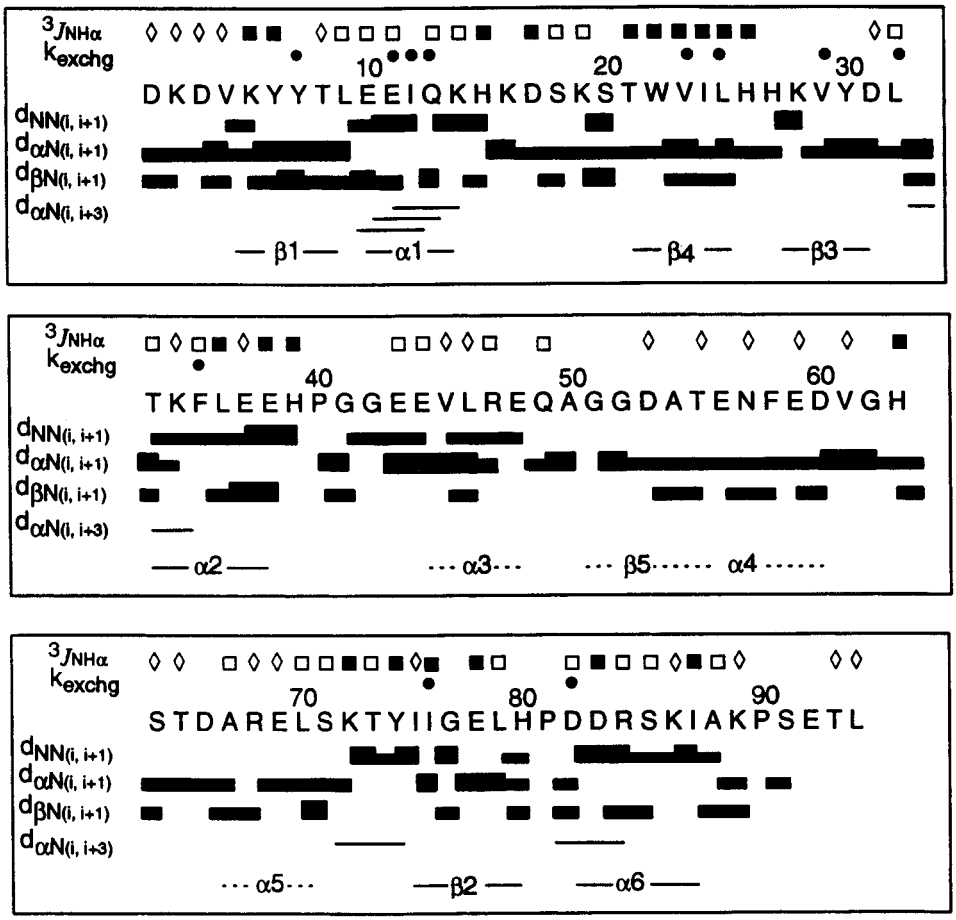

FIGURE 2: Summary of sequential information for apocyt b_5 at 298 K, pH 6.2. The top row contains $^3J_{NH-C\alpha H}$ data: (■) $J > 8$ Hz; (□) $J < 6$ Hz; (◊), intermediate values. In the next row, (●) mark the amide peaks remaining in the $^1H-^{15}N$ HSQC spectrum after apocyt b_5 was dissolved in ^2H_2O for 5 min at pH* 6.7. The bars below the sequence indicate the NOE connectivities: N_iN_{i+1}, α_iN_{i+1}, β_iN_{i+1}, α_iN_{i+3}. Thicker bars denote larger intensity. Regular secondary structure in the holoprotein is marked at the bottom. Dotted lines indicate structure not present or fully formed in the apoprotein structure.

to complete the ^1H and ^{15}N chemical shift analysis. With the new data, it was possible to assign approximately 98% of the backbone and 80% of the side chain resonances in spite of overlap in both ^1H and ^{15}N dimensions and the existence of a small population of a second form at the C-terminal end of the protein (Lecomte & Moore, 1991). Chemical shifts are listed in the supporting information. Backbone dihedral angle (ϕ) and backbone amide hydrogen exchange data are summarized in Figure 2. Sequential and medium-range NOEs derived from 2D and 3D data are also marked, along with the deduced stable secondary structure.

The apoprotein structure was calculated with over 800 restraints, or an average of nearly 9 restraints per residue over the 1−94 sequence. Figure 3 contains a family of 15 structures superimposed onto their average. Figure 3A shows the Cα trace from 3 to 87, whereas Figure 3B excludes the segment defining the heme-binding domain (41 to 67). For comparison, the trace from the X-ray structure of the bovine holoprotein (Mathews et al., 1979) is drawn in dashed lines, with heme group included. In what follows, the secondary structure elements of the apoprotein are denoted with "a" to distinguish them from those assigned in the X-ray structure.

In apocyt b_5, as in holocyt b_5 (Mathews et al., 1979), the N-terminal residues are disordered up to residue 3. The first β-strand (β1a, Val-4 to Tyr-7) is well defined and provides the leading edge of the β-sheet. From β1a, the polypeptide chain proceeds to α1a, the longest α-helix in the cytochrome. This helix is composed of residues Leu-9 to Lys-14. It starts with a capping box (Harper & Rose, 1993) and ends with a histidine C-cap (Presta & Rose, 1988). The backbone NH's at positions 11, 12, and 13 are protected and taken to form helical H-bonds. Helix α1a crosses over the sheet to lead into β4a (Thr-21 to Leu-25), which delineates the outer edge of the β-sheet. Poorly constrained loops connect α1a to β4a and β4a to β3a. The β3a strand is short (His-27 to Leu-32) and docks onto the C-terminal β-strand (β2a, Tyr-74 to His-80) in an antiparallel fashion. On the other side of β2a, and parallel to it, lies the first strand, β1a. The elements β1a, α1a, β4a, β3a, and β2a correspond closely to their holoprotein counterparts, even in detail such as the β bulge at position 75 (Moore & Lecomte, 1993). Doubled resonances are observed at the end of β2a; the slow conformational exchange takes place between two or more conformations (Lecomte & Moore, 1991), and only the major one is considered here.

Other secondary features are observed in the apoprotein. From strand β3a, an irregular α-helix is formed (α2a, Thr-33 to Glu-38). The X-ray structure of the holoprotein describes α2 as a series of turns, a characteristic which seems to be maintained in the apoprotein. Helix α2a contains at least one slowly exchanging backbone NH at position 35. The NMR fold suggests that 32 is the acceptor, in agreement with the X-ray structure of beef holocyt b_5. Beyond residue 40, the protein does not adopt sustained holoprotein-like

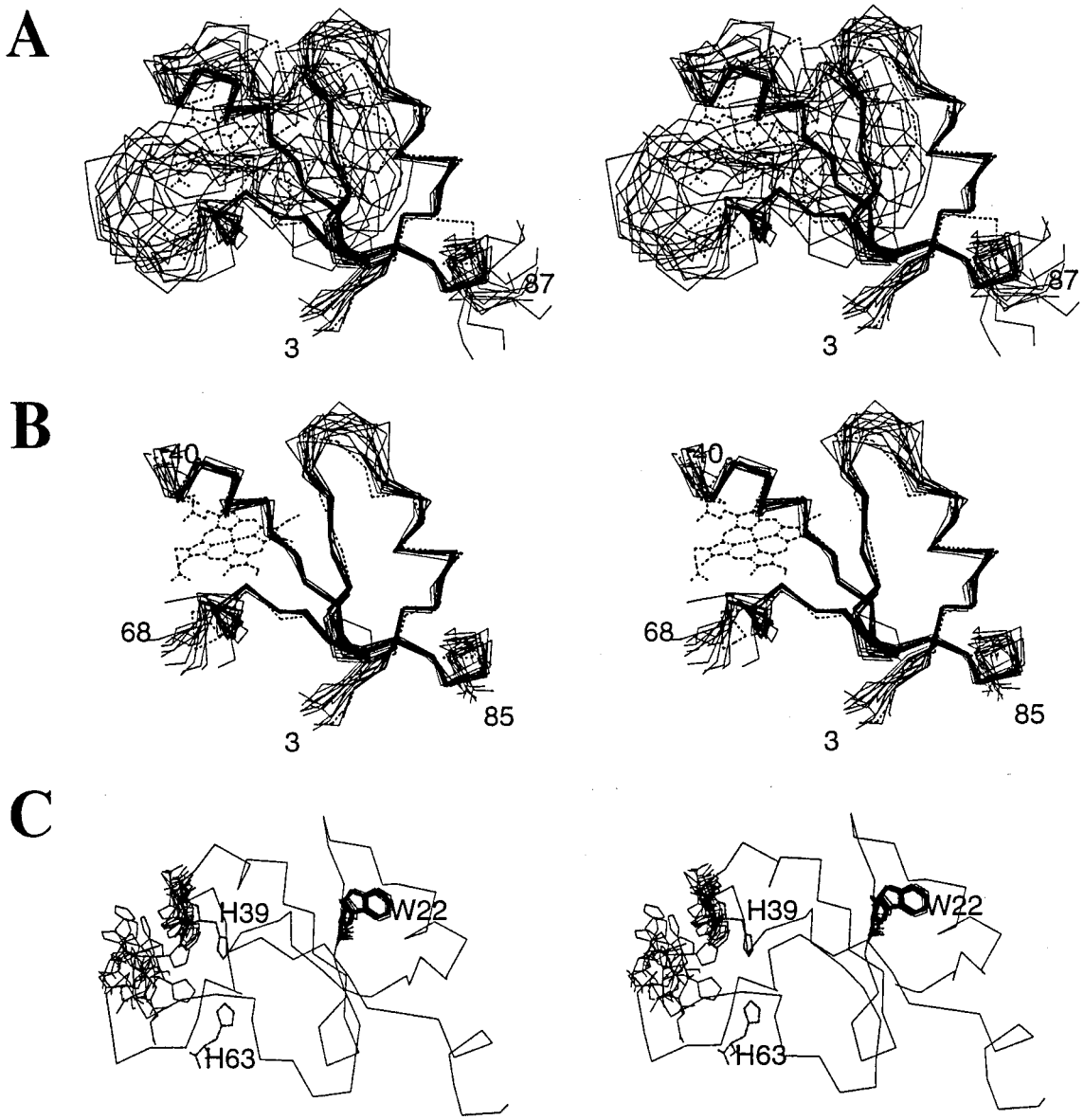

FIGURE 3: Stereoviews of the best-fit superposition of 15 apocyt b_5 structures. The fit was performed on the well-defined regions of the average structure (see Table 1). A. Cα trace from residue 3 to residue 87. The dashed line represents the bovine holoprotein in the crystalline state (PDB file 3b5c; Mathews et al., 1979) and with the heme group for reference. B. Ordered regions (3−40 and 68−85) of the same 15 structures. C. Conformational distribution for Trp-22, and the liganding histidines His-39 and His-63. The thin line is the Cα trace of the average structure to guide the eye, and the labeled side chains are from the X-ray structure. This is a conservative representation of His-63 as the structure calculations assumed a single conformer for the ill-defined loop. Time-averaging over residues 42−68 produces families of conformers with increased disorder.

structure until Glu-69. From residue 69, a single turn of α-helix is detected (α5a), leading into the C-terminal β-strand (β2a). Spectral overlap in the 65−69 region is such that N_iN_{i+1} NOEs would not be resolved if they were present. The starting turn of the last, short α-helix (α6a, Pro-81 to Arg-84) is detected; there is a kink at Ser-85, manifested by weak N_iN_{i+1} NOEs between 84 and 85, and strong to medium α-helical NOEs for residues Ser-85 to Ile-87. The C-terminal segment (Ala-88 to Leu-94) has no detectable medium- or long-range NOEs and appears disordered. Helix α2a matches well bovine α2, whereas α6a matches α6 only partly so. Two conformations are also observed in this C-terminal region (Lecomte & Moore, 1991).

The region extending from 40 to 69 is poorly defined. NOEs indicative of folded structure are found only at the ends of this segment and in the mid-forties, notably for residues Val-45 and Leu-46. On both sides of these, N_iN_{i+1} NOEs suggest a propensity for α-helical formation. This helix (α3a) does not form fully, or alternatively, samples several conformations rapidly as medium α_iN_{i+1} NOEs are also observed. Helix α3a is an altered version of α3 in the holoprotein (Glu-44 to Arg-47).

The elements of structure that are disrupted upon prosthetic group removal include the fifth strand of the β-sheet (β5, Gly-51 to Ala-54). Although resonance assignments could be made to this region, no strong dipolar contact was detected between its residues and the strand onto which they dock in the presence of the heme group. $^3J_{NH-C\alpha H}$ values are consistent with conformational averaging. In holocyt b_5, β5 is an irregular strand contributing side chains to hydrophobic core 1 but not to hydrophobic core 2, and in the apoprotein the absence of β5 does not interfere with the formation of core 2. In addition, no NOEs consistent with helix formation and intermediate $^3J_{NH-C\alpha H}$ values are seen for α4a (Asp-53

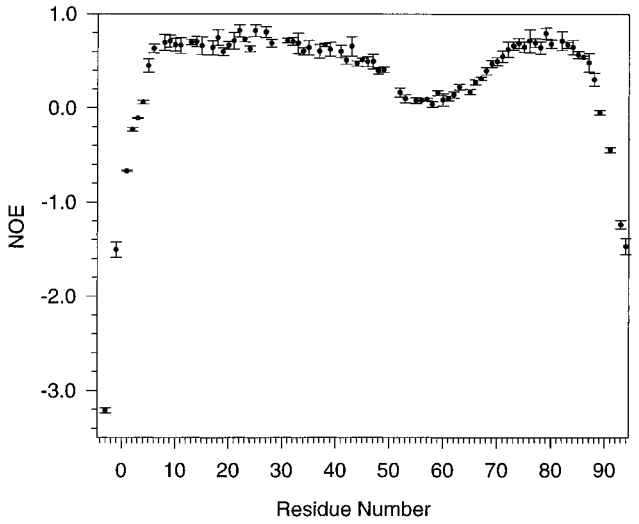

FIGURE 4: Amide backbone $^1H-^{15}N$ NOEs as a function of residue number. The average of three measurements is presented. Error bars indicate one standard deviation about the average. Several values are missing because of overlap or occurrence of prolines (at 40, 81, and 90).

to Asp-60). From Arg-47 to Arg-68, *i.e.*, the larger part of the heme pocket or core 1, apocyt b_5 displays little stable holo-like tertiary and secondary structure. Left as the only clearly recognizable elements near the heme-binding site are α2a, a distorted and possibly fluctuating α3a, and an abbreviated form of α5. The absence of long-range experimental constraints at the beginning of α5 distinguishes the two heme ligands (His-39 and His-63). Figure 3C illustrates these side chains and Trp-22. Whereas the orientation of Trp-22 in core 2 is well defined, that of His-39 appears less so, and His-63 samples a comparatively large conformational volume.

$^1H-^{15}N$ NOE Measurements. Many ^1H resonances in apocyt b_5 are unresolved, and it is necessary to ascertain the origin of the lack of constraints in the 40−69 region of the structure. A $^1H-^{15}N$ NOE experiment was carried out to expose possible differences in backbone amide $^{15}N-^1H$ vector correlation times (Kay et al., 1989). This experiment is part of a complete dynamic analysis to be published at a later time. The NOE values, plotted in Figure 4, are seen to vary along the polypeptide chain. For segments 4−40 and 70−87, they are approximately 0.8, as expected if the $^{15}N-^1H$ vectors tumble at the same rate as the protein as a whole. As is often the case, there is a gradual decrease in correlation time as the ends of the polypeptide chain are approached. This acceleration of internal motion explains the weak or absent ^1H NOEs for residues in the N- and C-termini.

Figure 4 also illustrates that from 43 to 49, the $^1H-^{15}N$ NOEs decrease below the slow motion value, ranging between 0.60 and 0.35; at residue 55, a drop to nearly zero is observed. According to these NOEs, the most mobile segment outside of the termini is located from 52 to 62. A gradual increase back to 0.72 is observed from 62 to 76. Thus, a large backbone segment in the middle of the protein displays accelerated internal motions. This segment encompasses α3, β5, α4, and α5, with a maximum at β5, *i.e.*, it extends over the heme-binding site except for α2 and the back wall provided by the rest of the sheet. A comparison of amide NOEs for the two heme ligands (His-39 and His-63) is revealing: His-39 has an NOE of 0.6, whereas His-63 has a value of 0.2. These two residues are clearly distinct in their dynamic properties, and the $^1H-^{15}N$ NOEs support the view that multiple conformations indeed coexist in solution for a large part of the empty heme-binding site.

A favorable feature of a protein capable of heme binding is the hydrophobicity of the heme cavity. To test this property, ANS was added to a solution of the apoprotein. At neutral pH and low salt, no fluorescence enhancement was detected up to a 50-fold excess of ANS. In addition, apocyt b_5 has no apparent affinity for the iron-free heme group, protoporphyrin-IX (M. Cocco, personal communication). The failure to bind tightly these molecules indicates that the heme pocket is not particularly hydrophobic in the apoprotein state.

DISCUSSION

Apocytochrome b_5 behaves as a typical, small water-soluble globular protein in several respects. Its thermal unfolding is cooperative and satisfactorily modeled with a two-state transition characterized by measurable heat capacity change and denaturational enthalpy (Pfeil, 1993).[2] Apocyt b_5 contains detectable secondary structure and a well-organized core, formed mostly by hydrophobic residues upon which N- and C-segments dock. Several residues of this core are not accessible to HyTEMPO (Moore & Lecomte, 1993), but the rapid rate of backbone amide H-exchange implies that locally or globally unfolded conformations are often adopted over the minute time scale of the experiment. The core arises from elements of a rudimentary β-barrel complemented by α1 and part of α6. With α2 and part of α5, this ensemble constitutes a stabilizing and organizing module for the rest of the protein. The remainder of the polypeptide chain (α3, β5, α4, and the start of α5) defines a heme binding site lid, and although it is poorly constrained, there are few but sufficient interproton NOEs to confine it in the general vicinity of its holoprotein location.

The liganding histidines have different chemical properties, possibly due to different local amino acid sequences (Moore et al., 1991). It is now seen that the two imidazole groups are not structurally or dynamically equivalent. Although on the average they both point toward solvent rather than inward as they do in the heme-bound form, His-39 is attached to a less flexible backbone segment than His-63. Thus, the heme-binding site is selectively distorted from its heme-containing conformation. This might be the case in apomyoglobin as well, where the distal side of the heme-binding site is formed but the proximal side appears to experience wider conformational fluctuations (Lecomte et al., 1995). In apocytochrome b_{562} also, distortion seems mostly localized to one side of the heme pocket (Feng et al., 1994).

The parallel with the oxygen storage protein myoglobin is imperfect. Myoglobin possesses a deep hydrophobic binding site for the heme. The apoprotein binds the metal-free PP-IX (Breslow & Koehler, 1965; Lecomte & Cocco, 1990) and ANS (Stryer, 1965; Cocco & Lecomte, 1994) in the heme cavity, and both molecules induce shifts in the conformational distribution. In contrast, apocyt b_5 has little or no affinity for PP-IX and ANS. Among the hydrophobic

[2] These data are for rabbit liver apocyt b_5, which has a stability of 7 kJ/mol at 298 K, pH 7.4, or 18 kJ/mol lower than the holoprotein. The rat liver proteins have a similar stability and melting temperatures (M. Mayer and J. T. J. Lecomte, unpublished results).

residues of core 1, Val-23 (in β4a) and Leu-32 (in β3a) occupy a well-defined location and the nearby Leu-25 (in β4a) and Phe-35 (in α2a) are somewhat constrained. It is possible that these four residues are not sufficient for the binding of hydrophobic molecules. This lack of affinity demonstrates that in designing bishistidyl *b* hemoproteins, the poising of hydrophobic side chains in the hemeless protein and the consequent risk of heme pocket collapse or intermolecular association can be reduced without necessarily compromising heme binding.

Apocyt b_5 can be viewed as constructed of two closely associated modules, one providing a stable well-defined core and minimal elements for heme binding, and the other, a partly folded, fluctuating ensemble of residues that make up the larger part of the heme pocket. It can be speculated that a favorable trait for the heme-binding site is to present a dynamic facade and another more structurally frozen facade to facilitate proper heme docking. This arrangement has two advantages for binding: it lowers the entropic cost by providing limited conformational range established prior to heme attachment, and it lowers the enthalpic cost by minimizing the number of strong protein−protein interactions to be broken for heme insertion. The exposure of the liganding histidines to solvent also ensures ready kinetic access to the binding site.

The architecture of apocyt b_5 suggests an efficient approach for prosthetic group binding protein design. Once a suitable starting two-module framework is established, iterations can concentrate separately on the stable secondary elements of the organizing module and the polypeptide segments with weak intrinsic folding propensities of the binding module. The partial fold option chosen by nature could be the key to avoiding unwanted aggregation, intramolecular collapse, low heme affinity, and various kinetic impasses.

Recent MD simulations of beef liver and rat liver apocyt b_5 raise the possibility that the dynamic properties of the binding module are sensitive to a small number of mutations (E. M. Storch and V. Daggett, personal communication). Our own experimental studies of mutants of α6 (in the organizing module) indicate that the stability of the apoprotein can be altered without impairing the affinity for the heme (M. Mayer and J. T. J. Lecomte, unpublished results). Although apocyt b_5 behaves as a single thermodynamic unit, these observations suggest that a certain level of structural independence between the two modules is possible. It is expected that the theoretical and experimental studies of cyt b_5 variants currently in progress will refine the design guidelines for modular construction and further our understanding of the relationship between structure, dynamics, and thermodynamics in partially folded proteins.

ACKNOWLEDGMENT

The authors thank Dr. Melanie J. Cocco for performing the PP-IX binding trials and Dr. Valerie Daggett for kindly sharing her MD results. We are grateful for the welcome critical comments of Drs. C. Robert Matthews and George D. Rose.

SUPPORTING INFORMATION AVAILABLE

List of ^{1}H and ^{15}N chemical shifts at pH 6.2 and 298 K (2 pages). Ordering information is given on any current masthead page.

REFERENCES

Beck von Bodman, A., Schuler, M. A., Jollie, D. R., & Sligar, S. G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9443−9447.
Billeter, M., Neri, D., Otting, G., Qian, Y. Q., & Wüthrich, K. (1992) *J. Biomol. NMR 2*, 257−274.
Bonvin, A. M. J., Boelens, R., & Kaptein, R. (1994) *J. Biomol. NMR 4*, 143−149.
Breslow, E., & Koehler, R. (1965) *J. Biol. Chem. 240*, PC2266−PC2268.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
Brünger, A. T. (1992) *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR*, Yale University Press, New Haven.
Cavanagh, J., & Rance, M. (1992) *J. Magn. Reson. 96*, 670−678.
Cocco, M. J., & Lecomte, J. T. J. (1990) *Biochemistry 29*, 11067−11072.
Cocco, M. J., & Lecomte, J. T. J. (1994) *Protein Sci. 3*, 267−281.
Drobny, G., Pines, A., Sinton, S., Weitekamp, D. P., & Wemmer, D. (1979) *Symp. Faraday Soc. 13*, 49−55.
Falzone, C. J., Kao, Y.-H., Zhao, J., Bryant, D. A., & Lecomte, J. T. J. (1994a) *Biochemistry 94*, 6052−6062.
Falzone, C. J., Kao, Y.-H., Zhao, J., MacLaughlin, K., Bryant, D. A., & Lecomte, J. T. J. (1994b) *Biochemistry 94*, 6043−6051.
Feng, Y., Sligar, S. G., & Wand, A. J. (1994) *Nat. Struct. Biol. 1*, 30−35.
Gros, P., van Gunsteren, W. F., & Hol., W. G. J. (1990) *Science 249*, 1149−1152.
Guiles, R. D., Basus, V. J., Sarma, S., Malpure, S., Fox, K. M., Kuntz, I. D., & Waskell, L. (1993) *Biochemistry 32*, 8329−8340.
Harper, E. T., & Rose, G. D. (1993) *Biochemistry 32*, 7605−7609.
Huntley, T. E., & Strittmatter, P. (1972) *J. Biol. Chem. 247*, 4641−4647.
Kay, L. E., Torchia, D. A., & Bax, A. (1989) *Biochemistry 28*, 8972−8979.
Lecomte, J. T. J., & Cocco, M. J. (1990) *Biochemistry 29*, 11057−11067.
Lecomte, J. T. J., & Moore, C. D. (1991) *J. Am. Chem. Soc. 113*, 9663−9665.
Lecomte, J. T. J., Kao, Y.-H., & Cocco, M. J. (1996) *Proteins: Struct., Funct., Genet.* (in press).
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989) *J. Magn. Reson. 85*, 393−399.
Mathews, F. S., Czerwinski, E. W., & Argos, P. (1979) *The Porphyrins*, Vol. 7, pp. 107−147, Academic Press, New York.
Moore, C. D., & Lecomte, J. T. J. (1990) *Biochemistry 29*, 1984−1989.
Moore, C. D., & Lecomte, J. T. J. (1993) *Biochemistry 32*, 199−207.
Moore, C. D., Al-Misky, O. N., & Lecomte, J. T. J. (1991) *Biochemistry 30*, 8357−8365.
Neri, D., Otting, G., & Wüthrich, K. (1990) *J. Am. Chem. Soc. 112*, 3663−3665.
Pfeil, W. (1993) *Protein Sci. 2*, 1497−1501.
Piotto, M., Saudek, V., & Sklenár, V. (1992) *J. Biomol. NMR 2*, 661−665.
Presta, L. G., & Rose, G. D. (1988) *Science 240*, 1632−1641.
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. (1994) *Nature 368*, 425−432.
Sklenár, V., Piotto, M., Leppik, R., & Saudek, V. (1993) *J. Magn. Reson., Ser. A 102*, 241−245.
Stryer, L. (1965) *J. Mol. Biol. 13*, 482−495.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
Teale, F. W. J. (1959) *Biochim. Biophys. Acta 35*, 543.
Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. L., & Matthews, C. R. (1995) *Biochemistry 34*, 12812−12819.

BI960501Q