Solution ^{1}H NMR Study of the Heme Cavity and Folding Topology of the Abbreviated Chain 118-Residue Globin from the Cyanobacterium *Nostoc commune*†

Deok Cheon Yeh,‡ Marc V. Thorsteinsson,§,|| David R. Bevan,§ Malcolm Potts,§ and Gerd N. La Mar*,‡

Department of Chemistry, University of California, Davis, California 95616, and Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Received September 3, 1999; Revised Manuscript Received November 29, 1999

ABSTRACT: The globin from the cyanobacterium *Nostoc commune*, abbreviated GlbN, which appears to serve as a part of a terminal oxidase rather than as a respiratory pigment, displays relatively normal O_2 binding properties, despite the highly abbreviated polypeptide chain, (118 residues) relative to more conventional globins [Thorsteinsson, M. V., Bevan, D. R., Potts, M., Dou, Y., Eich, R. F., Hargrove, M. S., Gibson, Q. H., and Olson, J. S. (1999) *Biochemistry 38*, 2117−2126]. The nature of the heme cavity and the general folding topology of this cyanoglobin were investigated by solution ^{1}H NMR to establish the extent to which, and the manner in which, this compact globin adheres to the standard globin fold. This represents by far the smallest globin subjected to structural analysis. The paramagnetic cyanomet derivative was selected because its characteristically large magnetic anisotropy imparts significant dipolar shifts which both improve resolution to greatly facilitate assignments and serve as indicators of the folding topology of the globin. Identification of the axial His 70 and highly conserved Phe 35 (CD1) determined the absolute orientation of the heme and proximal His. Sequential assignments of four helical and one loop segments, which exhibit dipolar contacts to the heme and among each other, confirm the presence of well-conserved F, G, and H helices and the FG corner. The majority of the abbreviation of the chain relative to the more conventional length globins is accommodated in the A−D helices, of which the last is completely missing. The distal residue which provides a H-bond to bound ligand is identified as Gln 43, but the expected helical position E7 could not be confirmed. His 46, placed at position E10, is found to adopt alternate orientations into, and out of, the heme cavity depending on protonation state, suggesting the presence of a Bohr effect at low pH. It is shown that the dipolar shifts exhibited by backbone protons for the assigned residues conform well to those observed for other cyanomet globins and further support a conserved Mb fold. Perturbed medium-range dipolar contacts and the pH-independent backbone proton lability of the F helix are interpreted in terms of a holoprotein which is less stable than a conventional length globin.

The photoautotrophic heterocystous cyanobacterium *Nostoc commune* UTEX 584 is capable of aerobic nitrogen fixation and synthesizes a monomeric hemoglobin (Hb)[1] with unusually high O_2 affinity under dual O_2 and N_2 deprivation (*1*). The putative role of this Hb, named cyanoglobin, also denoted GlbN, is to sequester O_2 and present it to a terminal oxidase complex (*2*). Cyanoglobin, like other, but not all, bacterial Hbs (*3*−*6*), contains fewer residues (118) than a typical monomeric globin (140−160 residues) (*1*). These compact globins with their abbreviated chains pose a challenge in establishing the sequence homology to the conventional globins, and it has been proposed that such compact globins may not adhere to the highly conserved globin fold of eight helices as typified by that of sperm whale or *Aplysia limacina* Mbs (*3*, *7*−*10*). The amino acid sequence of *Nostoc* cyanoglobin is similar to that of the protozoan Hbs and chloroplast Hb of the alga *Chlamydomonas eugametus* (*11*). A sequence alignment of *N. commune* GlbN with that of sperm whale has been proposed (*7*) and is shown in Figure 1, where the sequences of *A. limacina* and the bacterium *Paramecium caudatum* (*12*) are also included. Considerable sequence variability is seen among these globins, though the proximal ligands to the heme (His 70 in GlbN and His 93 in sperm whale Mb) are identical as is the highly conserved Phe CD1 (Phe 35 in GlbN and Phe 43 in sperm whale Mb). The likely distal residue which provides a hydrogen bond to the bound ligand is Gln 43 in GlbN, but His 64 in sperm whale Mb. A crystal structure of this GlbN is not yet available, but functional studies reveal an O_2 on-rate of ∼390 μM^{-1} s^{-1} which is more than twice as fast as

† This research was supported by grants from the National Institutes of Health (HL 16087 to G.N.L.) and the National Science Foundation (IBN953157 to M.P.).

* To whom correspondence should be addressed: Department of Chemistry, University of California, Davis, CA 95616. Phone: (916) 752-0958. Fax: (916) 752-8995. E-mail: lamar@indigo.ucdavis.edu.

‡ University of California.

§ Virginia Polytechnic Institute and State University.

|| Present address: Department of Bacteriology, University of Wisconsin, Madison, WI 53706.

[1] Abbreviations: COSY, two-dimensional bond correlation spectroscopy; DSS, 2,2′-dimethyl-2-silapentane-5-sulfonate; Hb, hemoglobin; Mb, myoglobin; MCOSY, magnitude COSY; metMbCN, cyanometmyoglobin; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser spectroscopy; TOCSY, two-dimensional total correlation spectroscopy; *Nc*, *N. commune*; GlbN, cyanoglobin; metGlbNCN, cyanometcyanoglobin.

10.1021/bi992081l CCC: $19.00
Published on Web 01/20/2000

FIGURE 1: Amino acid sequence alignments of *N. commune* (*Nc*) cyanoglobins with those of sperm whale (SW), *A. limacina* (*Al*), and *Paramecium caudatum* (*12*) (*Pc*). The previously reported alignment for *Nc* is labeled with a dagger, and that determined herein is labeled with an asterisk (*7*). The helical positions are given on the top line; the helices are identified on the second line from the top, and sequence numbers for *N. commune* cyanoglobin are given on the bottom line. Residues assigned in the study presented here, for which a position in the Mb fold is provided, are bold.

the next fastest of *Glycera* Hb (*13*), a "normal" O_2 off-rate, very effective discrimination for O_2 against CO, and a very loosely held heme that dissociates $\sim 10^2$ times faster than in sperm whale Mb (*14*, *15*).

Solution ^{1}H NMR analysis of globins is capable of providing the sequence identity and spatial disposition of residues in contact with the heme and establishing whether these residue are located on helices or loops (*16*); under favorable circumstances, the complete solution structure is attainable (*17*, *18*). The procedure is severely complicated by problems in spectral resolution, so that effective studies of a globin require ^{15}N and ^{13}C labeling (*17*, *18*). The paramagnetic, low-spin ferric and cyanomet forms of globins, however, exhibit large hyperfine shifts that significantly improve the resolution for residues in the active site, and hence greatly facilitate the assignment and location of residues relative to the heme (*19*−*33*). The ligated His and heme exhibit both dipolar and contact shift, for which the dominant latter contribution provides information about the orientation of the axial His relative to the heme (*34*−*36*). The characteristic large magnetic anisotropy of cyanomet globins, moreover, imparts large dipolar shifts to nonligated residues given by

$$\delta_{dip} = \frac{1}{12\pi}\left[\Delta\chi_{ax}(3\cos^2\theta - 1)R^{-3} + \frac{3}{2}\Delta\chi_{rh}(\sin^2\theta\cos 2\Omega)R^{-3}\right] \quad (1)$$

where $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic anisotropy of the diagonal paramagnetic susceptibility tensor, χ, respectively, and R, θ, and Ω are the polar correlation of a residue proton in the magnetic coordinate system. The anisotropies reflect the iron-ligand bonding (*37*), and the orientation of the magnetic axes reflects the axial His orientation (*38*−*43*) and the tilt and/or bend of the FeCN unit (*23*, *27*−*29*, *33*, *44*). However, both the anisotropies and magnetic axes are relatively conserved among cyanomet globins (major magnetic axes tilt within $\sim$15° of the heme normal), so that the backbones of residues in the conserved Mb fold exhibit δ_{dip} values highly characteristic of their position in the conserved Mb fold.

We report herein on a solution ^{1}H NMR study of *N. commune* metGlbNCN designed to provide information about the level of identity in the sequence of the residues in contact with the heme, and the nature of the secondary structural elements on which these residues are placed. Comparisons of these heme contacts and their δ_{dip} values with those of sperm whale (*19*, *45*, *46*) and *A. limacina* Mb (*23*, *47*) are subsequently used to determine the sequence alignment with other globins. To the degree allowed by resolution and sample purity, assignments are pursued by standard sequence-specific methods generally applied to diamagnetic systems (*16*).

MATERIALS AND METHODS

Protein. Overexpression and isolation of *N. commune* strain UTEX 584 cyanoglobin were performed as described previously (*15*). The cyanide complex was prepared by adding an $\sim$10-fold excess of KCN to an $\sim$2 mM metcyanoglobin solution. The pH was adjusted with NaO^2H or ^2HCl (^2HDCl) in ^2H_2O (^2D_2O) solution and measured with an Ingold microcombination electrode and Beckman model 3550 pH meter. Low-pH solutions were buffered with acetate, and neutral to alkaline solutions were buffered with phosphate. Exchange of buffer was performed using an Amicon ultrafiltration cell.

NMR Spectra. ^{1}H NMR spectra were collected on GE Omega 500 MHz and Bruker Avance 600 MHz spectrometers, and chemical shifts, δ_{DSS}(obs), are referenced to 2,2′-dimethyl-2-silapentane-5-sulfonate (DSS) via the solvent signal. Solvent suppression, when needed, was achieved by direct saturation during the relaxation delay. Spectra were collected over the pH range 5.1−9.6 and temperature range 5−35 °C. Labile protons were monitored using the "jump and return" pulse sequence (*48*) with, and without, saturation of the solvent resonance. Strongly relaxed resonances were optimally detected using the WEFT pulse sequence (*49*). Steady-state nuclear Overhauser effect (NOE) difference spectra were obtained in a manner described in detail previously (*19*). Nonselective T_1 values were obtained from the initial linear position of the magnetization recovery using the standard inversion−recovery pulse sequence. The position of a paramagnetically relaxed (with T_{1i}) proton *i* relative to the heme iron R_{Fe-i} can be estimated from the relation

$$T_{1i}/T_{1j} = R^6_{Fe-i}/R^6_{Fe-j} \quad (2)$$

where proton *j* is a resolved reference proton with a known R_{Fe-j} (i.e., axial His F8 ring $N_\delta H$ with an R_{Fe} of ∼5 Å) with a readily measured T_1 of ∼50 ms.

NOESY spectra (*50*) consisted of 512 t_1 blocks × 2048 t_2 points over 20 kHz using 50 and 100 ms mixing times at a repetition rate of 2 s^{-1}. The intensity of several labile protons was enhanced in a 1:1 echo NOESY spectrum (*51*). CLEAN-TOCSY spectra (*52*) consisted of 512 t_1 blocks × 2048 t_2 points over 10 kHz using 20, 40, and 100 ms spin-lock times at a repetition rate of 1 s^{-1}. One-dimensional NMR data were processed on a Sun Sparcstation using GE software, and two-dimensional data were processed on a Silicon Graphics Indigo II workstation using FELIX 950 or Bruker XWIN NMR software. The two-dimensional processing consisted of apodizing with 15°- and 30°-shifted sine-bell and zero-filling to 2048 × 2048 points prior to Fourier transformation.

Dipolar Shift Estimates. The backbone $C_\alpha H$ observed dipolar shifts, δ_{dip}(obs), for metGlbNCN, were estimated (*23*, *24*) using

$$\delta_{dip}(obs) = \delta_{DSS}(obs) - \delta_{DSS}(dia) \quad (3)$$

where δ_{DSS}(dia) is the observed shift for the residue in a diamagnetic complex. On the basis of showing that the globin fold in GlbN is largely conserved relative to either sperm whale or *Aplysia* Mbs, we estimate the backbone δ_{DSS}(dip) for GlbN to be similar to the observed values for sperm whale MbCO (*53*) or the calculated δ_{DSS}(dia) values for *Aplysia* metmbCN (*23*), which are very similar to each other. The calculated dipolar shift, δ_{dip}(calc), for the $C_\delta H$ of His at position E10 in sperm whale Mb was obtained using the published magnetic axes for sperm whale metMbCN (*46*). The Thr E10 was replaced with His with the ring either oriented into the heme pocket or rotated 90° about the $\alpha-\beta$ bond to orient the ring toward the protein surface.

RESULTS

NMR Spectra. The 500 MHz ^{1}H NMR reference spectra of metGlbNCN at pH 7.0 in ^2H_2O and ^1H_2O are illustrated in Figures 2A and 3A, respectively, and the WEFT spectra that emphasize strongly relaxed resonances appear in Figures 2B and 3B, respectively. The spectra reveal the presence of at least two molecular species in an ∼4:1 ratio, each with obvious methyl (labeled M_i and m_i for major and minor isomers, respectively) and single-proton peaks (labeled H_i and h_i for major and minor isomers, respectively, with *i* being heme position or residue sequence position); labile protons are shown in italics, i.e., *H*$_i$. The ratio of isomers does not appear to change with either pH (5.1−9.6) or temperature (5−35 °C). We subsequently consider only those resonances which can be demonstrated to arise from the major isomer. The molecular nature of the minor isomer will be addressed briefly later. Comparison of Figures 2A and 3A reveals one obvious moderately relaxed labile proton peak at 16 ppm, labeled *H*$_{70\delta}$, as well as several weakly relaxed labile protons in the 8−12 ppm window. The WEFT spectrum in ^2H_2O (Figure 2B) reveals two very strongly relaxed peaks (T_1 ∼ 3 ms), labeled $H_{70\delta}$ and $H_{70\epsilon}$, and one moderately relaxed resonance (T_1 ∼ 20 ms), labeled $H_{35\zeta}$, while the same spectra in ^1H_2O (Figure 3B) exhibit two additional moderately relaxed (T_1 ∼ 15−25 ms) labile proton peaks, one each on the high-, *H*$_{43\epsilon}$, and low-field side, *H*$_{43\epsilon'}$, of the diamagnetic envelope. Additional weakly relaxed labile proton signals are detected in the 12−15 ppm window in a 1:1 pulse sequence (*48*), but not in a hard pulse spectrum (not shown; see the Supporting Information).

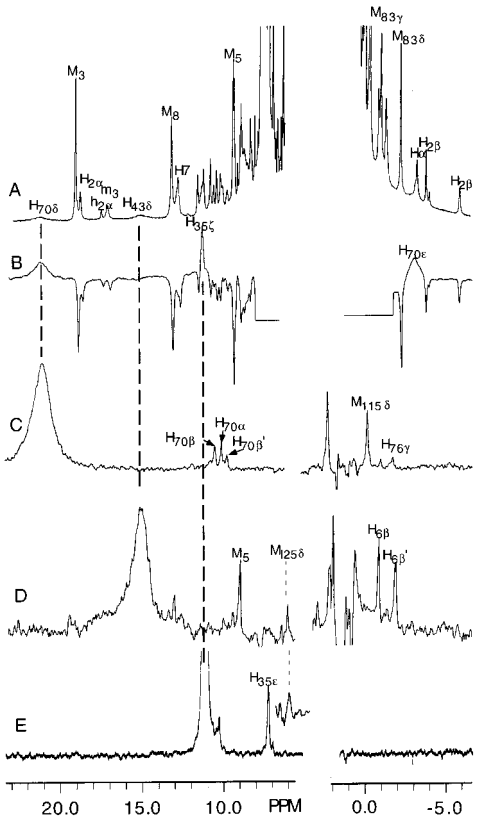

FIGURE 2: 500 MHz ^{1}H NMR (A) reference spectrum (collected at a repetition rate of 2 s^{-1}) and (B) WEFT spectrum (collected at a repetition rate of 10 s^{-1} with a relaxation delay of 20 ms) of *N. commune* metGlbNCN in ^2H_2O at pH 6.1 and 30 °C. (C−E) Steady-state NOE difference spectra upon saturation: (C) the His 70 (F8) $C_\delta H$, (D) the putative His 46 (E10) $C_\delta H$ resonance, and (E) the Phe 35 (CD1) $C_\zeta H$ signal. Resonances are labeled M_i (methyl) and H_i (single-proton) for the major isomers and m_i (methyl) and h_i (single-proton) for the minor isomers with the subscript *i* being heme position 1−8, α−*δ*, or the amino acid sequence position and proton identity.

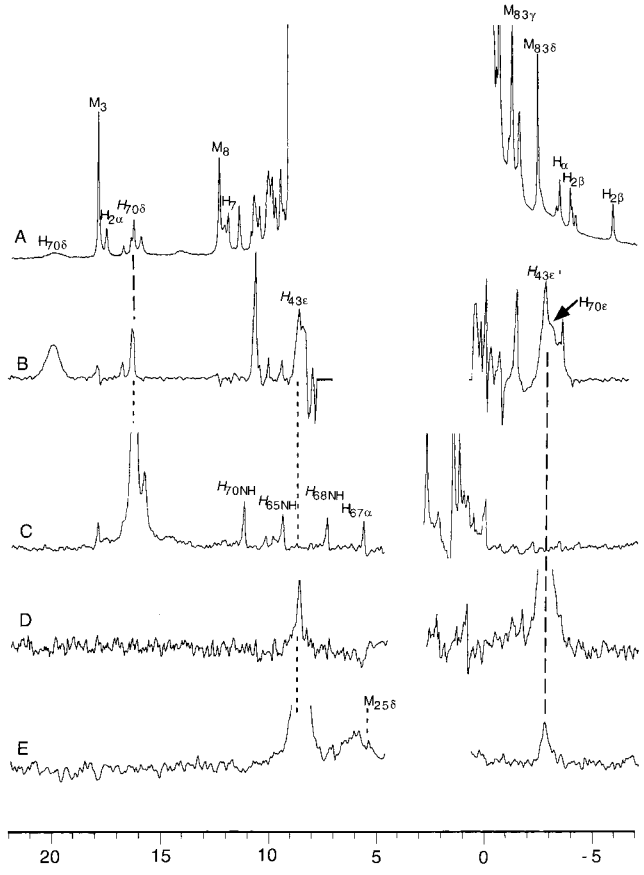

FIGURE 3: 500 MHz ^{1}H NMR (A) reference spectrum (repetition rate of 2 s^{-1}) and (B) WEFT spectrum (repetition rate of 10 s^{-1}, and relaxation delay of 20 ms) of *N. commune* metGlbNCN in 90% ^1H_2O/10% ^2H_2O at pH 5.8 and 30 °C. (C−E) Steady-state NOE difference spectra upon saturation: (C) the His 70 (F8) ring $N_\delta H$ signal $H_{70\delta}$, (D) the upfield Gln 43 (E7) $N_\epsilon H$ peak $H_{43\epsilon}$, and (E) the low-field Gln 43 (E7) $N_\epsilon H$ peak $H_{43\epsilon'}$. Resonances for the major isomers are labeled M_i (methyl) and h_i (single-proton) with the subscript *i* being heme positions 1−8, α−*δ*, or, for residues, the sequence position and proton identity (italics, i.e., *H*, identifies labile protons).

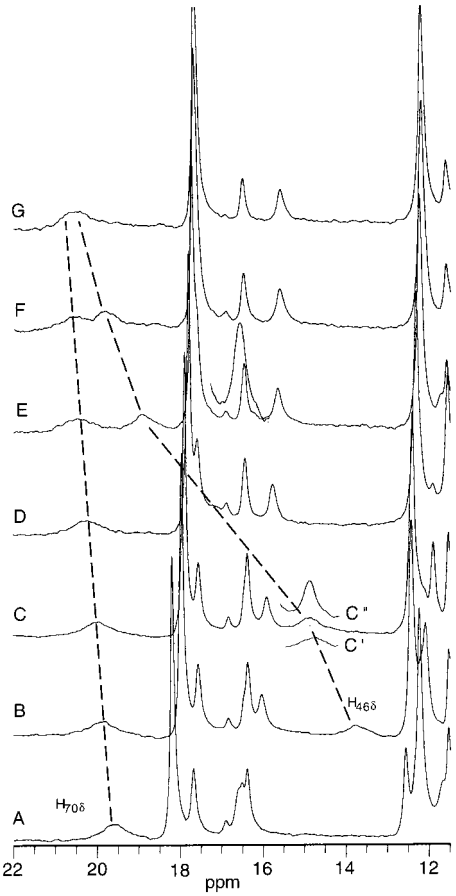

FIGURE 4: Effect of solution pH on the low-field portion of the 500 MHz ^{1}H NMR spectrum of *N. commune* metGlbNCN in 90% ^1H_2O/10% ^2H_2O at 30 °C and pH (A) 4.97, (B) 5.71, (C) 5.90, (D) 6.30, (E) 7.01, (F) 7.51, and (G) 8.61. Insets C′ (at 15 °C) and C″ (at 35 °C) show the effect of temperature on the peak $H_{70\delta}$ line width at a constant pH of 5.90.

The chemical shifts for the majority of the resolved resonances exhibit small to moderate (≤2 ppm) pH influences indicative of a single p*K* of ∼6 (not shown; see the Supporting Information). The chemical shifts at 30 °C for heme and axial His 70 signals in the pH-independent region (8.1−8.6) and the observed changes upon lowering the pH to 5.1 are provided in the Supporting Information. Similar chemical shift data for 34 assigned heme cavity residues are also listed in the Supporting Information. Nonselective T_1 values were determined at several pH values and were found to be pH-independent; the T_1 values for the resolved signals for nonligated residues are included in the Supporting Information. The majority of the signals exhibited temperature-dependent chemical shifts; the slope and intercepts in a Curie plot (shift vs the reciprocal of the absolute temperature) for the heme, His 70, and nonligated residues are given in the Supporting Information. One notable resonance, labeled $H_{46\delta}$ in Figures 2A and 3A, exhibits a very strong pH influence on its shift (∼10 ppm), as shown in Figure 4A−F, and excessive line width (>200 Hz) despite a relatively long T_1 of ∼100 ms. The effect of temperature on peak $H_{46\delta}$, shown in Figure 4, is to selectively narrow this peak above 25 °C (Figure 4C′) and further broaden it below 25 °C (Figure 4C″), which indicates that the signal is experiencing severe line broadening due to fast exchange on the NMR time scale (>3 × 10^4 s^{-1}) at 10 °C between alternate environments (*54*).

Heme Assignments. Completely standard ^{1}H NMR protocols (*21*, *24*, *36*) show for the major isomer two each hyperfine-shifted, characteristic three-proton vinyl, and four-proton propionate spin systems by TOCSY (not shown). NOESY connections to adjacent methyls (with no TOCSY connection) differentiate the 1-CH_3−2-vinyl and 3-CH_3−4-vinyl pairs from the 5-CH_3−6-propionate and 8-CH_3−17-propionate pairs. An NOE between a methyl from each of the two sets identifies the 1-CH_3−2-vinyl and 8-CH_3−7-propionate signals and, hence, assigns all pyrrole substituents. NOESY cross-peaks from adjacent pyrrole substituents to moderately relaxed protons (*21*) identify the four meso Hs (with the α-meso H resolved with a T_1 of ∼50 ms).

The minor isomer exhibits one well-resolved, obvious methyl peak (peak m_3 in Figures 2A and 3A) that exhibits a NOESY cross-peak to a resolved single proton h_2, identified as the H_α of a vinyl group in TOCSY spectra (not shown). A minor component CH_3−propionate-CH_2 pair is similarly located in the 8−10 ppm window (not shown). The remaining heme resonances for the minor isomer could not be located due to spectral congestion. However, inspection of the NOESY slices reveals that the cross-peak patterns are

essentially identical for the major 3-CH_3 and minor vinyl H_α resonances and for the major $2H_\alpha$ and minor methyl resonances (not shown; see the Supporting Information), which clearly establishes that the minor isomer has the heme rotated by 180° about the $\alpha-\gamma$ meso axis with respect to the major isomer, as has been established for numerous other globins (*24*, *36*, *55*, *56*).

Proximal Side. The two strongly relaxed ($T_1 \sim 3$ ms), nonlabile protons labeled $H_{70\delta}$ and $H_{70\epsilon}$ in Figures 2B and 3B are diagnostic for the ring CHs of the axial His in all low-spin ferric hemoproteins (*19*, *34*−*36*). TOCSY spectra exhibit a low-field, hyperfine-shifted AMX spin system which is coupled to its $N_\text{P}H$ (not shown; see the Supporting Information). Saturating the extreme low-field relaxed proton, $H_{70\delta}$, in ^2H_2O (Figure 2C) yields strong NOEs to protons of this AMX spin system, while saturating the low-field labile proton peak $H_{70\delta}$ (with a T_1 of ∼50 ms) in ^1H_2O (Figure 3C) yields NOEs to both the AMX spin system and its labile proton, uniquely identifying $C_\delta H$, $N_\delta H$, and the $NHC_\alpha HC_\beta H_2$ fragment of the axial His. The remaining upfield, relaxed proton with a T_1 of ∼3 ms, $H_{70\epsilon}$, can be confidently assigned to the remaining $C_\epsilon H$ of axial His 70 and is confirmed by an expected NOE to the His 70 $N_\delta H$ (not shown).

Distal Side. Residues with protons in proximity to the iron ($T_1 < 50$ ms, $R_\text{Fe} \leq 5$ Å), which do not exhibit detectable NOEs to the proximal His, most likely reside on the distal side. The definitive location on the distal side for such residues is established by the detection of a network of NOEs among numerous residues, none of which are in contact with the proximal His (see below). A moderately relaxed ($T_1 \sim 28$ ms, $R_\text{Fe} \sim 4.2$ Å), nonlabile proton at ∼11 ppm, $H_{35\zeta}$, fails to exhibit NOEs to the axial His, and TOCSY spectra in ^2H_2O (not shown; see the Supporting Information) connect $H_{35\zeta}$ to two additional signals in the aromatic window. The TOCSY connections, variable-temperature behavior, and T_1 for the resolved single proton are essentially the same as those observed for Phe CD1 in each cyanomet globin previously characterized by two-dimensional ^1H NMR (*21*, *24*, *57*), and can be confidently assigned as such here. Saturating the Phe CD1 $C_\zeta H$, as shown in Figure 2E, results in an NOE to its $C_\epsilon H$, as well as to a weakly relaxed proton signal at 5.3 ppm (labeled $M_{25\delta}$) whose shift is extraordinarily temperature dependent (see Table 2S in the Supporting Information). The two strongly relaxed, labile proton peaks, labeled $H_{43\epsilon}$ and $H_{43\epsilon'}$, located at either edge of the diamagnetic envelope in the ^1H_2O WEFT trace in Figure 3B, when saturated (Figure 3D,E), exhibit strong NOEs to each other but not to the axial His, dictating their origin as a geminal NH_2 group of a distal Asn or Gln. The previous sequence alignment proposed (see Figure 1) Gln 43 was at position E7, and in the absence of a plausible Asn or other Gln, it is so assigned. This NH_2 group exhibits hyperfine shifts and relaxation properties very similar to those of the E7 Gln $N_\epsilon H_2$ in numerous cyanomet globins with Gln E7 (*26*, *27*, *30*, *31*, *33*).

Orientation of the Heme. The assignment of the heme, axial His, and Phe CD1 allows the determination of the absolute orientation of the heme in the normal Mb fold, which is necessary to allow comparison of heme contacts in metGlbNCN with those of structurally characterized globins. The Phe CD1 ring in all globins makes dipolar contact with either pyrroles C/D or pyrroles D/B, depending on the heme

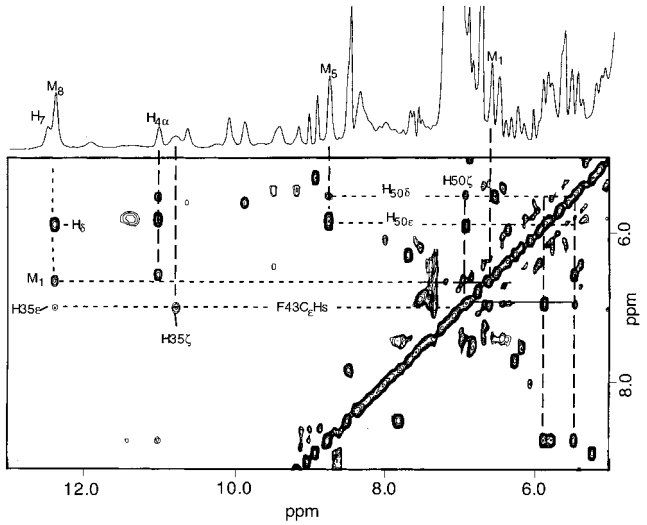

FIGURE 5: Portion of the 500 MHz NOESY spectrum (mixing time of 30 ms) of *N. commune* metGlbNCN in ^2H_2O at pH 5.8 and 30 °C illustrating the dipolar contacts between the Phe 35 (CD1) ring and the heme 1-CH_3 and 8-CH_3 and between the ring of Phe 50 (E14) and the heme 5-CH_3 that determine the absolute orientation of the heme in the heme pocket. The two-dimensional data set was apodized by a 30°-shifted sine-bell window function and zero-filled to 2K × 2K points prior to Fourier transformation.

orientation (*21*, *23*, *27*, *28*, *33*, *45*, *47*, *55*, *56*, *58*−*60*). The Phe CD1 identified above exhibits NOESY cross-peaks to 8-CH_3 and 1-CH_3 (Figure 5), as expected and observed in *Chironomus* metHbIIICN (*29*, *56*) and *A. limacina* metMbCN (*23*), in contrast to the expected and observed Phe CD1 contacts to 5-CH_3 and 4-vinyl in mammalian (*19*, *27*, *57*) and *Lucina pectinata* metHbICN (*33*). Hence, the heme orientation, as viewed from the proximal side, is that depicted in Figure 6.

Sequence-Specific Assignments on Helical Fragments. Characteristic dipolar connectivities among subsets of the TOCSY-detected $N_\delta HC_\alpha H$ fragments of 25 residues (not shown) lead to the location of four sequential fragments of nine, five, eight, and four residues, which the available side chain TOCSY data characterize as Z_j-AMX_{j+1}-Z_{j+2}-Thr_{j+3}-AMX_{j+4}-Ala_{j+5}-G_{j+6}-Z_{j+7}-AMX_{j+8} (**I**, Z being long side chains; see the Supporting Information), AMX_k-AMX_{k+1}-Ala_{k+2}-Ile_{k+3}-Ala_{k+4} (**II**), Ala_l-Ala_{l+1}-Leu_{l+2}-X_{l+3}-Z_{l+4}-Val_{l+5}-Thr_{l+6}-AMX_{l+7} (**III**, X being an unidentified residue; not necessary to identify it since α_i-N_{i+3} and $\alpha_i-\beta_{i+3}$ cross-peaks connect the two segments), and Ala_i-AMX_{i+1}-Leu_{i+2}-Ala_{i+3} (**IV**). Sequential fragments **II**−**IV**, as well as the sequence up to Ala_{j+5} for fragment **I**, exhibit the N_i-N_{i+1}, α_i-N_{i+1}, $\alpha_i-\beta_{i+3}$, and α_i-N_{i+3} (except **I**; see Discussion) NOESY cross-peak (*16*) and $C_\alpha H$ chemical shift (*61*) patterns shown schematically in Figure 7 that are characteristic of α-helices. The sequence unequivocally shows fragment **I** arises from Met 66−Asn 74, where AMX_{j+4} is the backbone of the axial His identified above, and confirms the predicted axial ligand role for His 70. The first five residues of fragment **I** represent the F4−F9 helical positions if we assume axial His occupies the conventional F8 position (*62*); the last three residues of segment **I** likely constitute part of the FG corner (see below). Dipolar contacts between F helix residues Met 66−Ala 31 and the heme were not observed and [except for the Met 66 (F4) side chain terminus] not expected in the Mb fold; the Met 66 (F4) terminus which

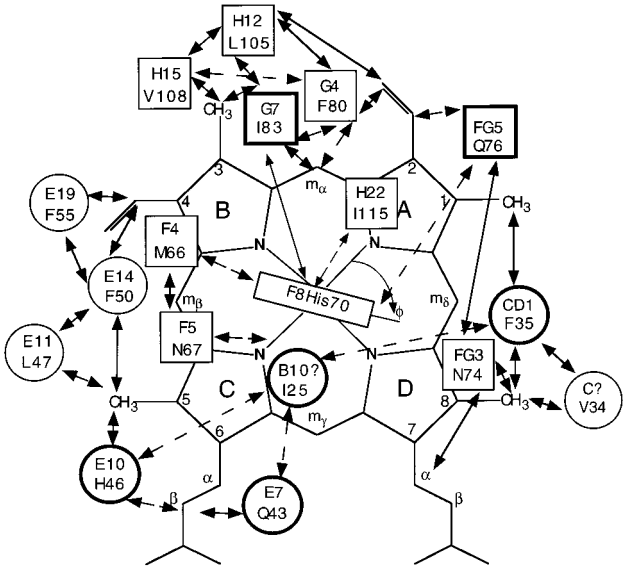

FIGURE 6: Schematic diagram of the heme (labeled in Fisher notation) and heme pocket of *N. commune* GlbN as determined herein with NMR data. The view is from the proximal side, with proximal residues shown as squares and rectangles and distal side residues shown as circles. The residues are identified by both sequence number and the helical or loop position established by comparison to the Mb fold. Solid and dashed double-headed arrows illustrate the dipolar contacts observed by NOESY cross-peaks and steady-state NOE, respectively. The nonligated residues outlined in bold exhibit significant dipolar shifts and paramagnetic relaxation. The orientation of the His 70 (F8) imidazole plane relative to the pyrrole A−pyrrole C vector is defined by the angle ϕ.

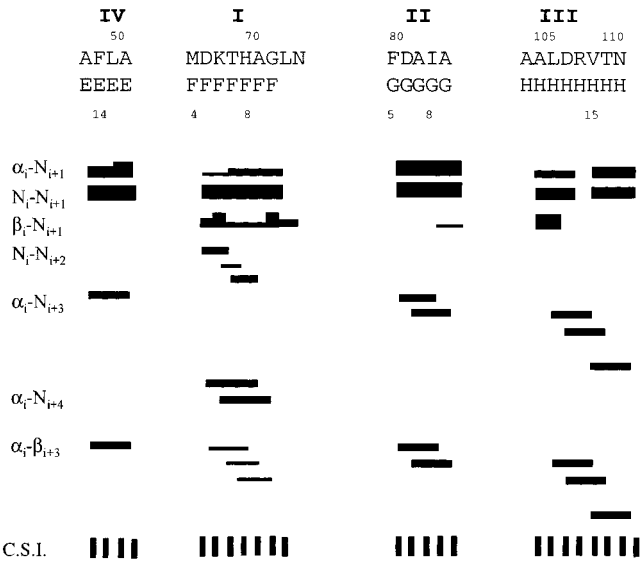

FIGURE 7: Schematic representation of the pattern and intensity (wide line = strong, medium line = moderate, and narrow line = weak) NOESY cross-peaks that provide sequence-specific assignments and establish α-helical structure (*16*) for four segments of the polypeptide chain for *N. commune* metGlbNCN determined in ^1H_2O at pH 5.8 and 30 °C. The sequence positions of the four fragments are given on the top line, and the helix position is given on the second line. A residue whose $C_\alpha H$ shift satisfies the criterion for α-helical positions (*61*) is denoted by a vertical bar at the bottom.

should make a heme contact could not be assigned due to spectral congestion.

Fragment **II** is uniquely identified in the sequence as Phe 80−Ala 85, which should correspond to a part of the G helix (Figure 1); NOESY cross-peaks from a TOCSY-detected (see the Supporting Information) aromatic ring to AMX_k (not shown) confirm Phe 80. Both the Phe 80 ring and Ile 83 side chain protons exhibit significant dipolar shifts (see the Supporting Information), with the resolved Ile 83 $C_\delta H_3$ ($T_1 \sim 100$ ms) and $C_\gamma H_3$ ($T_1 \sim 70$ ms) ~5.8 and ~5.5 Å from the iron, respectively. The Phe 80 ring exhibits NOESY cross-peaks to 2-vinyl and α-meso-H, while the Ile 83 side chain exhibits NOESY cross-peaks to the heme 3-CH_3 and α-meso-H, as well as axial His 70 (F8) (as shown schematically in Figure 6, which shows both of these residues on the proximal side of the heme). The G helix generally provides two strong contacts with the heme for residues G5 and G8 as in sperm whale (*19*, *45*) or *Aplysia* Mb (*23*, *47*) or G4 and G7 as in *Scapharca* Hb (*28*, *59*).

Helical fragment **III** corresponds to residues Ala 103−Asp 110 in the sequence, with the Leu 105 and Val 108 side chains exhibiting weak dipolar shifts (see the Supporting Information) and minimal paramagnetic relaxation ($T_1 \geq 200$ ms, $R_{Fe} \geq 7$ Å). This portion of the sequence near the C-terminus corresponds to the H helix (Figure 1). The NOESY cross-peaks from Leu 105 to the 2-vinyl, and from Val 108 to the 3-CH_3, as well as the dipolar contacts of both of these residues with the proximally located Ile 83 (G7 or G8) or Phe 80 (G4/G5) (as depicted in Figure 6) confirm the expected proximal placement of the H helix. The H helix generally provides two to three important contacts to the heme and axial His (*45*, *58*). Residues H22 (in *Chironomus* Hb) (*29*, *58*) or H23 (in sperm whale) (*19*, *45*) make strong contacts with the axial His, while residue H15 makes contacts with pyrroles A/B. In *Aplysia* Mb (*47*), an additional heme contact exists for residue H12 (while the H22 side chain is too short to contact the heme). The observed contacts for Leu 105 and Val 108 satisfy the contacts for positions H12 and H15 in *Aplysia* Mb (*47*).

The available sequence assigns the last fragment **IV** to Ala 49−Ala 52, and the NOESY connection of a TOCSY-detected (see the Supporting Information), weakly hyperfine-shifted aromatic ring to AMX_{i+1} confirms Phe 50. The other three residues exhibit negligible hyperfine shifts and paramagnetic relaxation. The alignments agree on the placement of these residues somewhere on the E helix. The NOESY cross-peak between the ring of Phe 50 and the 4-vinyl group (see Figure 6) is that expected for the E14 position in the Mb fold, which is consistent with the sequence predicted on the basis of a Gln at position E7, and confirms Gln43 at position E7 (see Figure 1). Sequential assignments toward the N-terminus from helical fragment **IV** could not be attained due to either spectral congestion or peptide proton lability. Moreover, spectral congestion precluded locating the E helix backbone for Gln 43 (E7) starting from the relaxed $N_\epsilon H_2$ fragment identified above. The observed residue−heme and inter-residue dipolar contacts are depicted schematically in Figure 6.

Heme Contacts at Interhelical Loops. The strongly relaxed and hyperfine-shifted Phe ring in contact with the heme 1-CH_3 and 8-CH_3 has been assigned above to Phe CD1, but this did not identify which Phe in the sequence. NOESY cross-peaks between the Phe CD1 to a complete TOCSY-detected Val side chain are those expected between the CD1 and C7 residues in sperm whale or *Aplysia* Mb, and lead to the assignment of Phe 35 as CD1, with Val 34 possibly completing the C helix. It was not possible to locate the

signals for the other residue adjacent to Val 34 to verify the presence of a C helix. The Asn 74 identified above as the likely third nonhelical residue (AMX_{j+8}) at the end of fragment **I** exhibits NOESY cross-peaks to moderately relaxed ($T_1 \sim 100$ ms, $R_{Fe} \sim 5.8$) C_γHs of a dipolar-shifted, TOCSY-detected AMPTX spin system (not shown). Both Asn 74 and the AMPTX spin systems are in dipolar contact with the proximal His 70 (F8) (shown schematically in Figure 6), and exhibit the expected contacts to pyrrole A and pyrrole D of residues FG3 and FG5 of sperm whale Mb (or FG2 and FG4 of *Aplysia* Mb) (see Figure 6) and are thus assigned as Asn 74 (FG3) and Gln 76 (FG5).

Other Active Site Residues. A six-spin, TOCSY-detected, methyl-containing residue with only very weak upfield dipolar shifts exhibits NOESY cross-peaks to the heme 5-CH_3 and to Phe 50 (placed at position E14 above) (not shown). Failure to locate the backbones of adjacent residues precludes sequence-specific assignment and establishing whether the residue is located on a helix. However, the likely Leu exhibits some (but not all; see Discussion*)* properties of the E11 side chain in the Mb fold, and hence is tentatively assigned as Leu 47. A weakly dipolar-shifted, TOCSY-detected aromatic ring (see the Supporting Information) exhibits NOESY cross-peaks to the 4-vinyl and the identified Phe 50 (E14) (not shown). Despite the inability to assign its backbone, a reasonable assignment to Phe55 can be made. The analogous residue in *Chironomus* HbIII or *Aplysia* Mb (*23*, *29*, *47*, *58*) occupies position E19, and makes similar contacts to residue E14 and the heme. However, it was not possible to confirm a helical position for this residue in GlbN.

The B helix generally provides a key residue at position B10 that lines the distal pocket (*27*, *45*, *47*, *58*, *60*, *62*). Unfortunately, no additional helical fragments could be resolved. A strongly hyperfine-shifted, TOCSY-detected, six-spin system (not shown) with a resolved methyl peak at 5.3 ppm (labeled $M_{25\delta}$ in Figure 2B) can arise from only Leu or Ile. Comparably strong NOESY and TOCSY cross-peaks from $M_{25\delta}$ to a pair of strongly coupled upfield protons (not shown) show the resolved methyl arises from an ethyl, rather than an isopropyl, fragment, and hence identify the residue as Ile rather than Leu. Weak NOEs to this Ile from Phe 35 (CD1) C_ζH (Figure 2D) and Gln 43 (E7) N_ϵH (Figure 3E) place it in the distal cavity, but not in contact with the heme. A similarly strongly low-field dipolar-shifted Leu (B10) exhibits these dipolar contacts in a variety of globins (*19*, *23*, *26*, *27*, *29*, *33*), and hence, this Ile is the most likely candidate for the B helix residue that corresponds to the Leu (B10). No adjacent residues could be identified, but Ile 25 is the only reasonable candidate.

The NOE to an upfield, partially resolved, methyl signal labeled $M_{115\delta}$, observed upon saturating the proximal His 70 C_δH (Figure 2C), is part of a TOCSY-detected, six-spin system (not shown) with an obvious weakly dipolar-shifted methyl partially resolved on the upfield edge of the diamagnetic envelope. Hence, this proximal residue must be a Leu or an Ile. Spectral congestion prevented both sequence-specific assignment and establishing a helical position for this residue. The proximal location so close to the axial His for a Leu or an Ile is consistent only with residue H22 [*Chironomus* Hb (*29*, *58*)] or H23 [sperm whale Mb (*19*, *45*)] in the Mb fold. Leu 115 in the sequence accounts for this residue (see Discussion).

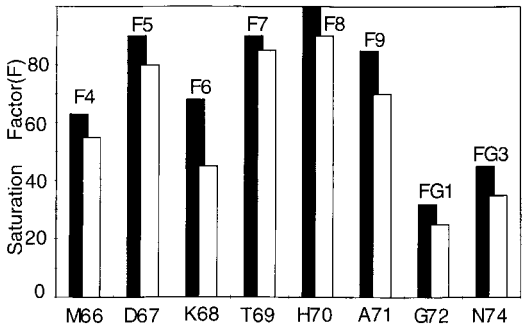

FIGURE 8: Bar diagram of the saturation factor, F_i, defined in eq 5, for the peptide backbone for F helix residues Met 66 (F)−Ala 71 (F9) and FG corner residues Gly 72 (FG1) and Asn 74 (FG3) at pH 5.8 (black bar) and pH 8.6 (white bar) in 90% ^1H_2O/10% ^2H_2O at 30 °C.

Last, the broadened and moderately relaxed proton whose chemical shift changes by more than 10 ppm in the pH titration (labeled $H_{43\delta}$ in Figures 2A and 4) is too broad even at the highest practical temperature (Figure 4C″) to characterize in detail. However, saturating this peak in ^2H_2O leads to NOEs only to 5-CH_3, 6-C_βHs, and the methyl of our proposed B helix Ile 25, as shown in Figure 2D, confirming it as a distal residue. The NOE pattern to the heme is that expected for the residue at position E10 in the Mb fold, which should be occupied by His 46 in this protein on the basis of the sequence assignment of Gln 43 (E7) and Phe 50 (E14) above (see Figure 1). Hence, a tentative assignment of this residue is made to one of the ring protons (most likely the C_δH) of His 46 (E10). The large variation in shift (∼10 ppm; see Figure 4) with pH strongly suggests that the residue is changing orientation with pH, and it is noted that the NOEs to 5-CH_3 and 6-H_β become more intense as the pH is raised into the alkaline region (not shown). Unfortunately, the excessive line width of this peak and the failure to resolve the alternate environments preclude a more definitive description of the alternate environments for this residue.

Labile Proton Exchange. Comparison of the peptide NH signal *i* intensity in the 1:1 spectra (*48*) in ^1H_2O with (I_i) and without (I_i^o) saturation of the solvent signal (not shown; see the Supporting Information) shows that all resolved peptide NH signals exhibit saturation transfer due to exchange with solvent, with the rate k_i given by (*54*, *63*−*65*)

$$F_i = I_i/I_i^o = \rho_{1i}/(\rho_{1i} + k_i) \qquad (4)$$

where ρ_{1i} is the intrinsic relaxation rate of proton *i* in the absence of exchange. The saturation factors, F_i, vary from 0.3 to 0.9 for F helix and FG turn residues, as shown in Figure 8; the saturation factor *F* for these resolved NH is independent of pH in the range 5.3−9.5 within uncertainties in *F* (±10%).

DISCUSSION

Heme Electronic and Molecular Structure. The heme methyl T_1 values of ∼150 ms are similar to those of comparably contact-shifted methyls in other cyanometglobins (*34*−*36*) and support a very similar electronic structure. The dominant contact shift pattern of pyrrole substituents is controlled, in large part, by the orientation of the axial His imidazole plane defined by ϕ in Figure 6 (*39*−*42*). Comparison of the observed heme methyl shift pattern (see the

Supporting Information) with predictions (*42*) of the relative contact shifts as a function of ϕ indicates $\phi \sim -60 \pm 10°$, an orientation similar to that found in *Ascaris suum* Hb (*66*). The plots of the heme methyl shifts versus the reciprocal of the temperature deviate from Curie behavior in a fashion that is similar to that in other cyanometglobins, with the deviation that can be attributed to thermal population of the excited orbital state which has the spin density pattern rotated by 90° from that of the ground state (*41*, *43*, *46*, *67*). The axial His ring hyperfine shift pattern, which reflects comparable contact and dipolar contributions (*46*), is similar to that in *L. pectinata* metHbICN (*33*), which exhibits a major magnetic axis minimally tilted from the heme normal, or the model compound with imidazole and cyanide as ligated (*68*), where the major magnetic axis is very likely normal to the heme. Hence, a major magnetic axis orientation close to the normal to the heme can be anticipated for cyanoglobin. The pattern of NOEs from its axial His ring $N_\delta H$ to the F helix backbone (not shown) is consistent with an orientation of the F helix similar to that found in other cyanometglobins characterized by ^{1}H NMR (*23*, *32*, *69*).

Folding Topology and Sequence Alignment. This study confirms His 70 as the axial ligand, Gln 43 as the distal E7 residue, and Phe 35 at the highly conserved CD1 position. The placement of the axial His as next to last on the F helix and the location of Asn 74 and Gln 76 at nonhelical positions characteristic of the FG corner, together with the observation of a helical fragment that exhibits the characteristics (*62*) of the G helix, argue for a conserved F helix and FG corner and at least beginning of the G helix placement in cyanoglobin relative to the Mb fold. With Gln 76 at position FG5, the two G helix contacts to the heme are located at G4 (Phe 80) and G7 (Ile 83), like in *Scapharca* HbI (*28*, *59*), rather than at G5 and G8 as in mammalian (*19*, *45*) and *Aplysia* Mbs (*23*, *47*). However, in sperm whale Mb (*45*) [and *Scapharca* HbI (*59*)], the G helix is approximately normal to the heme and the two major heme contact residues at G5 (G4) and G8 (G7) are clearly on the proximal and distal sides, respectively. In *Aplysia* Mb, the G axis is tilted with respect to the heme with the G5 residue clearly proximal, but the G8 residue is now peripheral rather than distal to the heme (*47*). In metGlbNCN, even Ile 83 (G7) *is clearly proximal*, since it exhibits a NOESY cross-peak to His 70 (F8) (see Figure 1). Hence, the G helix is translated toward the proximal side in GlbN relative to that in sperm whale or *Aplysia* Mb.

The weak hyperfine shift and relaxation (see the Supporting Information), as well as the heme contacts of H helix residues Leu 105 and Val 108 (Figure 6), are completely consistent with those of residues at positions H12 and H15 in *Aplysia* Mb. The unique contact in metGlbNCN of an Ile or a Leu with proximal His 70 (F8) is characteristic of helical position H23 in sperm whale (Tyr) or H22 in *Chironomus* (Met), but the analogous position H22 in *Aplysia* Mb has a side chain that is too short to make contact with His F8. Hence, this Ile or Leu must arise from Ile 115 at a position analogous to helical position H22. The heme contact for Leu 105 (H12) reflects an H-like helix in *Aplysia* Mb, which is translated slightly toward its C-terminus with respect to that in sperm whale Mb. The alignment in Figure 1 thus supports the presence of "normal" G and H helices with positions and orientations closer to those in *Aplysia* than those in sperm whale Mb. The apparent translations of helix G in GlbN toward its N-terminus compared to either *Aplysia* or sperm whale Mb noted above are confirmed in the altered "register" of G−H interhelical contacts. Thus, side chain interactions occur between residues G5 and H15, and G8 and H12, in sperm whale Mb that are shifted to G4 (Phe 80) with both H12 (Leu 105) and H15 (Val 108), and G7 (Ile 83) with H12 (Leu 105), in cyanoglobin (see Figure 6).

It was possible to obtain evidence for the E helix only for a short fragment (Ala 49−Ala 52), for which Phe 50 exhibited the heme contact at position E14. While this alignment is supported by the detection of a Gln side chain close to the ligated cyanide that is usually occupied by residue E7, the backbone for the Gln (E7) residue could not be located and direct evidence for the continuation of the helix below Ala 49 could not be obtained. Moreover, Phe 55 exhibits the heme contact generally observed for position E18 (with distal residue E7), but which cyanoglobin would place at position E19 which is not oriented toward the heme in either sperm whale or *Aplysia* Mb. Hence, the possibility of a broken E helix near the N-terminus for achieving these usual properties cannot be discounted. The E helix, moreover, must be truncated at its N-terminus, as restricted by the identification of Phe 35 at position CD1, and Gln 43 at position E7 (Figure 1), as proposed for *Paramecium* Hb (*12*).

Spectral congestion prevented identification of any helical fragment that could correspond to the A, B, or C helix, although the identification of Phe 35 at position CD1 suggests a nearly "complete" C-helix initiated by Ser 29 (Figure 1). The D helix is clearly absent and may account, in part, for the ready loss of heme by GlbN (*15*). The occupation by an Ile of the position generally found for the B10 residue in the Mb fold suggests the alignment of A and B helices as shown in Figure 1. However, direct confirmation that those residues reside on helices is absent.

Backbone Dipolar Shifts. An indirect but convincing argument for a largely conserved Mb fold in the heme environment in cyanoglobin is provided by the $C_\alpha H$ dipolar shifts exhibited by the various assigned residues, which are given by eq 1. The magnetic axes and resultant dipolar shifts (eq 1) have been determined for several structurally characterized metMbCNs (*21*, *28*, *29*, *44*), and in each case, a very similar pattern of δ_{dip} *for backbone protons* is observed which correlates very well with the conserved helical and loop positions. The predicted $\delta_{dip}(C_\alpha H)$ as a function of sequence number for sperm whale metMbCN is reproduced in Figure 9A. The positions of the helices and interhelical loops are shown at the bottom. Panels B−E of Figure 9 plot the $\delta_{dip}(C_\alpha H)$ observed in cyanoglobin and those of sperm whale (*46*) and *Aplysia* Mb (*23*) for the residue located on homologous helical or loop positions. The very good correlation [except for Ile 83 (G7)] supports both the conserved disposition of the residues relative to the heme (geometric factors in eq 1) and the conserved electronic and magnetic properties of the heme iron (magnetic axes and anisotropies). The deviation of the Ile 105 (G7) $C_\alpha H$ may reflect changes in the G helix noted above.

E Helix and pH Influences on Structure. The relaxation properties of the Gln 43 (E7) $N_\epsilon Hs$ place one 4.5 Å from the iron, and the $N_\epsilon H_2$ δ_{dip} values are very similar to those for elephant metMbCN (*26*, *27*) and *L. pectinata* metHbICN (*33*). Hence, Gln 43 very likely contributes a hydrogen bond

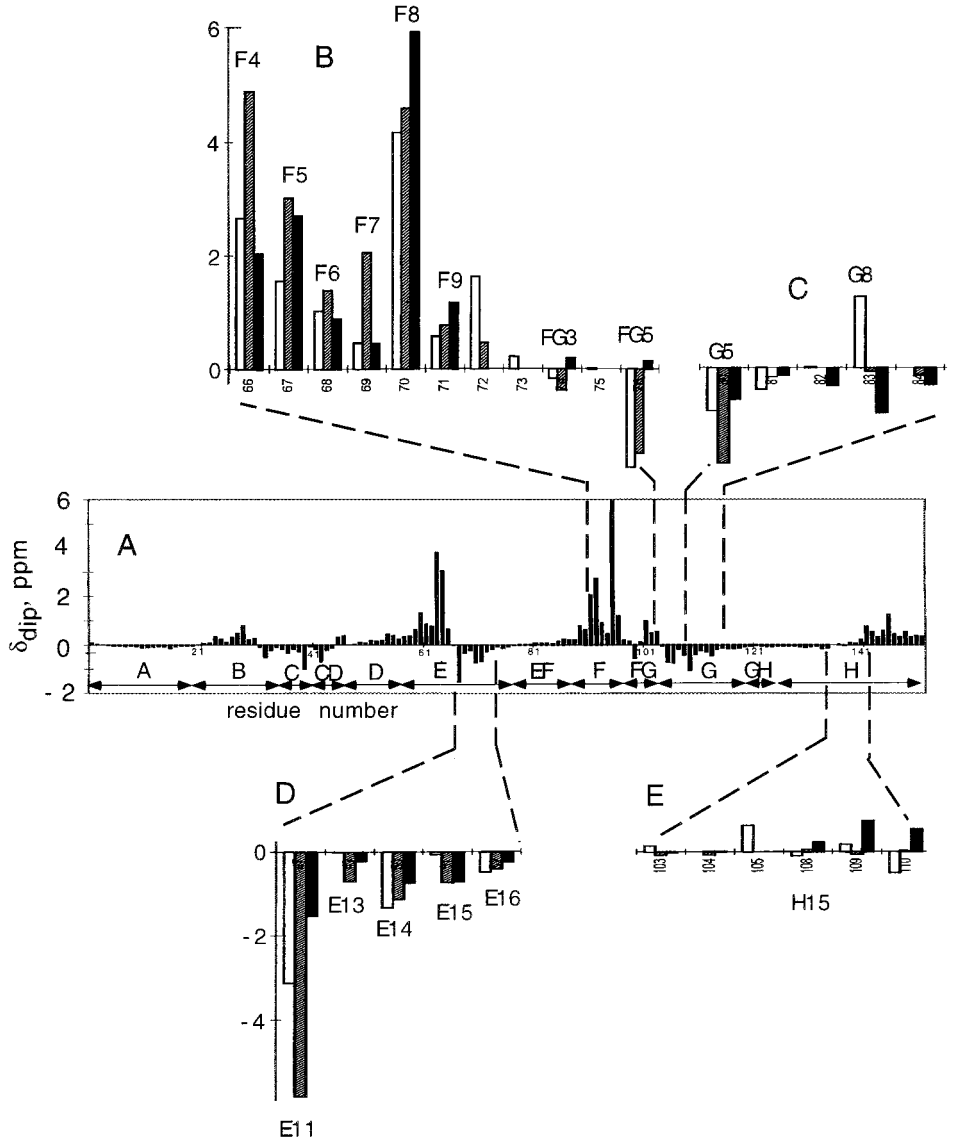

FIGURE 9: (A) Predicted backbone $C_\alpha H$ dipolar shifts for sperm whale metMbCN as a function of sequence number using the published magnetic axes and anisotropies (*46*) and the MbCO crystal coordinates (*45*); the helical or interhelical loop positions are also shown at the bottom. Expanded portions of the helical section where spatial dipolar shifts are observed are shown for (B) the F helix and FG corner, (C) a portion of the G helix, (D) the position of the E helix, and (E) the portion of the H helix where the $C_\alpha H$ δ_{dip}(obs) values are compared for sperm whale metMbCN (*46*) (gray bars), *A. limacina* metMbCN (*23*) (white bars), and *N. commune* metGlbNCN (black bars). Note the similarities of the $C_\alpha H$ δ_{dip} values for the homologous residues in the three proteins.

to the bound cyanide in metGlbNCN. The residues expected to reside on the E helix exhibit anomalous properties that include the absence of the characteristic strong dipolar contact between the backbone of the Leu 47 (E11) and the pyrrole B−C junction, the observation of a heme contact with Phe 55 that the alignment places at position E19 but is expected by residue E18, and the apparent strong pH effects on the ring resonance of the proposed His 46 which would occupy position E10.

The large (∼10 ppm) chemical shift change with pH, but relatively little change in the T_1 of ∼100 ms with pH for $H_{46\delta}$, argues for alternate orientations of the side chain into, and out of, the heme pocket depending on its protonation states. Placement of a His at position E10 in *Aplysia* (*23*) or sperm whale Mb (*46*), with magnetic axes whose major axis is close to the heme normal, predicts an ∼10 ppm *low-field* δ_{dip} when a His at position E10 is oriented into the heme pocket, and only an ∼2 ppm δ_{dip} when oriented toward the surface, with the ring CHs that are ∼6 Å from the iron that do not dramatically change with rotation about χ_1. Hence, the pH effect on shift and the large line width for the proposed His 46 (E10) ring CH are consistent with a dynamic equilibrium[2] for the side chain orientation between a neutral side chain oriented into the pocket at alkaline pH and a protonated side chain that is expelled from the pocket at acidic pH. The observed p*K* of ∼6 is thought to arise from a His side chain (*16*). The excessive line width arises from an exchange rate (*54*) with a half-life of <30 μs. In accord with this proposal, the steady-state NOEs from the proposed His 46 (E10) $C_\delta H$ to the heme 5-CH_3 and $6H_\alpha$ appear

[2] Resonance $H_{46\delta}$, when resolved, exhibits single-proton intensity within experimental error. However, irradiation of $H_{46\delta}$ also leads to saturation transfer to another position under the aromatic envelope (not shown). These results suggest yet a third, very minor, orientation for His 46 (E10) that is in "slow exchange" with the single $H_{46\delta}$. In ^2H_2O, this "third" species signal appears to exhibit an intensity much lower than that of a single proton at the position where magnetization transfer is observed.

stronger at alkaline than at acidic pH (not shown). The functional role of the titration of the His is not known (*15*) but suggests the possibility of a Bohr effect.

The very similar heme cavity molecular and electronic structure of cyanomet *Aplysia* Mb and cyanoglobin is consistent with the similar O_2 off-rates ($\sim$80 s^{-1}) for these two proteins, which are only a factor of 5 slower in sperm whale Mb (*15*). The data presented here, however, provide little information about the structural basis for the much faster O_2 on-rate in GlbN (390 mM^{-1} s^{-1}) than in either *Aplysia* or sperm whale Mbs ($\sim$15 mM^{-1} s^{-1}) (*15*). The NMR data presented here identify a pH-modulated structural transition with a p*K* of $\sim$6 that likely involves alternate orientations of the distal His 46 (E10) side chain. It is not known, however, whether the transition alters the O_2 binding properties.

Structure and Dynamic Stability of the F Helix. While the majority of the characteristic (*16*) α-helical NOESY cross-peaks (α_i-N_{i+1} and $\alpha_i-\beta_{i+3}$) are observed for the F helix fragment of residues Met 66 (F4)−Ala 71 (F9), the conventional α_i-N_{i+3} cross-peaks are not detected, while α_i-N_{i+4} cross-peaks are observed, as shown in Figure 7. The α_i-N_{i+3} and α_i-N_{i+4} cross-peaks (weaker and stronger than expected for an α-helix, respectively) are indicative of a helix with a slightly lower pitch than an α-helix. The reason for the "looser" helix is unknown. However, a possible consequence for this perturbed helix is also manifested in its dynamic behavior (see below).

The exchange of "buried" labile protons, due to transient unfolding, with opening and closing rates k_{op} and k_{cl}, respectively (and hence equilibrium constraints for unfolding $K_{op} = k_{op}/k_{cl}$), and base-catalyzed exchange for the "exposed" group ($k_{ex} = k_2[OH^-]$), leads to an observed labile proton exchange rate, k_{obs}, which, for primarily folded proteins ($k_{cl} \gg k_{op}$), is given by (*70*)

$$k_{obs} = k_{op}k_{ex}/(k_{cl} + k_{ex}) \quad (5)$$

One of two simple, limiting cases for stable proteins, with $k_{cl} \gg k_{ex}$ (EX2), simplifies eq 5 to

$$k_{obs} = (k_{op}/k_{cl})k_{ex} = K_{op}k_{ex} \quad (6)$$

where k_{obs} represents base-catalyzed exchange. Both the F helix backbone and His F8 ring NH in cyanometglobins have been reported to exhibit base-catalyzed exchange indicative of the EX2 mechanism (*63*−*65*). The cyanomet complex of GlbN, however, exhibits essentially pH-independent exchange for both F helix and FG corner peptide NHs (see Figure 8), such as that predicted for the case where $k_{cl} \ll k_{ex}$ (EX1), for which eq 5 reduces to

$$k_{obs} = k_{op} \quad (7)$$

The conversion from the EX2 to the EX1 exchange mechanism for a conserved structural element within a given protein folding topology could arise if the protein stability (particularly that of the F helix) is decreased significantly, i.e., k_{cl} increased dramatically. A reported notable property of cyanoglobin, compared to more conventional globins, is the ease with which it loses heme (*15*). Moreover, we have noted a pattern of NOEs that suggest a more flexible F helix in GlbN than in conventional length globins. Since the F helix is the last helix formed on generating holoMb from apoMb and heme (*71*), the accelerated heme loss and the pH-independent observed exchange rate are likely related (*72*).

CONCLUSIONS

Solution ^1H NMR of the cyanomet derivative of *N. commune* cyanoglobin provides the spectral resolution via the characteristic dipolar shifts arising from magnetic anisotropy that allows sequence-specific assignment of numerous helical and interhelical fragments near the heme. The combination of the characteristic heme contacts and the direction and magnitude of the dipolar shifts for residues on helical and interhelical loop fragments allows the determination of the sequence alignment for a large portion of the protein and the demonstration that the folding topology of the compact globin adheres well to the conventional globin fold.

SUPPORTING INFORMATION AVAILABLE

Four figures showing the effect of temperature on 1:1 ^1H NMR spectra of labile protons, NOESY slices of 3-CH_3 and $2H_\alpha$ for major and minor heme orientations, a plot of δ_{DSS}-(obs) versus pH, and NOESY−TOCSY spectra for fragment **I** and aromatic ring assignments and two tables of spectral parameters for the heme ligands and for nonligated residues. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Potts, M., Angeloni, S. V., Ebel, R. E., and Bassam, D. (1992) *Science 256*, 1690−1692.
2. Hill, D. R., Belbin, T. J., Thorsteinsson, M. V., Bassam, D., Brass, S., Ernst, A., Boger, P., Paerl, H., Mulligan, M., and Potts, M. (1996) *J. Bacteriol. 178*, 6587−6598.
3. Webster, D. A. (1988) in *Advances in Inorganic Biochemistry* (Eichhorn, G. L., and Marzilli, L. G., Eds.) pp 246−262, Elsevier, Amsterdam.
4. Vasudeuvan, S. G., Armarego, W. L. F., Shaw, D. C., Lilley, P. E., Dixon, N. E., and Poole, R. K. (1991) *Mol. Gen. Genet. 226*, 49−58.
5. Cramm, R., Siddiqui, R. A., and Fredrich, B. (1994) *J. Biol. Chem. 269*, 7349−7354.
6. Lacelle, M. M., Kumaro, M., Kurita, K., Yamane, K., Zuber, P., and Nakano, M. M. (1996) *J. Bacteriol. 178*, 3803−3808.
7. Vinogradov, S. N., Walz, D. A., Pohajdak, B., Moens, L., Kapp, O. H., Suzuki, T., and Trotman, C. N. A. (1993) *Comput. Biochem. Physiol. 106B*, 1−26.
8. Kapp, O. H., Moens, L., Vanfleteren, J., Troman, C., Suzuki, T., and Vinogradov, S. (1995) *Protein Sci. 4*, 2179−2190.
9. Moens, L., Vanfleteren, J., Van de Peer, Y., Peeters, K., Kapp, O., Czelusniak, J., Goodman, M., Blaxter, M., and Vinogradov, S. (1996) *Mol. Biochem. Evol. 13*, 324−333.
10. Suzuki, T., and Imai, K. (1998) *Cell. Mol. Life Sci. 54*, 979−1004.
11. Tsubamoto, Y., Matsuoka, A., Yusa, K., and Shikama, K. (1990) *Eur. J. Biochem. 193*, 55−59.
12. Iwaasa, H., Takagi, T., and Shikama, K. (1989) *J. Mol. Biol. 208*, 355−358.
13. Parkhurst, L. J., Sima, P., and Goss, D. J. (1980) *Biochemistry 19*, 2688−2692.
14. Thorsteinsson, M. V., Bevan, D. R., Ebel, R. E., Weber, E., and Potts, M. (1996) *Biochim. Biophys. Acta 1292*, 133−139.
15. Thorsteinsson, M. V., Bevan, D. R., Potts, M., Dou, Y., Eich, R. F., Hargrove, M. S., Gibson, Q. H., and Olson, J. S. (1999) *Biochemistry 38*, 2117−2126.
16. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley-Interscience, New York.

17. Osapay, K., Theriaut, Y., Wright, P. E., and Case, D. A. (1994) *J. Mol. Biol. 244*, 183−197.
18. Volkman, B. F., Alam, S. L., Saterlee, J. D., and Markeley, J. L. (1998) *Biochemistry 37*, 10906−10917.
19. Emerson, S. D., and La Mar, G. N. (1990) *Biochemistry 29*, 1545−1556.
20. Yamamoto, Y., Iwafune, K., Nanai, N., Osawa, A., Chujo, R., and Suzuki, T. (1991) *Eur. J. Biochem. 198*, 299−306.
21. Qin, J., and La Mar, G. N. (1992) *J. Biol. NMR 2*, 597−618.
22. La Mar, G. N., and De Ropp, J. S. (1993) in *Biological Magnetic Resonance* (Berliner, L. J., and Reuber, J., Eds.) pp 1−78, Plenum Press, New York.
23. Qin, J., La Mar, G. N., Ascoli, F., and Brunori, M. (1993) *J. Mol. Biol. 231*, 1009−1023.
24. Qin, J., La Mar, G. N., Cutruzzolá, F., Travaglini Allocatelli, C., Brancaccio, A., and Brunori, M. (1993) *Biophys. J. 65*, 2178−2190.
25. Yamamoto, Y., and Suzuki, T. (1993) *Biochim. Biophys. Acta 1163*, 287−296.
26. Vyas, K., Rajarathnam, K., Yu, L. P., Emerson, S. D., and La Mar, G. N. (1993) *J. Biol. Chem. 268*, 14826−14835.
27. Zhao, X., Vyas, K., Nguyen, B. D., Rajarathnam, K., La Mar, G. N., Li, T. S., Piontek, K., Erich, R. F., Olson, J. S., Ling, J., and Bocian, D. F. (1995) *J. Biol. Chem. 270*, 20763−20774.
28. Wu, Y., Basti, M., Gambacurta, A., Chiancone, E., Ascoli, F., and La Mar, G. N. (1996) *Biochim. Biophys. Acta 1298*, 261−275.
29. Zhang, W., La Mar, G. N., and Gersonde, K. (1996) *Eur. J. Biochem. 237*, 841−853.
30. Yamamoto, Y., and Suzuki, T. (1996) *Biochim. Biophys. Acta 1293*, 129−139.
31. Zhang, W., Cutruzzolá, F., Allocatelli, C. T., Brunori, M., and La Mar, G. N. (1997) *Biophys. J. 73*, 1019−1030.
32. Wu, Y., Chien, E. Y. T., Sligar, S. G., and La Mar, G. N. (1998) *Biochemistry 37*, 6979−6990.
33. Nguyen, B. D., Zhao, X., Vyas, K., La Mar, G. N., Lile, R. A., Bruker, E. A., Phillips, G. N., Jr., Olson, J. S., and Wittenberg, J. B. (1998) *J. Biol. Chem. 273*, 9517−9526.
34. Bertini, I., Turano, P., and Villa, A. J. (1993) *Chem. Rev. 93*, 2833−2932.
35. Yamamoto, Y. (1998) *Annu. Rep. NMR Spectrosc. 36*, 1−77.
36. La Mar, G. N., Satterlee, J. D., and de Ropp, J. S. (2000) in *The Porphyrin Handbook* (Kadish, K., Smith, K., and Guilard, R., Eds.) pp 185−298, Academic Press, San Diego.
37. Horrocks, W. D., Jr., and Greenberg, E. S. (1973) *Biochim. Biophys. Acta 322*, 38−44.
38. Shulman, R. G., Glarum, S. H., and Karplus, M. (1971) *J. Mol. Biol. 57*, 93−115.
39. Traylor, T. G., and Berzinis, A. P. (1980) *J. Am. Chem. Soc. 102*, 2844−2846.
40. Yamamoto, Y., Nanai, N., Chujo, R., and Suzuki, T. (1990) *FEBS Lett. 264*, 113−116.
41. Turner, D. L. (1995) *Eur. J. Biochem. 227*, 829−837.
42. Shokhirev, N. V., and Walker, F. A. (1998) *JBIC, J. Biol. Inorg. Chem. 3*, 581−594.
43. Banci, L., Bertini, I., Luchinat, C., Pierattelli, R., Shokhirev, N. V., and Walker, F. A. (1998) *J. Am. Chem. Soc. 120*, 8472−8479.
44. Emerson, S. D., and La Mar, G. N. (1990) *Biochemistry 29*, 1556−1566.
45. Kuriyan, J., Wilz, S., Karplus, M., and Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133−154.
46. Nguyen, B. D., Xia, Z. C., Yeh, D. C., Vyas, K., Deaguero, H., and La Mar, G. N. (1999) *J. Am. Chem. Soc. 121*, 208−217.
47. Bolognesi, M., Coda, A., Frigerio, F., Gatii, G., Ascenzi, P., and Brunori, M. (1990) *J. Mol. Biol. 213*, 621−625.
48. Plateau, P., and Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.
49. Gupta, P. K. (1976) *J. Magn. Reson. 24*, 461−465.
50. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
51. Sklenar, V., and Bax, A. (1987) *J. Magn. Reson. 74*, 469−471.
52. Griesinger, C., Otting, G., Wutrich, K., and Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 355−360.
53. Theriault, Y., Pochapsky, T. C., Dalvit, C., Chiu, M. L., Sligar, S. G., and Wright, P. E. (1994) *J. Biomol. NMR 4*, 491−504.
54. Sandström, J. (1982) *Dynamic NMR Spectroscopy*, Academic Press, New York.
55. Lecomte, J. T. J., Johnson, R. D., and La Mar, G. N. (1985) *Biochim. Biophys. Acta 829*, 268−274.
56. Peyton, D. H., La Mar, G. N., and Gersonde, K. (1988) *Biochim. Biophys. Acta 954*, 82−94.
57. Emerson, S. D., Lecomte, J. T. J., and La Mar, G. N. (1988) *J. Am. Chem. Soc. 110*, 4176−4182.
58. Steigemann, W., and Weber, E. (1979) *J. Mol. Biol. 127*, 309−338.
59. Royer, W. E., Heard, K. S., Harrington, D. J., and Chiancone, E. (1995) *J. Mol. Biol. 253*, 168−186.
60. Rizzi, M., Wittenberg, J. G., Coda, A., Fasano, M., Ascenzi, P., and Bolognesi, M. (1994) *J. Mol. Biol. 244*, 86−89.
61. Wishart, D. S., Sykes, B. D., and Richard, F. M. (1991) *J. Mol. Biol. 222*, 311−333.
62. Bashford, D., Chothia, C., and Lesk, A. M. (1987) *J. Mol. Biol. 196*, 199−216.
63. Cutnell, J. D., La Mar, G. N., and Kong, S. B. (1981) *J. Am. Chem. Soc. 103*, 3567−3572.
64. Lecomte, T. J. L., and La Mar, G. N. (1985) *Biochemistry 24*, 7388−7395.
65. Lambright, D. G., Balasubramanian, S., and Boxer, S. G. (1989) *J. Mol. Biol. 207*, 289−299.
66. Yang, J., Kloek, A. P., Goldberg, D. E., and Mathews, F. S. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 4224−4228.
67. Shokhirev, N. V., and Walker, F. A. (1995) *J. Phys. Chem. 99*, 17795−17804.
68. Chacko, V. P., and La Mar, G. N. (1982) *J. Am. Chem. Soc. 104*, 7002−7007.
69. Nguyen, B. D., Xia, Z., Cutruzzolá, F., Travaglini Allocatelli, C., Brunori, M., and La Mar, G. N. (2000) *J. Biol. Chem.*, in press.
70. Englander, S. W., and Mayne, L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 243−265.
71. Eilezer, D., and Wright, P. E. (1996) *J. Mol. Biol. 263*, 531−538.
72. Hargrove, M. S., Wilkinson, A. J., and Olson, J. S. (1996) *Biochemistry 35*, 11300−11309.

BI992081L