Biochemistry **1998,** *37,* 10906−10919

Solution Structure and Backbone Dynamics of Component IV *Glycera dibranchiata* Monomeric Hemoglobin−CO[†,‡]

Brian F. Volkman,[§] Steve L. Alam,[||,⊥] James D. Satterlee,[||] and John L. Markley*[,§]

National Magnetic Resonance Facility at Madison, Department of Biochemistry, University of Wisconsin−Madison, 420 Henry Mall, Madison, Wisconsin 53706, and Department of Chemistry, Washington State University, Pullman, Washington 99164-4630

Received April 10, 1998; Revised Manuscript Received May 27, 1998

ABSTRACT: The solution structure and backbone dynamics of the recombinant, ferrous CO-ligated form of component IV monomeric hemoglobin from *Glycera dibranchiata* (GMH4CO) have been characterized by NMR spectroscopy. Distance geometry and simulated annealing calculations utilizing a total of 2550 distance and torsion angle constraints yielded an ensemble of 29 structures with an overall average backbone rmsd of 0.48 Å from the average structure. Differences between the solution structure and a related crystal structure are confined to regions of lower precision in either the NMR or X-ray structure, or in regions where the amino acid sequences differ. ^{15}N relaxation measurements at 76.0 and 60.8 MHz were analyzed with an extended model-free approach, and revealed low-frequency motions in the vicinity of the heme, concentrated in the F helix. Amide proton protection factors were obtained from H−D amide exchange measurements on ^{15}N-labeled protein. Patterns in the backbone dynamics and protection factors were shown to correlate with regions of heterogeneity and disorder in the ensemble of NMR structures and with large crystallographic *B*-factors in the X-ray structures. Surprisingly, while the backbone atoms of the F helix have higher rmsds and larger measures of dynamics on the microsecond to millisecond time scale than the other helices, amide protection factors for residues in the F helix were observed to be similar to those of the other helices. This contrasts with H−D amide exchange measurements on sperm whale myoglobin which indicated low protection for the F helix (S. N. Loh and B. F. Volkman, unpublished results). These results for GMH4 suggest a model in which the F helix undergoes collective motions as a relatively rigid hydrogen-bonded unit, possibly pivoting about a central position near residue Val^{87}.

The globin family of oxygen-binding heme proteins has been the focus of structure−function analysis by X-ray crystallography, molecular dynamics, and NMR spectroscopy for nearly 40 years (*1*−*3*). Interest in these proteins remains high, partly because a precise understanding of the mechanism of ligand binding to the heme iron is lacking. Because the heme pocket is effectively inaccessible to solvent, ligand entry must involve some amount of protein structural rearrangement. Time-resolved X-ray crystallography methods have recently been used in an attempt to directly observe the internal motions associated with ligand association−dissociation in sperm whale myoglobin (Mb)[1] (*4*). Recently developed NMR methods for measuring orientation-dependent nuclear dipolar couplings have also been applied to Mb and implicate the H helix in slow collective motions (*5*).

Interest in *Glycera dibranchiata* monomeric hemoglobins (GMH), a subgroup of the globin superfamily, stems from their unusual ligand binding properties and amino acid substitutions at sequence positions that are highly conserved in other members of the globin family. For example, the E7 Leu found in all GMH proteins replaces the highly conserved globin distal His. Furthermore, sequencing and modeling studies (*6*) predict that the protein studied here, GMH4, has a B10 Phe, which is normally a Leu in the globin family (*7*). These hydrophobic side chains create an extremely apolar ligand-binding pocket and may contribute to the altered ligand binding properties of these proteins (*8*−*11*). NMR analysis first revealed the reversed orientation

---

[†] This work was supported by the National Institutes of Health (Grants GM47645 to J.D.S. and GM35976 to J.L.M.). Equipment in the National Magnetic Resonance Facility at Madison (NMRFAM) was purchased with funds from the University of Wisconsin, the NSF Biological Instrumentation Program (Grant DMB-8415048), the NIH Biomedical Research Technology Program (Grant RR02301), the NIH Shared Instrumentation Program (Grant RR02781), and the U.S. Department of Agriculture. Support for the NMR instrumentation at Washington State University was from NIH (Grant RR06312011) and from Battelle Pacific Northwest National Laboratory.

[‡] All measured relaxation parameters, calculated model-free parameters, backbone *J* coupling constants, H−D amide exchange rates and protection factors, X-PLOR NOE and dihedral constraint files, coordinates for the family of 29 conformers, and the refined average structure for GMH4CO have been deposited in the BioMagResBank database, under accession number 4096. Coordinates for the family of 29 conformers (1VRE) and the refined average structure for GMH4CO (1VRF) have been deposited in the Brookhaven Protein Data Bank.

* To whom correspondence should be addressed. E-mail: markley@nmrfam.wisc.edu. Phone: (608) 263-9349. Fax: (608) 262-3759.

[§] University of Wisconsin−Madison.

[||] Washington State University.

[⊥] Present address: Department of Human Genetics, University of Utah, 15N 2030 E RM 6160, Salt Lake City, UT 84112-5330.

[1] Abbreviations: GMH4, component IV *G. dibranchiata* monomeric hemoglobin; GMH4CO, CO-ligated form of component IV *G. dibranchiata* monomeric hemoglobin; Hb, hemoglobin; Mb, myoglobin; HSQC, heteronuclear single-quantum coherence; SE, sensitivity enhancement; INEPT, insensitive nuclear enhancement by polarization transfer; NOESY, nuclear Overhauser effect spectroscopy.

S0006-2960(98)00810-1 CCC: $15.00

of the heme cofactor in GMH proteins, relative to that of myoglobin (Mb) (*12*).

Recent studies have elucidated the sequences of three naturally isolated GMH proteins (GMH2, GMH3, and GMH4) and have led to the production of recombinant wild-type GMH4 (*13*, *14*). These sequence determinations, in combination with a comprehensive mass spectrometry analysis (*6*, *15*, *16*), indicate that the amino acid sequence of the published crystal structures (*17*, *18*) does not correspond to an actual GMH protein and that the crystals contained a complex mixture of many GMH proteins.

The work reported here reveals the structure of a true GMH protein, CO-ligated GMH4. The amino acid sequence of GMH4 differs at 11 positions from that of the crystal structures. We recently reported comprehensive ^{1}H, ^{15}N, and ^{13}C resonance assignments for GMH4CO and numerous heme−protein contacts obtained by means of isotope-filtered NOESY experiments (*14*). Here we continue this work by presenting the high-resolution solution structure of GMH4CO, along with an analysis of backbone dynamics from ^{15}N relaxation measurements and amide H−D exchange rates.

NMR relaxation provides a sensitive probe of internal molecular dynamics at atomic resolution (*19*). We have analyzed the ^{15}N relaxation measurements of GMH4CO within the model-free framework of Lipari and Szabo (*20*, *21*) to obtain generalized order parameters (S^2) and correlation times for internal motion (τ_e) for the N−H bond vectors of the backbone. In addition, residues that show evidence of chemical exchange broadening due to slow motions on the microsecond to millisecond time scale required fitting of an additional parameter (R_{ex}). A number of residues in the F helix of GMH4CO located on the proximal side of the heme were found to experience slow motions of this type; however, no evidence was seen for similar dynamic properties for residues comprising the distal heme pocket.

EXPERIMENTAL PROCEDURES

Sample Preparation. Previously published procedures (*14*) were used in overexpressing *G. dibranchiata* component IV monomeric hemoglobin (GMH4) in *Escherichia coli*, U-^{15}N and U-^{13}C,^{15}N labeling, purifying, and reconstituting the apoprotein with *b*-type hemin. The reduced carbon monoxide-ligated (CO) form of GMH4 was produced from pure ferric GMH4 by reduction with an excess of sodium dithionite. All experiments discussed below were performed on samples in 100 mM potassium phosphate buffer with 100 mM KCl at pH 5.0 in 90% H_2O/10% D_2O or 99.9% D_2O saturated with CO. Final protein concentrations were 3.5 and 3 mM for the ^{15}N- and ^{13}C/^{15}N-labeled samples, respectively. A sample for amide H−D exchange experiments was prepared by lyophilizing a sample of fully reduced ^{15}N-labeled GMH4CO in 90% H_2O/10% D_2O and resuspending it in 99.9% CO-saturated D_2O with 1 equiv of sodium dithionite. All NMR sample tubes were purged with CO and capped.

NMR Spectroscopy. All NMR spectra were recorded at 20 °C on Bruker DMX750, DMX600, and DMX500 spectrometers equipped with triple-resonance ^{1}H, ^{13}C, and ^{15}N probes and *Z*- or three-axis pulsed field gradient capabilities. Quadrature detection in the indirectly detected dimensions was obtained in the echo−antiecho mode (*22*) or with the States−TPPI method (*23*). Gradient pulses as a combination of *X*- and *Z*-gradients at the “magic angle” (*24*, *25*) were used for coherence selection in all three-dimensional (3D) and four-dimensional (4D) NOESY experiments to maximize water supression.

Quantitative heteronuclear *J* correlation experiments (*26*) were used to obtain stereospecific assignments of prochiral methyl and methylene groups in the protein, as well as values for ϕ, χ^1, and χ^2 dihedral angles, as described previously (*14*). A 4D ^{13}C/^{15}N HMQC−NOESY−HSQC spectrum was collected at 500.13 MHz with a mixing time of 100 ms and spectral widths and time domain sizes (where * indicates complex points) of 4000 Hz and 16* ^{13}C points in f_1, 5701.25 Hz and 64* ^{1}H points in f_2, 1666.67 Hz and 16* ^{15}N points in f_3, and 3004.81 Hz and 512* ^{1}H points in f_4. A 100 ms 3D ^{15}N NOESY−HSQC spectrum was acquired with spectral widths and time domain sizes of 1666.67 Hz and 36* ^{15}N points in f_1, 7716.05 Hz and 128* ^{1}H points in f_2, and 7716.05 Hz and 1024* ^{1}H points in f_3. A 3D ^{13}C NOESY−HSQC spectrum was acquired at 600.13 MHz with an 80 ms mixing time and spectral widths and time domain sizes of 6250 Hz and 48* ^{13}C points in f_1, 9090.90 Hz and 237* ^{1}H points in f_2, and 8333.33 Hz and 512* ^{1}H points in f_3. Since the heme *b* used during reconstitution of the holoprotein was not isotopically enriched, isotope-edited and -filtered NOESY experiments were used to identify intra-heme and heme−protein NOEs, as described previously (*14*).

A series of 22 2D ^{15}N HSQC experiments were carried out at 499.84 MHz in 99.9% D_2O over a period of 14 days for the determination of amide H−D exchange rates and protection factors. Each spectrum was acquired with 256* points and a spectral width of 1666.67 Hz in the ^{15}N dimension and 1024* points and a spectral width of 8333.33 Hz in the ^{1}H dimension. Spectra were acquired 18, 48, 78, 108, 400, 460, 520, 580, 640, 700, 820, 880, 940, 1000, 1060, 1120, 12 840, 17 429, 17 533, 18 433, and 18 520 min after dissolving the lyophilized protein in D_2O.

Relaxation parameters were measured at 750.13 MHz (^{15}N T_1 and T_2 and ^{1}H−^{15}N NOE) and 600.13 MHz (T_2 only) ^{1}H frequencies. All pulse schemes utilized gradients for sensitivity enhancement and selective pulses for water flip-back (*27*). All spectra were acquired with 16 scans per FID, 200* ^{15}N points, and 1024* ^{1}H points. Spectral widths at 750 (600) MHz were 10 000 (8333.33) Hz in the ^{1}H dimension and 2500 (2000) Hz in the ^{15}N dimension. T_2 values at 750 MHz were obtained with nine different durations of the delay *T*: 8, 16, 33, 50, 66, 99, 132, 198, and 264 ms. T_2 values at 600 MHz were obtained with *T* values of 0, 8, 16, 32, 48, 64, 96, and 128 ms. T_1 values at 750 MHz were obtained with *T* values of 10, 60, 120, 200, 400, 800, 1400, and 2200 ms. Duplicate spectra were recorded for *T* values of 10 and 200 ms (T_1), *T* values of 8 and 66 ms (T_2, 750 MHz), and *T* values of 0 and 64 ms (T_2, 600 MHz) for estimation of peak height uncertainties.

All Fourier transformation of NMR data was performed with FELIX95 (Molecular Simulations). In most cases, the indirect dimension with the least amount of digitization was extended with linear prediction by no more than 50% of the original data size. Typically, an 85°-shifted squared sine bell window function was applied to each FID prior to zero filling to the final matrix size, Fourier transformation, and phase correction. The initial values for incremented delays in multidimensional experiments were set in a manner that allowed predictable phasing in each dimension and minimized roll and offset of the baseline (*28*).

All ^{1}H dimensions were indirectly referenced to a separate 10 mM DSS (2,2-dimethyl-2-silapentane-5-sulfonate) sample at 20 °C in the NMR buffer discussed above. Indirect ^{13}C and ^{15}N dimensions were referenced indirectly to DSS as previously described (*29*). Chemical shifts and peak intensities of all NOESY spectra were tabulated with the PPFLX package (*30*), while Felix macros provided by M. Akke were used to obtain peak intensities of 2D ^{15}N HSQC spectra for relaxation and amide H−D exchange measurements.[2]

Structure Calculations and Analysis. The hybrid distance geometry−simulated annealing approach (*31*) was used for determination of the three-dimensional structure of GMH4CO from NOE and dihedral constraints with the program X-PLOR 3.843 (*32*). Calculations were performed in parallel on a Silicon Graphics (Mountain View, CA) Onyx computer equipped with 6 R10000 processors. The covalent geometry of the heme and carbon monoxy (CO) ligand and XPLOR patch residues defining the CO−heme and heme−His^{90} coordinate−covalent bonds were derived from the XPLOR parameter files parallh22x.pro and topallh22x.pro. The C−Fe and Fe−N$^{\epsilon 2}$ distances were fixed to 1.90 and 2.20 Å, respectively, and defined as colinear. Aside from the linkage between the side chain of His^{90} and the heme iron, only experimental NOE constraints (158) were used to define the position of the heme within the protein matrix. In each round of refinement, a total of 50 starting structures were generated with substructure embedding and regularized with 1000 steps of simulated annealing at 2000 K and 1000 cooling steps. The final refinement stage included 2500 cooling steps from a temperature of 2000 K.

Distance constraints were obtained by means of a series of Perl (*33*) scripts. For each NOESY cross-peak, a list of ^{1}H−^{1}H pairs was generated by comparison with the database of ^{1}H, ^{15}N, and ^{13}C GMH4CO resonance assignments[3] (*14*, *34*). This output was filtered against the structure resulting from the previous round of refinement, removing all combinations with an interatomic distance greater than a chosen threshold. Unambiguously assigned NOEs were added to the list of constraints for the next stage of calculations. Constraints for the initial round of refinement were obtained from the 4D ^{13}C−^{15}N NOESY cross-peaks, which were often unambiguously assigned by chemical shift alone, as well as from other NOEs, which were evaluated conservatively (using a 12 Å threshold) against a previously constructed model structure of GMH4CO (*6*). Backbone ϕ and side chain χ^1 and χ^2 dihedral angle constraints were obtained from quantitative heteronuclear *J* correlation experiments as described elsewhere (*14*). The Procheck/NMR (*53*) and X-PLOR software packages were used for analysis of the final family of 29 accepted structures.

Determination of ^{15}N Relaxation Parameters and the Initial Estimate of τ_m. Processed ^{15}N T_1 and T_2 and NOE data were analyzed with a suite of macros and programs available from A. Palmer.[4] Maximum ^{1}H−^{15}N peak intensities were obtained from each spectrum for all residues whose signals were resolved sufficiently. Uncertainties in peak intensitites were estimated from duplicate measurements for the T_1 and T_2 experiments, and from the rms baseplane noise for the NOE experiment. The *cpmgr2* program was used for fitting the data to obtain values for $1/T_1$ and $1/T_2$ by the Levenberg−Marquardt method (*35*). Uncertainties for $1/T_1$ and $1/T_2$ were obtained from Monte Carlo simulations and were typically 2−3% larger than those obtained from the least-squares fit. An initial estimate of the overal rotational correlation time, τ_m, was obtained from the ratio of the trimmed mean values of $1/T_2$ and $1/T_1$ and the program *tmest*.

Table 1: CSA and B_0 Dependence of ^{15}N $1/T_1$ and $1/T_2$ and Heteronuclear NOE[a]

	750 MHz		600 MHz	
CSA	R_1	% change	R_1	% change
−172	0.918	5.2	1.222	3.7
−160	0.873	−	1.178	−
−148	0.832	−4.7	1.139	−3.3

	750 MHz		600 MHz	
CSA	R_2	% change	R_2	% change
−172	15.737	5.2	13.882	3.8
−160	14.963	−	13.378	−
−148	14.244	−4.8	12.911	−3.5

	750 MHz		600 MHz	
CSA	NOE	% change	NOE	% change
−172	0.853	0.9	0.829	0.8
−160	0.845	−	0.822	−
−148	0.838	−0.8	0.816	−0.7

[a] Values calculated for an N−H bond vector with $S^2 = 0.9$, $\tau_e = 1$ ps, and $\tau_m = 10.35$ ns at 750 and 600 MHz (^{1}H frequency) using the simplified spectral density function [$J(\omega)$] of Lipari and Szabo (*21*) and standard equations for ^{15}N $1/T_1$ and $1/T_2$ and heteronuclear NOE (*49*).

Model-Free Analysis of ^{15}N Relaxation Data. The program ModelFree version 3.1 (*35*) was used to calculate dynamical parameters from the ^{15}N T_1 and T_2 and NOE values for 138 of the 147 residues of GMH4CO, according to the “model-free” (*20*, *21*) and “extended model-free” (*36*) formalism. Five different models were evaluated for each amide ^{1}H−^{15}N pair in the initial stages of analysis as described by Palmer and co-workers (*37*) with the following fitted parameters for each model: (1) S^2, (2) S^2 and τ_e, (3) S^2 and R_{ex}, (4) S^2, τ_e, and R_{ex}, and (5) S_f^2, S^2, and τ_e. The N−H bond length was taken to be 1.02 Å, and a value of −160 ppm was used for the ^{15}N chemical shift anisotropy (CSA). Initial model-free calculations which used uncertainties for the experimental values from Monte Carlo analysis of the exponential fits of T_1 and T_2 resulted in large sum-squared errors (SSEs) for many residues, even for models with three adjustable parameters (models 4 and 5), suggesting the possibility of errors in the measured relaxation values, underestimation of experimental uncertainties, or deviation from model spectral densities.

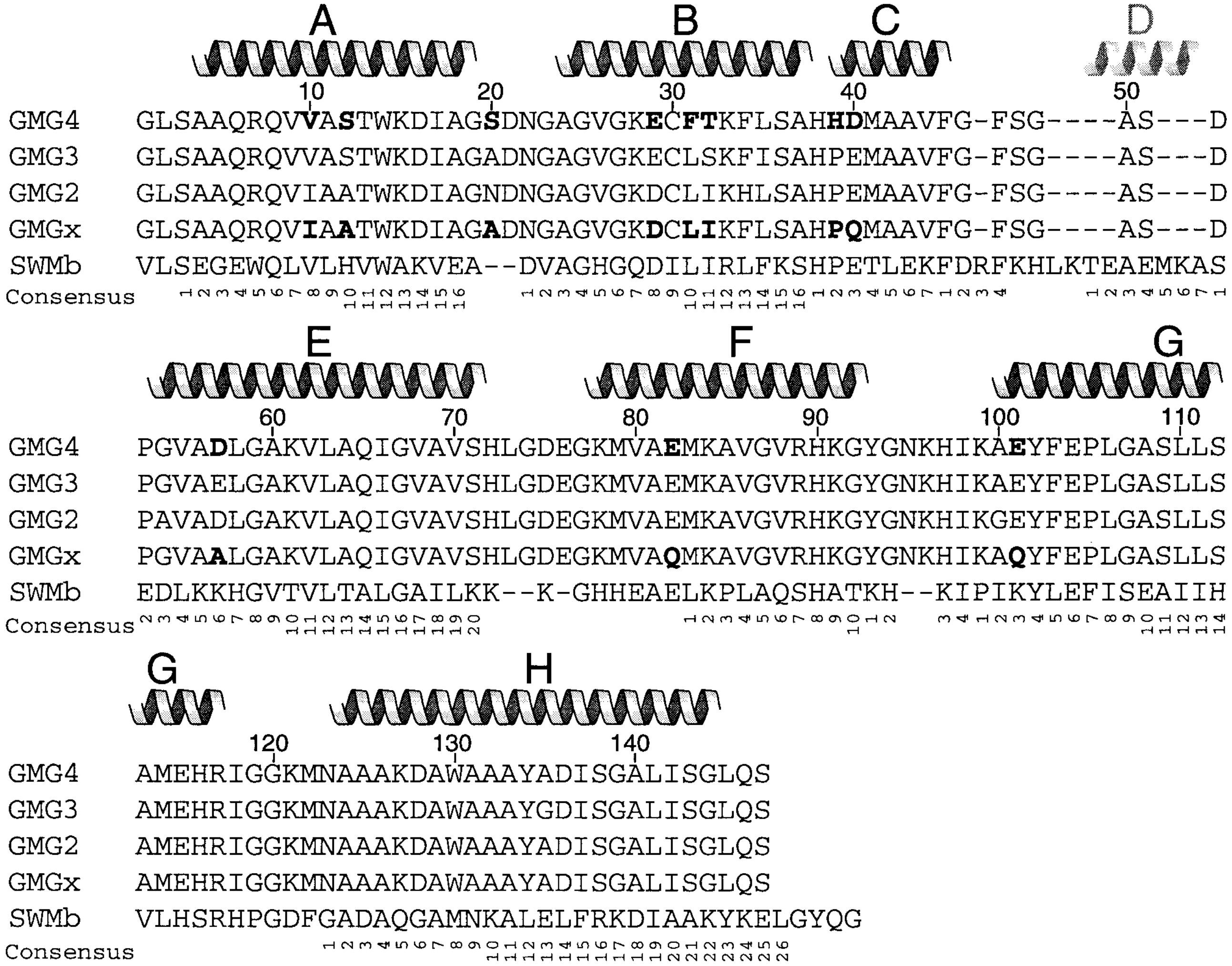
Systematic errors in the measurement of ^{15}N T_2 values and their propagation into model-free analysis have been discussed elsewhere (*38*−*40*). Underestimation of T_2 values

---

[2] Peakpick software may be downloaded from the NMRFAM WWW site at http://www.nmrfam.wisc.edu/roger/Software/peakpick/pp_main.html.

[3] BMRB accession number 4038 (^{1}H, ^{15}N, and ^{13}C chemical shifts of GMH4CO).

[4] ModelFree 3.1, fitting programs, and Felix macros may be obtained from A. Palmer at http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/.



GMG4 GLSAAQRQV**VAS**TWKDIAG**S**DNGAGVGK**E**C**FT**KFLSAH**HD**MAAVFG-FSG----AS---D
GMG3 GLSAAQRQVVASTWKDIAGADNGAGVGKECLSKFISAHPEMAAVFG-FSG----AS---D
GMG2 GLSAAQRQVIAATWKDIAGNDNGAGVGKDCLIKHLSAHPEMAAVFG-FSG----AS---D
GMGx GLSAAQRQV**IAA**TWKDIAG**A**DNGAGVGK**D**C**LI**KFLSAH**PQ**MAAVFG-FSG----AS---D
SWMb VLSEGEWQLVLHVWAKVEA--DVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKAS

GMG4 PGVA**D**LGAKVLAQIGVAVSHLGDEGKMVA**E**MKAVGVRHKGYGNKHIKA**E**YFEPLGASLLS
GMG3 PGVAELGAKVLAQIGVAVSHLGDEGKMVAEMKAVGVRHKGYGNKHIKAEYFEPLGASLLS
GMG2 PAVADLGAKVLAQIGVAVSHLGDEGKMVAEMKAVGVRHKGYGNKHIKGEYFEPLGASLLS
GMGx PGVA**A**LGAKVLAQIGVAVSHLGDEGKMVA**Q**MKAVGVRHKGYGNKHIKA**Q**YFEPLGASLLS
SWMb EDLKKHGVTVLTALGAILKK--K-GHHEAELKPLAQSHATKH--KIPIKYLEFISEAIIH

GMG4 AMEHRIGGKMNAAAKDAWAAAYADISGALISGLQS
GMG3 AMEHRIGGKMNAAAKDAWAAAYGDISGALISGLQS
GMG2 AMEHRIGGKMNAAAKDAWAAAYADISGALISGLQS
GMGx AMEHRIGGKMNAAAKDAWAAAYADISGALISGLQS
SWMb VLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG

FIGURE 1: Amino acid sequences of *G. dibranchiata* monomeric globins and sperm whale myoglobin (SWMb). The sequences of the three GMG proteins (GMG2, GMG3, and GMG4) as well as the sequence (GMGx) used in refinement of *Glycera* monomeric hemoglobin crystal structures are shown and aligned with SWMb and the globin consensus numbering as in Bashford et al. (*7*). The location of helices in GMH4CO is shown above sequences with residue numbering according to the GMG4 sequence. The globin consensus numbering is shown below sequences. The D helix, missing in GMG proteins and some other members of the globin family, is shaded gray.

(the most likely systematic error) will result in the overestimation of τ_m, and the subsequent overestimation of the generalized order parameter, S^2.

Recent measurements of the ^{15}N CSA for ubiquitin (*41*) suggest it may vary significantly from the average of −160 ppm, which is generally used in analysis of ^{15}N relaxation parameters. Because the CSA mechanism contributes significantly to the overall transverse and longitudinal ^{15}N relaxation rates, especially at high field strengths, deviation of the CSA from the assumed value of −160 ppm will decrease the goodness of fit for model-free calculations. Table 1 shows the effect of variation in CSA for the ^{15}N R_1 and R_2 and NOE values at 600 and 750 MHz for a typical N−H bond vector of GMH4CO. Variation of the CSA within the range observed for proteins will change the observed values of $1/T_1$ and $1/T_2$ at 750 MHz by up to 5%, significantly more than the experimental uncertainties of 1% for the $1/T_1$ and $1/T_2$ measurements obtained for this protein at that field strength from Monte Carlo simulations.

To account for the variability in the ^{15}N CSA and improve the model-free analysis of the ^{15}N relaxation parameters for GMH4CO, the uncertainties for R_1 and R_2 at 750 MHz were set to 5% of their values. A similar approach has been described by Zhang et al. (*42*). Significantly lower SSE values were obtained with model-free calculations as a result, as well as slightly increased uncertainties in the values of the fitted parameters. In addition, data for a number of residues that had initially required a two-parameter model (model 2 or 3) were now fitted adequately by the one-parameter model, model 1 (S^2 only). Only one- and two-parameter fits (models 1−3) were used in the final model-free analysis, as low errors (SSE < 15) were obtained for results from all residues but Leu^2. Data from this residue were not adequately fitted with any of the five models.

Determination of Amide Exchange Rates and Protection Factors. Amide H−D exchange decay curves were fitted using the Levenberg−Marquardt algorithm to a three-parameter single-exponential function with the program proFit 5.0.1 (Cherwell Scientific) to yield amide exchange rates (k_{obs}). The program HXPRED was used to calculate intrinsic exchange rates (k_{int}) from the primary protein sequence at pH 5.0 and 20 °C (*43*). Protection factors (P) were calculated with $P = k_{int}/k_{obs}$.

RESULTS

Solution Structure of GMH4CO. Figure 2 shows the family of 29 GMH4CO DG−SA conformers that represent

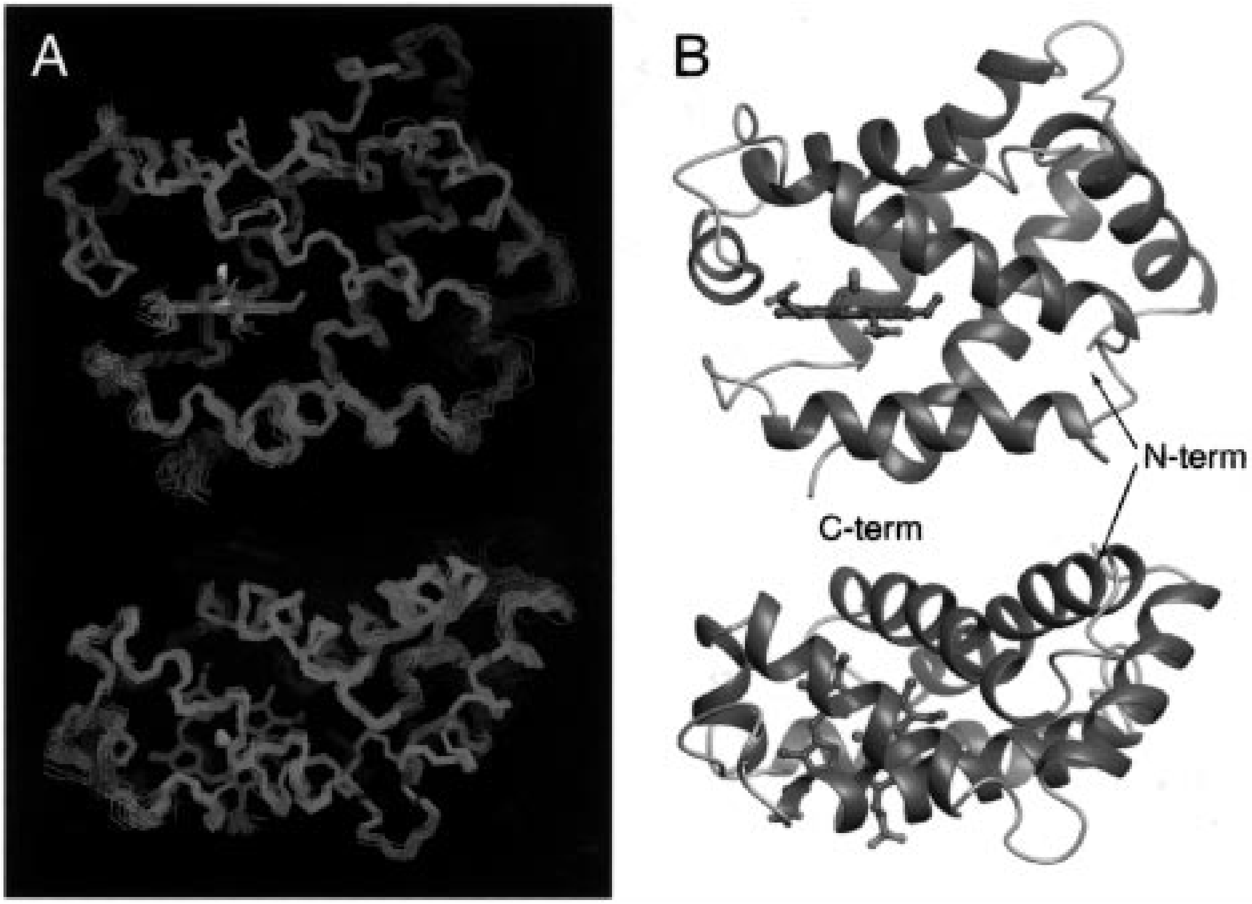


FIGURE 2: Ensemble and refined average NMR structures of GMH4CO created with MOLMOL (*51*). (A) Family of 29 GMH4CO conformers superimposed on the average backbone coordinates. Two views are shown, related by a 90° rotation about the horizontal axis (i.e., front and top views). Helices are colored green and connecting loops gray, and the heme cofactor is shown in red. The carbon monoxide ligand is also shown. (B) Ribbon diagram representations of the refined average structure of GMH4CO oriented as in panel A.

the structure, including the heme cofactor and carbon monoxide ligand. The familiar globin fold is observed, with seven helices connected by six loops of various lengths. The D helix of the globin family is not present in GMH4CO (Figure 1), as with the α-chain of human hemoglobin (*7*). The protein backbone is well-defined throughout the structure of GMH4CO, although residues of the seven helices are slightly more ordered than the loops and turns, with average backbone rms deviations of 0.39 (helices) and 0.61 Å (all other nonterminal residues). The approximate locations of the seven helices in GMH4CO are indicated at the top of Figure 1. The precise extent of each helix (residues 4−17, 24−37, 39−45, 54−71, 78−92, 100−117, and 124−144) defined by the Procheck-NMR program (*53*) for the family of structures is found to correspond well to the definitions obtained previously from the chemical shift index and analysis of NOEs (residues 4−20, 24−37, 40−47, 55−73, 77−93, 102−120, and 125−147) (*14*). Most side chains which are not completely solvent-exposed are also highly ordered in the family of conformers, as reflected by an rmsd of 0.92 Å for all non-hydrogen atoms of the protein.

The orientation of the heme cofactor in GMH4CO is reversed relative to that in myoglobin (*12*, *14*). The position of the heme within the protein is defined very well, with an average rmsd from the mean for all non-hydrogen heme atoms of 0.50 Å. When the atoms of the propionate groups are excluded, the average heme heavy atom rmsd is 0.20 Å. While the CO ligand in the crystal structure of MbCO is tilted by either 40 or 60° (*44*), in the carbon monoxy GMH crystal structure the ligand is nearly perpendicular to the heme plane and slighly displaced from the iron atom (*18*). Because there are no NMR spectroscopic constraints on the position of the CO ligand, distances taken from the MbCO crystal structure were used to fix its position, and the CO bond was taken to be normal to the plane of the heme as seen in the GMH crystal structure.

A total of 2550 NOE and dihedral angle constraints (from quantitative *J* correlation) were used in the final refinement of the structure of GMH4CO, as summarized in Table 2. This corresponds to an average of 17 constraints per residue, while the actual distribution of constraints by residue is indicated in Figure 3A. Of the 201 constraints involving the heme cofactor, 158 are heme−protein NOEs identified in 2D and 3D heteronuclear edited and filtered NOESY experiments (*14*). Constraints between the heme and protein, which constitute more than 25% of the 618 long-range NOEs, are essential for positioning the heme cofactor within the protein and contribute significantly to determining the global fold of the protein.

Of the 50 structures calculated in the final round of refinement, an ensemble of 29 conformers was selected for analysis, consisting of those for which no NOE constraint was violated by more than 0.2 Å and no dihedral constraint was violated by more than 2°. The average atomic coordinates of these 29 conformers were refined to regularize the structure, and Table 2 shows the final X-PLOR energies and

Table 2: X-PLOR Statistics for 29 GMH4CO Conformers[a]

constraint type	number
long-range NOEs	618
medium-range NOEs	536
sequential NOEs	506
intraresidue NOEs	745
dihedrals (ϕ, χ^1, and χ^2)	145
total constraints	2550

X-PLOR energy	$\langle \mathrm{SA} \rangle$	$\langle \overline{\mathrm{SA}} \rangle_r$
E_{total}	76 ± 10	66
E_{bond}	4.2 ± 0.7	2.8
E_{angle}	42 ± 4	39
E_{improper}	12 ± 1	11
E_{vdw}[b]	11 ± 3	7
E_{noe}[c]	5 ± 2	5
E_{cdih}[d]	0.5 ± 0.5	1.8
$E_{\mathrm{Lennard-Jones}}$[e]	-540 ± 18	-535

rmsd from ideal geometry	$\langle \mathrm{SA} \rangle$	$\langle \overline{\mathrm{SA}} \rangle_r$
bonds (Å)	0.0014 ± 0.0001	0.0011
angles (deg)	0.26 ± 0.01	0.25
impropers (deg)	0.26 ± 0.01	0.25

rmsd from experimental constraints	$\langle \mathrm{SA} \rangle$	$\langle \overline{\mathrm{SA}} \rangle_r$
distance constraints[f]	0.006 ± 0.001	0.006
dihedral constraints	0.23 ± 0.10	0.45

atomic rmsd (Å)	N, C^{α}, and C′	all non-H
$\langle \mathrm{SA} \rangle$ vs $\langle \overline{\mathrm{SA}} \rangle$	0.48 ± 0.10	0.92 ± 0.11
$\langle \mathrm{SA} \rangle$ vs $\langle \overline{\mathrm{SA}} \rangle_r$	0.71 ± 0.12	1.28 ± 0.16
$\langle \overline{\mathrm{SA}} \rangle$ vs X-ray	1.36	N/A
$\langle \overline{\mathrm{SA}} \rangle_r$ vs X-ray	1.45	N/A

[a] Notations are as follows. $\langle \mathrm{SA} \rangle$ is the ensemble of 29 X-PLOR DG−SA conformers. $\langle \overline{\mathrm{SA}} \rangle$ is the average coordinates obtained from a least-squares superposition of all backbone N, C^{α}, and C' atoms, except those of residues 1 and 147. $\langle \overline{\mathrm{SA}} \rangle_r$ is the refined average coordinates. [b] The X-PLOR F_{repel} function was used to simulate the van der Waals potential with atomic radii ranging from 0.9 times their CHARMM (*50*) values at high temperatures to 0.77 times their CHARMM values at low temperatures (*32*). [c] NOE-derived distance restraints were applied with a force constant of 50 kcal mol^{-1} Å^{-2}, using a soft square-well potential with a switching distance of 0.5 Å, an asymptote of 0.1, and a soft exponent of 1. [d] Dihedral angle constraints were given force constants of 200 kcal mol^{-1} rad^{-2} and applied at the beginning of the annealling−refinement stage. [e] The Lennard-Jones potential was not used during refinement but was calculated after the final stage using nonbonded parameters from the PARMALLH6 parameter set. [f] No NOE constraints were violated by more than 0.2 Å, and no dihedral constraints were violated by more than 2° in the family of 29 conformers (accepted from a total of 50 calculated conformers).

structural statistics for both the family of structures and the refined average structure. The overall average rms deviations for the family from the average coordinates for the backbone (N, C^{α}, and C′) and all non-hydrogen atoms were 0.48 and 0.92 Å, respectively. The small residual violations of experimental constraints and covalent geometry, in addition to the small atomic rms deviations, reflect the quality of the calculated structures. A Ramachandran plot analysis of all 29 structures places 83% of non-glycine residues in the most favored regions, with only 1% in disallowed regions. Unusual or poorly defined backbone conformations in the loop regions preceding (72−78) and following (93−98) the F helix account for most of the residues found outside the most favored regions of the Ramachandran plot.

Table 3: Approximate Backbone ϕ and ψ Values for Residues That Display Conformational Heterogeneity in Calculated Structures of GMH4CO[a]

residues	ψ_i	ϕ_{i+1}	n[b]
H^{72}−L^{73}	−70	+60	16
H^{72}−L^{73}	+60	+70	13
H^{97}−I^{98}	+60	−70	22
H^{97}−I^{98}	−80	+60	7
G^{119}−G^{120}	−30	+120	24
G^{119}−G^{120}	+110	−120	5

[a] Values were estimated by visual inspection of individual Ramachandran plots for each residue showing backbone dihedral values for all 29 NMR structures. [b] Number of structures out of 29 that contain the conformational subpopulation specified by ψ_i and ϕ_{i+1}.

Figure 3B shows the average backbone atomic rms deviations for each residue in GMH4CO, which are generally below 0.5 Å. Disordered regions include only the amino and carboxyl termini, and residues Gly^{74}−Lys^{78}, which connect the E and F helices. Slightly increased backbone rmsds are observed for residues of the CD loop (Ser^{48}−Pro^{53}), F helix, and FG loop (Met^{79}−Ile^{98}). The circular variances for the backbone (ϕ and ψ) and side chain (χ^1 and χ^2) torsion angles are shown in panels C and D of Figure 3, respectively. Increased rmsds and circular variances of the backbone atoms and dihedral angles are correlated in some instances.

Both the backbone atom rmsd values (Figure 3B) and the ϕ and ψ circular variances (Figure 3C) reflect the precision with which the backbone is defined in the structural ensemble, but the latter suggests clearly that three positions in the sequence display significant backbone angle heterogeneity: His^{72}−Leu^{73}, His^{97}−Ile^{98}, and Gly^{119}−Gly^{120}. These residues are located in the EF, FG, and GH loops, but only in the EF loop are markedly higher atomic rmsds observed (Figure 3B). In each instance, two conformational subclasses are observed within the structural ensemble which are related by a concerted change in the values of ψ_i and ϕ_{i+1}. Table 3 shows the average ϕ and ψ values for the residues in each conformer and the size of the subpopulation. Residues 119 and 120 are solvent-exposed glycines defined by relatively few experimental constraints (each has only one long-range NOE), and each can undergo a backbone rearrangement without significant displacement of the surrounding structure or violation of constraints. The other pairs of residues (His^{72}−Leu^{73} and His^{97}−Ile^{98}) are located at opposite ends of the F helix.

The 29 NMR solution conformers of GMH4CO are compared with the X-ray crystal structure of a mixture of GMH proteins (*18*) in Figure 3E. The solid line shows the C^{α} rmsd at each residue between the ensemble of 29 solution conformers and the X-ray structure. The dashed line shows a similar comparison between the family of NMR structures and a model structure, which was constructed by substitution of the residues of GMH4CO into the crystal structure (*6*). Sizable differences between the NMR and X-ray structures are limited to the residues of the N and C termini and the AB, EF, and FG loops. It is interesting to note that the general profile of the rmsd plots resembles that of the crystallographic *B*-factors (dotted line).

^{15}N Relaxation Measurements. Data from 138 of the 147 residues of GMH4CO were included in the analysis of ^{15}N relaxation parameters and backbone dynamics. Overlap in

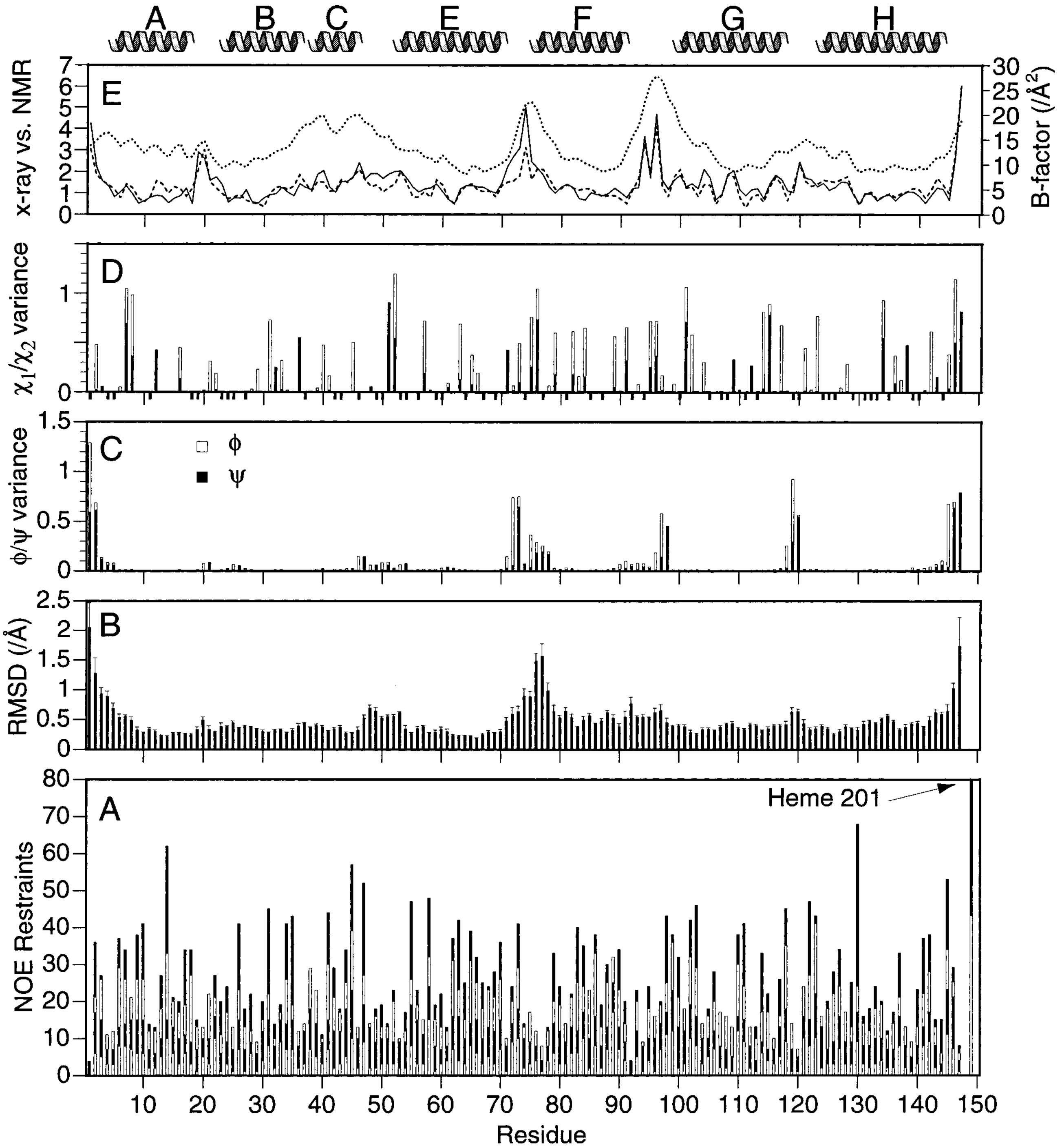


FIGURE 3: Structural statistics for GMH4CO as a function of residue. The location of helices is indicated at the top. (A) NOE constraints, with contributions from intraresidue, sequential, medium-range ($1 < i - j < 5$), and long-range ($i - j > 4$) constraints shown as individual white and black bars from bottom to top, respectively. Intraresidue heme and heme−protein NOEs are represented as residue 149. (B) Average backbone (C', C^{α}, and N) atomic rms deviations of the family of 29 conformers from their average coordinates. (C) Backbone ϕ and ψ dihedral circular variances, shown as stacked black and white bars, respectively. The circular variance indicates the range of rotation observed in the family of conformers for a torsion angle from 0 (completely restricted to one conformation) to 1 (essentially all values observed, from 0 to 360°). (D) Side chain χ^1 and χ^2 dihedral angle circular variances, shown as stacked black and white bars, respectively. Black bars below the horizontal axis indicate Gly or Ala residues for which no circular variance values are obtained. (E) (Solid line) Average rms deviation (in angstroms) of the 29 NMR GMH4CO structures (C^{α} atoms only) from the crystal structure of a *G. dibranchiata* monomeric hemoglobin (PDB entry 1HBG) (*18*) and (dashed line) from the model structure of GMH4CO derived from 1HBG (*6*). (Dotted line) The average crystallographic *B*-factors for the backbone atoms of each residue of the crystal structure (*18*).

the $^{15}N-^{1}H$ HSQC spectrum prevented the determination of ^{15}N T_2 values for six residues at 600 MHz, and the N-terminal amino group and two prolines account for the other three residues. Figure 4 shows the values of ^{15}N R_1 and R_2 relaxation rates and heteronuclear $^{1}H-^{15}N$ NOEs obtained for GMH4CO. Values for all parameters are relatively uniform for the helices, with slight decreases in R_2 and NOE observed for some of the loop residues. The average R_2 and R_1 values at 750 MHz for residues with values within one standard deviation of the overall average R_2 and R_1 values (the trimmed means) were 15.0 and 0.884 s^{-1}, respectively. An intial estimate of the overall correlation time of the

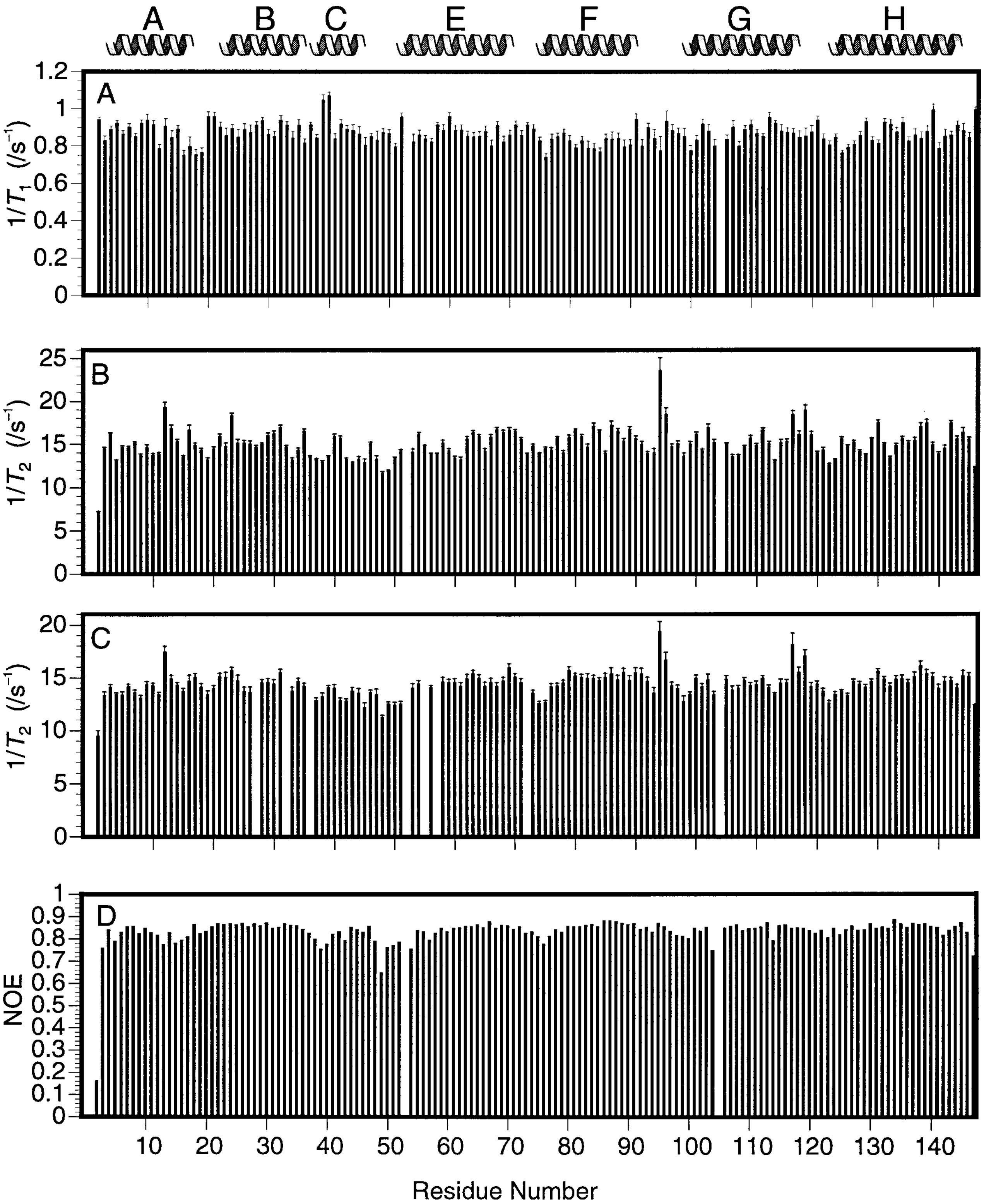


FIGURE 4: Backbone ^{15}N relaxation parameters for individual residues of GMH4CO at 20 °C: (A) ^{15}N 1/T_1 and (B) 1/T_2 values at 750 MHz (^{1}H frequency) and (C) ^{15}N 1/T_2 values at 600 MHz and (D) ^{1}H−^{15}N NOE values at 750 MHz. The location of helices is indicated at the top. Gaps correspond to prolines or to residues whose values could not be measured because of spectral overlap.

protein (τ_m = 10.3 ns) was obtained from the ratio of the trimmed means of the R_2 and R_1 values at 750 MHz (R_2/R_1 = 17.0). Some residues displayed larger than average ^{15}N R_2 values with a pronounced field dependence. Examples include Thr^{13}, Lys^{96}, Gly^{119}, and most notably Asn^{95}, which has R_2 values of 19 and 23 s^{-1} at 600 and 750 MHz, respectively. These line-broadening effects are presumably due to chemical exchange processes on the microsecond to millisecond time scale (vide infra).

Model-Free Analysis. The final model-free analysis was performed after model selection for each residue on the basis of preliminary fits to each of the five models for all residues, a procedure similar to that descirbed by Mandel et al. (*37*); the results are summarized in Figure 5. Global optimization

Table 4: Average Values for S^2, Backbone Atomic rms Deviations, and log P (Amide Protection Factors) for the Seven α-Helices of GMH4CO[a]

helix (residues)	S^2	rmsd (Å)	log P
A (4−17)	0.92 ± 0.03	0.40 ± 0.20	3.8 ± 2.0
B (24−37)	0.95 ± 0.03	0.35 ± 0.15	5.3 ± 1.7
C (39−45)	0.89 ± 0.03	0.33 ± 0.12	1.9 ± 1.7
E (54−71)	0.95 ± 0.03	0.29 ± 0.11	5.1 ± 1.7
F (78−92)	0.92 ± 0.04	0.56 ± 0.08	4.7 ± 1.7
G (100−117)	0.94 ± 0.04	0.36 ± 0.07	3.7 ± 1.7
H (124−144)	0.94 ± 0.03	0.42 ± 0.07	5.6 ± 1.7

[a] Values represent averages taken over the residues indicated for each helix. Uncertainties are standard deviations of those averages, representing the distribution of values across the residues of each helix, rather than experimental uncertainties.

of the overall molecular tumbling correlation time produced a value for τ_m of 10.5 ns. Only one- and two-parameter fits (models 1−3) were used in the final analysis, with the choice between model 2 or 3 indicated by a non-zero value for either τ_e in Figure 5C (model 2) or R_{ex} in Figure 5D (model 3). For Leu^2, the two-parameter models proved to be insufficient, but neither of the three-parameter models gave a significantly lower sum-squared error (SSE). Of the date from the 138 residues analyzed, 20 were fitted adequately by model 3, 14 by model 2, and 104 by model 1. Figure 5E shows the individual SSE values for all residues from the final model-free calculation.

The majority of residues of GMH4CO have high generalized order parameters (S^2), as shown in Figure 5B, and very fast internal motions ($\tau_e \rightarrow 0$). Significant internal mobility on the picosecond to nanosecond time scale appears to be confined to three regions of the protein; in addition to the residues of the N terminus, residues that have slower internal motions (τ_e > 15 ps) are in the vicinity of the C helix and CD loop (residues 39, 40, 49−52, and 54). The first two residues of the C helix and the five residues preceding the start of the E helix, which were fitted adequately by model 2, display the lowest S^2 values found in the protein (Figure 5B). Slightly reduced values of S^2 (<0.85) also are observed for residues in other loop regions of GMH4CO (residues 75, 76, 99, and 123).

In addition to the picosecond to nanosecond motions characterized by reduced S^2 values and τ_e values greater than 15 ps, residues involved in slow (microsecond to millisecond) time scale motions are those whose relaxation data are best fitted by model 3, which includes a chemical exchange term (R_{ex}) to account for additional line broadening observed in T_2 measurements. Because its effect on T_2 scales with increasing field strength, the ability to discern the presence of chemical exchange broadening is significantly enhanced when T_2 values at multiple fields are available. Many residues of the F helix have T_2 values which scale with field strength and are therefore characterized by significant values of R_{ex} (Figure 5D). This correlates well with the average backbone atomic rmsd, which is higher for the F helix than any other, as shown in Table 4. Residues that have slower motions of this type generally have normal S^2 values; the average S^2 values for the F helix are not significantly different from those in the other helices (Table 4).

Hydrogen Exchange. All the helices of GMH4CO display significantly large amide protection (Figure 5A and Table 4). Residues in loops and at the start of each helix have minimal protection factors, consistent with the absence of hydrogen bonding for those regions. The first three residues of the α-helix are not H-bonded to backbone oxygens, and since the C helix extends only two turns, significant protection of only about four amide hydrogens is expected. Therefore, the lower average protection factor of the C helix relative to the others is consistent with the pattern of hydrogen bonding in the calculated structures, and does not necessarily reflect dynamic instability.

Table 5: Comparison of Side Chain Hydrogen Bonds Observed in NMR and X-ray Crystal Structures of GMH4CO

donor	acceptor	NMR D−A/H−A[a]	NMR (min) D−A/H−A[b]	X-ray D−A (Å)
Trp^{130} N$^{\epsilon1}$	Val^{10} O	3.08/2.20	3.17/2.36	−
Trp^{130} N$^{\epsilon1}$	Ile^{10} O	−	−	4.28
Trp^{130} N$^{\epsilon1}$	Thr^{13} O$^{\gamma1}$	2.97/2.89	2.79/2.89	2.82
His^{90} N	Val^{86} O	3.10/2.20	2.77/1.83	3.18
His^{90} N$^{\delta1}$	Val^{86} O	3.09/2.29	3.04/2.14	2.82

[a] Average hydrogen bond donor−acceptor and hydrogen−acceptor distances for 29 GMH4CO structures. [b] Hydrogen bond donor−acceptor and hydrogen−acceptor distances for the refined average GMH4CO structure.

The NH groups of two side chains display large protections factors. The H$^{\delta1}$ of His^{90} and the H$^{\epsilon1}$ of Trp^{130} are observed in all the 2D HSQC spectra of the H−D exchange series. Both of these exchanged too slowly over the time course (∼13 days) to allow fitting of their decays, suggesting that they are also involved in hydrogen bonds. Inspection of the family of conformers reveals that both the side chain H$^{\delta1}$ and backbone H^N of His^{90} are hydrogen-bonded to the main chain carbonyl of Val^{86}. It also appears that the backbone carbonyl oxygen of Val^{10} is the most likely hydrogen bond acceptor for the Trp^{130} side chain H$^{\epsilon1}$. Table 5 summarizes the average distances for these side chain···main chain hydrogen bonds.

DISCUSSION

The solution structure of the monomeric hemoglobin component IV−carbon monoxide complex (GMH4CO) is very precise, with most backbone and internal side chain atomic rms deviations from the average structure of less than 0.5 Å. The position of the heme cofactor is well defined by a large number (158) of heme−protein NOE constraints.

Heme Pocket Description. Structural features of the proximal side of the heme are highlighted in Figure 6A which includes the heme cofactor, carbon monoxide ligand, and surrounding side chains. Important heme−protein interactions include the coordinate−covalent bond between the proximal histidine (His^{90}) side chain and the heme iron, the His^{90}···Val^{86} hydrogen bonds, the hydrogen bonds between the Arg^{89} and Tyr^{93} side chains and the heme propionate of the D pyrrole (*18*). While providing the only covalent link between heme and protein, the side chain of the proximal-histidine forms a hydrogen bond directly with the F helix backbone which serves both to stabilize the position of the ligand side chain and couple the structure and dynamics of the heme pocket to the entire F helix. In the globin family consensus, the negatively charged D propionate normally forms a salt bridge to a basic residue (usually Lys) at the FG1 position (*7*), where Tyr^{93} is found in GMH4CO (Figure 1). The presence of Arg^{89} one turn of helix away in the F7

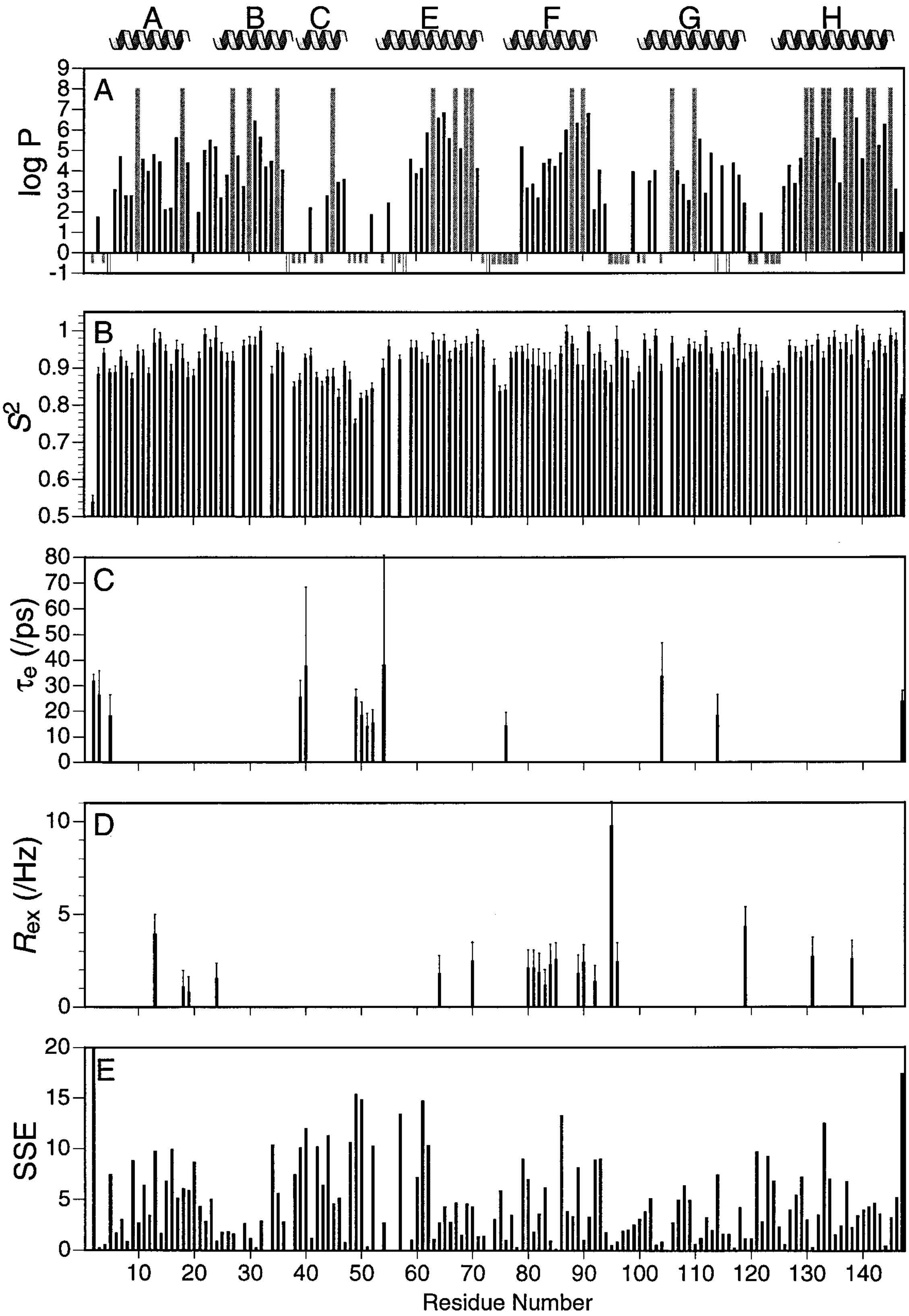


FIGURE 5: Comparison of backbone dynamics of GMH4CO from ^{15}N relaxation measurements and results from H−D amide exchange measurements. (A) Black bars indicate measured protection factors, and positive gray bars indicate residues for which exchange was too slow to detect over a 14 day period. Negative gray bars indicate residues for which complete exchange occurred during the dead time of the experiment (~30 min), and negative white bars indicate residues for which spectral overlap prevented unambiguous determination of protection factors. (B) Generalized order parameters, S^2, calculated from ^{15}N relaxation data using three different model-free parametrizations as described in the text. (C) Internal motional time constants, τ_e, for residues which were evaluated with model 2 in the final model-free analysis. (D) Chemical exchange terms, R_{ex}, for residues which were fit with model 3. (E) Sum-squared errors (SSE) from the final calculation of model-free parameters.

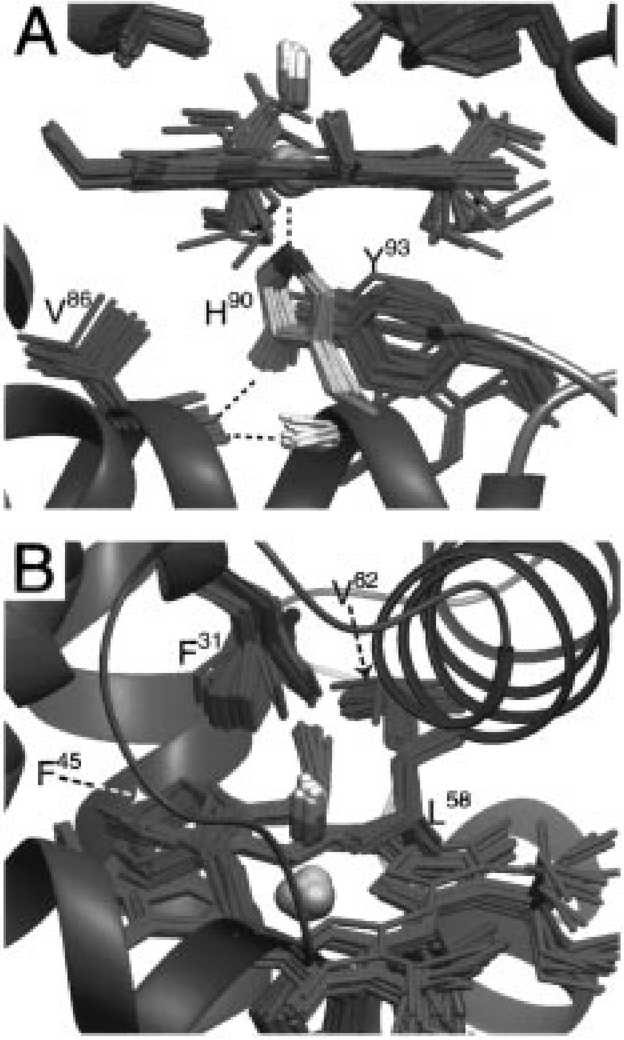


FIGURE 6: Detailed views of the heme pocket of GMH4CO created with MOLMOL (*51*). The heme and selected amino acid side chains from the family of 29 structures are shown. All structures were superimposed on the refined average structure, using backbone N, C^{α}, and C′ atoms. The backbone of the refined average structure is represented schematically by the ribbon. The non-hydrogen atoms of the heme (red), the heme iron (yellow), and the carbon monoxide ligand (red and white, above Fe^{2+}) are also shown. (A) Selected side chains from the proximal side of the heme. Side chains of Val^{86} and Tyr^{93} are shown in green. The proximal histidine, His^{90}, is shown in yellow, with the $N^{\epsilon 2}$ in blue, and the coordinate−covalent bond to the heme Fe is indicated by a dashed line. The hydrogen bonds between Val^{86} carbonyl oxygen and the H^{N} and $H^{\delta 1}$ of His^{90} are also indicated by dashed lines. (B) Distal side of the heme pocket. Hydrophobic side chains that surround the carbon monoxide ligand (Phe^{31}, Phe^{45}, Leu^{58}, and Val^{62}) are shown in green. The side chain of Leu^{31} from the X-ray crystal structure of a related GMH protein (*18*) is shown in blue.

position (which is usually leucine in the globin family) may compensate for the missing FG1 lysine. Unfortunately, no long-range NOEs define the orientation of the positively charged side chain of Arg^{89}; this makes it difficult to determine whether it interacts with the negatively charged D propionate of the heme. In the refined average structure, these carboxylate and guanidinyl groups are within 6 Å of each other, however.

Aromatic resonances of the Tyr^{93} side chain were not observed in any of the NMR spectra. This suggests that the signals are broadened by chemical exchange (ring flipping) on an intermediate time scale (*14*). Despite the complete lack of NOE constraints involving the ring protons, the side chain of Tyr^{93} is surprisingly well-ordered in the calculated structures. Its position in the structure is determined through NOEs to the H^{β} protons and a χ^1 dihedral angle constraint obtained from quantitative *J* correlation measurements of the $^3J_{H^{\beta}-C'}$ and $^3J_{H^{\beta}-N}$ couplings (*14*). Few unambiguous NOEs to the propionates could be obtained, since they fall in a highly degenerate region of the ^1H spectrum and could not be resolved by heteronuclear editing because the heme was not labeled with ^{13}C. This resulted in a wide range of rotamer states for the propionates in the family of conformers. Because neither the Tyr^{93} ring nor the heme propionates are well-defined, the average heme $O^{\delta 1}$−Tyr^{93} O^{ζ} distance (5.7 Å) in the ensemble of NMR structures is greater than that (2.5 Å) observed in the crystal structure. Despite this lack of precision, the general orientation of Tyr^{93} is still consistent with the formation of a hydrogen bond to the heme.

Figure 6B shows the distal heme environment, including four hydrophobic residues that surround the CO ligand creating an unusually apolar binding pocket. Two of these residues are ones that distinguish this protein from the globin family consensus, Leu^{58} (E7, normally the distal histidine, found in 97% of globin sequences) and Phe^{31} (B10, leucine with 90% frequency) (*7*). The other two are Phe^{45}, which is completely conserved in all globins, and Val^{62}, which is also highly conserved. These residues and the heme combine to completely isolate the ligand binding site from solvent access, and are themselves essentially buried in the interior of the protein, with only the heme propionate carboxylates having any solvent-exposed surface area. Lys^{61}, whose N^{ζ} atom is an average of 4.6 Å from a D propionate oxygen in the family of conformers (not shown), may be involved in a salt bridge, similar to the D propionate−Arg^{89} salt bridge on the proximal side.

Comparison of GMH4CO and a Related X-ray Structure. The crystals used in the X-ray structure determination were grown from protein isolated from the natural source which was known to be a mixture of isoforms (*18*). Recent mass spectrometric analysis has shown that these crystals contained a mixture of the three major monomeric forms, for which sequences are known (*15*), as well as a number of uncharacterized minor forms (*16*). Furthermore, the protein sequence used in refining the crystal structures has a calculated molecular mass different from those of all of the species observed by mass spectrometry (*17*).

The 11 amino acid differences between the sequence of GMG4 and that used in refining the crystal structure are highlighted with bold letters in Figure 1. The most notable difference is at the B10 position where GMG4 has Phe (whereas GMGx has Leu). This indicates that GMG4 was not the dominant component of the monomeric fraction used for crystallization. The positions of the two side chain types, which are compared in Figure 6B, have the same general orientation. Conformational disorder was described for the B10 leucine in the deoxy and CO-ligated crystal structures, and the map was interpreted in terms of multiple χ^2 rotamer populations (*18*, *45*). Since this space can accommodate the larger side chain of a phenylalanine, as in the GMH4CO structure described here, the rotational disorder of the smaller leucine side chain is not surprising. It is also interesting to note that, of the four distal side chains shown in Figure 6B, the side chain of Phe^{31} makes the closest contact with the

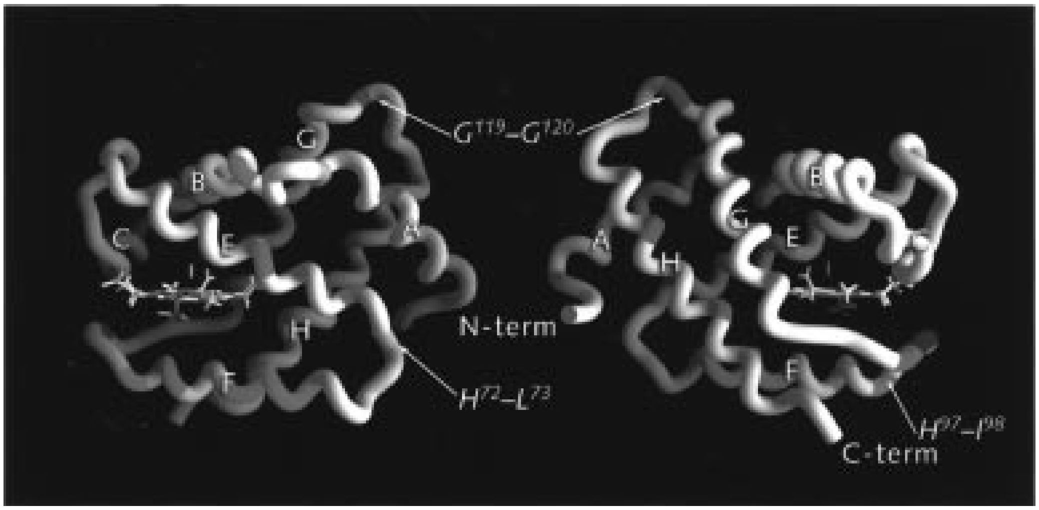


FIGURE 7: Ribbon diagram of GMH4CO highlighting of residues with significant R_{ex} contributions to ^{15}N T_2 values, created with the program GRASP (*52*). Residues that were best analyzed with models 1 and 2 in the final model-free calculations are shown in white, and those best analyzed with model 3 are shaded white → magenta → blue in proportion to increasing values of the fitted R_{ex} (Figure 5D). Views from front and rear are shown, with helices and termini labeled. In addition, the approximate locations of three sites in GMH4CO that display appreciable backbone heterogeneity in the family of structures are indicated (His^{72}−Leu^{73}, His^{97}−Ile^{98}, and Gly^{119}−Gly^{120}).

CO ligand, whereas in the crystal structure, the slight displacement of the CO from a ideal coordination geometry was found to be mediated by van der Waals contact with the methyl groups of Val^{62} and Leu^{58}. This is consistent with the idea that Phe^{31} contributes to the unusual ligand binding properties of this protein.

Most of the other amino acid differences may also be explained in terms of the sequence variability among the components of the monomeric hemoglobin fraction. Only Ala^{57} and the three glutamines at positions 40, 82, and 101 in the crystal structure are not represented in at least one of the sequenced components. The glutamines are likely all glutamic acid residues which could not be identified unambiguously by either chemical sequencing, amino acid analysis, or refinement of the electron density. The completely solvent-exposed side chain of residue 57 was not observed in the electron density, and alanine was substituted during refinement; residue 57 may actually be an aspartate or glutamate with a high degree of χ^2 rotational disorder, as in the NMR structure (Figure 3D).

Despite the differences in the protein sequences studied by crystallography and NMR, the resulting structures are very similar. The average backbone rms deviations from the X-ray crystal structure for both the family of 29 GMH4CO conformers and the refined average structure are slightly larger than the variation within the family of NMR structures (Table 2). Aside from the termini, differences in the backbone conformations of the NMR and crystal structures are confined to a few small regions between helices (Figure 2E).

Differences of ∼3 Å between the NMR and crystal structures are observed for the C$^{\alpha}$ atoms of Gly^{19} and Ser^{20}. These residues are part of a well-defined reverse turn formed in both structures by residues 18−21 at the end of the A helix. The discrepancy may be related to the substitution of Ser^{20} in GMH4CO for Ala^{20} in the crystal structure (Figure 1); however, in neither case does this solvent-exposed side chain appear to contact any other part of the structure. Tertiary interactions through which conformational perturbations could be transmitted to these residues include the conserved Asp^{21}−Arg^{117} and Asp^{16}−Lys^{121} salt bridges, but no clear indication of this type of long-range effect is observed in the structural comparison. Whereas the backbone rmsds of residues 18−21 are generally low (Figure 3B), a few residues in this region of the structure (Thr^{13}, Ala^{18}, Gly^{19}, Ala^{24}, and Gly^{119}) display evidence of slow motions (R_{ex}) (Figure 5D), and residues Gly^{119} and Gly^{120} were noted for displaying significant backbone heterogeneity in the familiy of NMR structures (Figure 3C and Table 3). The relative positioning of these residues in the structure of GMH4CO is highlighted in Figure 7. Perhaps the crystal structure represents one of two or more conformers that interchange on the microsecond to millisecond time scale.

Larger differences between the NMR and crystal structures are observed at the EF and FG loops. These residues display conformational heterogeneity in the NMR structure (Table 3), and they also have the highest *B*-factors in the crystal structure (Figure 3E). Residues Val^{70} and Leu^{73} of the EF loop make hydrophobic contacts with the A helix in which residues 10 and 12 are different in GMG4 and GMGx (Figure 1). Interestingly, for residues 72 and 73, the model of GMH4CO derived from the crystal structure (*6*) is much closer to the conformation found in the NMR structure than is the crystal structure, as indicated by the difference between the solid and dashed lines of Figure 3E. This may reflect a slight change in the packing of the end of the E helix against the A helix as a result of the change from Ile^{10} in the crystal structure to Val^{10} in GMH4CO. A similar scenario could be envisioned for the interaction of the FG loop with nearby residues of the BC loop, where His^{39} and Asp^{40} both differ from the sequence of the crystal structure; however, no corresponding perturbation in the FG loop is observed in the model structure. Internal dynamics may be a factor in the differences between the NMR and crystal structures for the residues in and around the F helix, as discussed below.

The NMR and X-ray structures differ in regard to one instance of side chain···main chain hydrogen bonding. Hydrogen bond donors are revealed by H−D exchange measurements, but the identity of the acceptor must be deduced from examination of the calculated structures.

While the pair of $His^{90}\cdots Val^{86}$ hydrogen bonds on the proximal side of the heme (Figure 6A) is consistent with previous crystal structures, the $Trp^{130}\cdots Val^{10}$ H bond observed here is not consistent with an earlier report of a H bond between the side chain of Trp^{130} and the side chain $O^{\gamma1}$ of Thr^{13} (*17*). As indicated in Table 5, while the Trp^{130} $N^{\epsilon1}\cdots Thr^{13}$ $O^{\gamma1}$ distance (2.97 Å) is reasonable, the $H\cdots O$ distance (2.89 Å) indicates that the geometry of this potential H bond is very poor compared to that of the $Trp^{130}\cdots Val^{10}$ arrangement. If this represents a real structural rearrangement, the substitution of Val^{10} for Ile^{10} could be a contributing factor.

Backbone Dynamics of GMH4CO. The overall molecular tumbling correlation time (τ_m) is an important parameter in the analysis of internal dynamics. All model-free $^{15}N-^{1}H$ dynamics calculations were performed with the assumption of an isotropic τ_m, estimated initially from an average T_1/T_2 ratio to be 10.3 ns and globally optimized in the final model-free analysis to be 10.5 ns. GMH4CO has a moderately nonspherical shape; the three principal components of the inertia tensor for GMH4CO calculated from the refined average NMR structure are 3.39×10^6, 4.70×10^6, and 5.96×10^6 Da Å^{-2}. It is possible that the model-free analysis would be improved by the inclusion of an anisotropic diffusion tensor and that some residues that showed better fits to the model that included contributions to T_2 from chemical exchange (R_{ex}, model 3) would then be characterized adequately by model 1 or 2. However, the inclusion of ^{15}N T_2 data at two field strengths makes it unlikely that the large R_{ex} terms required in the vicinity of the F helix are completely artifactual, since some of these residues (Asn^{95} and Lys^{96}) display large, field-dependent transverse relaxation rates (Figure 4B,C).

Trends observed in the model-free analysis of NMR relaxation can be correlated with a number of features of the NMR structures. Figure 7 shows the structure of GMH4CO with R_{ex} contributions highlighted in varying shades (white → magenta → blue) to indicate increasing values of R_{ex}. The lower precision of the F helix in the NMR structure (Figure 3B and Table 4) correlates with the number of its residues that exhibit motion on the slow time scale (magenta) in Figure 7. The locations of residue pairs with backbone dihedral angle heterogeneity (Figure 3C and Table 3) are also indicated in Figure 7. Asn^{95} has the largest R_{ex} term (colored blue in Figure 7) and is situated between Tyr^{93}, suspected of having ring-flipping dynamics in the intermediate exchange regime, and the ϕ and ψ heterogeneity of $His^{97}-Ile^{98}$. The residues in and following the C helix have reduced S^2 values (Figure 5B) and slightly larger rmsds (Figure 3B).

As shown in Figure 3E, the crystallographic *B*-factors have a profile similar to the rmsds between the NMR and X-ray structures. In contrast to some other comparisons of NMR relaxation-based backbone order parameters with crystallographic *B*-factors (*42*, *46*), high *B*-factors correlate here with regions of GMH4CO that have either lower order parameters (enhanced picosecond to nanosecond time scale internal motion) or significant R_{ex} factors (microsecond to millisecond time scale motion). Either type of internal motion may be expected to contribute to higher crystallographic thermal factors (*42*, *46*). The C helix and residues between the E and F helices, which have lower order parameters (Figure 5B), have correspondingly higher *B*-factors (Figure 3E). Residues Asn^{95} and Lys^{96} have the largest crystallographic *B*-factors in this protein, and exhibit similarly large R_{ex} contributions in solution (Figure 5D). Interestingly, unlike most residues of the F helix, residues 86−88 have no R_{ex} contributions and display low *B*-factors.

Slow Motions of the F Helix. The F helix amide protons of sperm whale myoglobin, which are structurally similar to those of GMH4, have very low protection factors (S. N. Loh, personal communication). This is, however, clearly not the case for the F helix of GMH4CO, whose protection factors are similar to those of the other seven α-helices (Figure 5A). At the same time, the combined results of the model-free analysis of ^{15}N relaxation data and structure calculations suggest that the F helix does have significant internal mobility, most likely on the microsecond to millisecond time scale. These motions may involve the helix as a rigid unit, rather than local unfolding events, to preserve the large protection factors over the entire length of the F helix. While a direct correlation between the dynamic and structural data is not warranted, the presence of localized disorder in the ensemble of calculated structures at each end of the F helix provides an example of a possible mechanism for this type of concerted motion. The lack of an R_{ex} contribution at Gly^{87} coincident with a minimum in the X-ray structure *B*-factors in the middle of the F helix seems to conflict with this interpretation. However, this could be explained by a model for F helix motion in which both ends pivot about a stationary point corresponding to the residues that lack the dynamic behavior of the surrounding regions. The exchange broadening of the ring protons of Tyr^{93} also could be accommodated by a model that postulates slow motions in the F helix.

Implications for Ligand Binding Kinetics. A large number of residues contact the heme in GMH4CO, forming both electrostatic and hydrophobic interactions. However, the coordinate−covalent bond between the side chain of His^{90} and the heme iron is the strongest single link between the protein and its cofactor. As a result, motions originating from the F helix can be expected to influence the heme environment to some degree. At this point, the motions of the F helix observed in this study cannot be correlated directly to the kinetics of ligand association−dissociation. Quantitative characterization of the time scales of these motions with measurements of ^{15}N $T_{1\rho}$ (*47*, *48*) values in GMH4CO, analysis of dynamics of other liganded forms (deoxy, O_2, and CN^-), and comparisons with myoglobin, GMH4 mutants, and other related proteins will permit a more meaningful comparison with association and dissociation rates.

ACKNOWLEDGMENT

We thank Dr. Andrew Lee, Dr. David LeMaster, and Dr. Arthur Palmer for helpful discussions and Dr. Jane Caldwell for generously providing software tools for structure refinement. We also thank Dr. Ed Mooberry and other members of the NMRFAM staff for their gracious help.

REFERENCES

1. Case, D. A., and Karplus, M. (1979) *J. Mol. Biol. 132*, 343−368.

2. Shulman, R. G., Ogawa, S., Wüthrich, K., Yamane, T., Peisach, J., and Blumberg, W. E. (1969) *Science 165*, 251−257.
3. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (1960) *Nature 185*, 422−427.
4. Srajer, V., Teng, T.-Y., Ursby, T., Pradervand, C., Ren, Z., Adachi, S.-I., Schildkamp, W., Bourgeois, D., Wulf, M., and Moffat, K. (1996) *Science 274*, 1726−1729.
5. Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1997) *Nat. Struct. Biol. 4*, 292−297.
6. Alam, S. L., Satterlee, J. D., and Edmonds, C. G. (1994) *J. Protein Chem. 13*, 151−164.
7. Bashford, D., Chothia, C., and Lesk, A. M. (1987) *J. Mol. Biol. 196*, 199−216.
8. Parkhurst, L. J., Sima, P., and Goss, D. J. (1980) *Biochemistry 19*, 2688−2692.
9. Rohlfs, R. J., Mathews, A. J., Carver, T. E., Olson, J. S., Springer, B. A., Egeberg, K. D., and Sligar, S. G. (1990) *J. Biol. Chem. 265*, 3168−3176.
10. Mintorovitch, J., and Satterlee, J. D. (1988) *Biochemistry 27*, 8045−8050.
11. Mintorovitch, J., Pelt, D. v., and Satterlee, J. D. (1989) *Biochemistry 28*, 6099−6104.
12. Cooke, R. M., and Wright, P. E. (1985) *Biochim. Biophys. Acta 832*, 365−372.
13. Alam, S. L., Dutton, D. P., and Satterlee, J. D. (1994) *Biochemistry 33*, 10337−10344.
14. Alam, S. L., Volkman, B. F., Markley, J. L., and Satterlee, J. D. (1998) *J. Biomol. NMR* (in press).
15. Teske, J. G., Edmonds, C. G., Deckert, G., and Satterlee, J. D. (1997) *J. Protein Chem. 16*, 139−150.
16. Green, B. N., Sannes-Lowery, K. A., Loo, J. A., Satterlee, J. D., Kuchumov, A. R., Walz, D. A., and Vinogradov, S. N. (1998) *J. Protein Chem. 17*, 85−97.
17. Arents, G., and Love, W. E. (1989) *J. Mol. Biol. 210*, 149−161.
18. Braden, B. C., Arents, G., Padlan, E. A., and Love, W. E. (1994) *J. Mol. Biol. 238*, 42−53.
19. Jardetzky, O. (1996) *Prog. Biophys. Mol. Biol. 65*, 171−219.
20. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4559−4570.
21. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.
22. Kay, L. E., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
23. Marion, D., and Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
24. Zijl, P. C. M. V., Johnson, M. O. N., Mori, S., and Hurd, R. E. (1995) *J. Magn. Reson., Ser. A 113*, 256−270.
25. Warren, W. S., Richter, W., Andreotti, A. H., and Farmer, B. T. (1993) *Science 262*, 2005−2009.
26. Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. C., and Zhu, G. (1994) *Methods Enzymol. 239*, 79−105.
27. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
28. Bax, A., Ikura, M., Kay, L. E., and Zhu, G. (1991) *J. Magn. Reson. 91*, 174−178.
29. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J. Biomol. NMR 6*, 135−140.
30. Chylla, R. A., and Markley, J. L. (1993) *J. Magn. Reson., Ser. B 102*, 148−154.
31. Nilges, M., Kuszewski, J., and Brünger, A. T. (1991) in *Computational Aspects of the Study of Biological Macromolecules by NMR* (Hoch, J. C., Ed.) Plenum Press, New York.
32. Brünger, A. (1992) *X-PLOR Manual Version 3.1*, Yale University Press, New Haven, CT.
33. Wall, L., and Schwartz, R. L. (1991) *Programming Perl*, O’Reilly and Associates, Inc., Sebastopol, CA.
34. Seavey, B. R., Farr, E. A., Westler, W. M., and Markley, J. L. (1991) *J. Biomol. NMR 1*, 217−236.
35. Palmer, A. G., Rance, M., and Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371−4380.
36. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 4989−4991.
37. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) *J. Mol. Biol. 246*, 144−163.
38. Simbrunner, J., and Stollberger, R. (1995) *J. Magn. Reson., Ser. B 109*, 301−309.
39. Ross, A., Czisch, M., and King, G. C. (1997) *J. Magn. Reson. 124*, 355−365.
40. Kemple, M. D., Yuan, P., Nollet, K. E., Fuchs, J. A., Silva, N., and Prendergast, F. G. (1994) *Biophys. J. 66*, 2111−2126.
41. Tjandra, N., Szabo, A., and Bax, A. (1996) *J. Am. Chem. Soc. 118*, 6986−6991.
42. Zhang, P., Dayie, K. T., and Wagner, G. (1997) *J. Mol. Biol. 272*, 443−455.
43. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins 17*, 75−86.
44. Kuriyan, J., Wilz, S., Karplus, M., and Petsko, G. A. (1986) *J. Mol. Biol. 192*, 133.
45. Archer, S. J., Ikura, M., Torchia, D. A., and Bax, A. (1991) *J. Magn. Reson. 95*, 636−641.
46. Dayie, K. T., Wagner, G., and Lefévre, J.-F. (1996) *Annu. Rev. Phys. Chem. 47*, 243−282.
47. Szyperski, T., Luginbühl, P., Otting, G., Güntert, P., and Wüthrich, K. (1993) *J. Biomol. NMR 3*, 151−164.
48. Akke, M., and Palmer, A. G. (1996) *J. Am. Chem. Soc. 118*, 911−912.
49. Abragam, A. (1961) *Principles of Nuclear Magnetism*, Clarendon Press, Oxford, England.
50. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
51. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
52. Nicholls, A., Sharp, K., and Honig, B. (1991) *Proteins: Struct., Funct., Genet. 11*, 281−296.
53. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.

BI980810B